

Article

Estimation and Inference for Spatio-Temporal Single-Index Models

Hongxia Wang¹, Zihan Zhao¹, Hongxia Hao¹ and Chao Huang^{2,*} ¹ School of Statistics and Data Science, Nanjing Audit University, Nanjing 211815, China² Department of Statistics, Florida State University, Tallahassee, FL 32306, USA

* Correspondence: chuang7@fsu.edu

Abstract: To better fit the actual data, this paper will consider both spatio-temporal correlation and heterogeneity to build the model. In order to overcome the “curse of dimensionality” problem in the nonparametric method, we improve the estimation method of the single-index model and combine it with the correlation and heterogeneity of the spatio-temporal model to obtain a good estimation method. In this paper, assuming that the spatio-temporal process obeys the α mixing condition, a nonparametric procedure is developed for estimating the variance function based on a fully nonparametric function or dimensional reduction structure, and the resulting estimator is consistent. Then, a reweighting estimation of the parametric component can be obtained via taking the estimated variance function into account. The rate of convergence and the asymptotic normality of the new estimators are established under mild conditions. Simulation studies are conducted to evaluate the efficacy of the proposed methodologies, and a case study about the estimation of the air quality evaluation index in Nanjing is provided for illustration.

Keywords: spatio-temporal correlation; spatio-temporal heterogeneity; reweighting estimation; local linear method; single-index models

MSC: 62H11



Citation: Wang, H.; Zhao, Z.; Hao, H.; Huang, C. Estimation and Inference for Spatio-Temporal Single-Index Models. *Mathematics* **2023**, *11*, 4289. <https://doi.org/10.3390/math11204289>

Academic Editors: Jin-Ting Zhang and Heng Lian

Received: 17 August 2023

Revised: 18 September 2023

Accepted: 12 October 2023

Published: 14 October 2023



Copyright: © 2023 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Recently, statistical analysis of spatio-temporal data has been widely used in environmental science, economics and social science, and other fields, and it plays an important role in both academic and industrial circles. Spatio-temporal data estimation methods have also been developed. Spatio-temporal correlation and spatio-temporal heterogeneity are two significant features of spatio-temporal data.

In the process of solving practical problems, spatio-temporal data models usually consider these two properties. Various spatio-temporal modeling methods have been applied to explore the effect of spatio-temporal correlation and heterogeneity. For example, ref. [1] proposed multiscale geographically weighted regression (MGWR). The authors of [2] extended geographically and temporally weighted regression (GTWR) ([3]) to multi-scale geographically and temporally weighted regression (MGTWR) in consideration of scale effects. The authors of [4] proposed the hetero-convolutional long short-term memory model based on the convolutional long short-term memory (ConvLSTM) neural network model. The authors of [5] adopted a local linear method to model spatio-temporal heterogeneity. A new nonparametric spatio-temporal inversion model was proposed by [6] for the gas emission problem considering both spatio-temporal heterogeneity and atmospheric inversion. Ref. [7] proposed to estimate the variance–covariance structure of the residuals by variography and removed the correlation by spatial filtering residuals.

In addition, the spatio-temporal trend is more complex, and spatio-temporal data can be regarded as a time series dataset with spatial information, which has the characteristics

of being dynamic, massive, and high-dimensional. In order to accurately predict the change in the spatio-temporal trend function, this paper combines the idea of a semi-parametric model, adds the influence behavior of multidimensional independent variables on dependent variables, and considers the spatio-temporal error model. The single-index model is one of the most popular semi-parametric models. The spatio-temporal single-index model that we are interested in here is:

$$Y_{i,t} = g(\beta_0^T X_{i,t}) + \varepsilon_{i,t} \quad i \in \mathcal{I}_n = \prod_{k=1}^N \{1, \dots, n_k\}, \quad t \in T_n = \{1, \dots, n_{N+1}\}, \quad (1)$$

where $Y_{i,t} \in \mathbb{R}$, $X_{i,t} \in \mathbb{R}^d$, $\mathbf{i} = (i_1 \cdots i_N)$ represents the spatial location, t represents the time, $n = (n_1, \dots, n_{N+1})$, $g(\cdot)$ is an unknown link function, β_0 is an unknown coefficient with $\beta_0^T \beta_0 = 1$ and the first component of β_0 is positive [8], and ε is a random error which has $E(\varepsilon|X) = 0$ almost surely.

The single-index model is one of the most popular semi-parametric models in applied statistics. Many authors have studied the estimation of the index coefficient β_0 , focusing on issues of \sqrt{n} estimability and efficiency. The methods include the average derivative method [9,10], the local linear method [11,12], the least squares method [13,14], the functional additive regression [15], and the empirical likelihood method [16,17]. Most regressions in single-index models assume that the random variables are independently and identically distributed. However, spatio-temporal correlation and heterogeneity are often found in the model error terms. For spatio-temporal correlation, we portray it through an α -mixing condition that is consistent with the spatio-temporal characterization. There are usually two types of assumptions for heterogeneity: the first assumption is that the variance function is purely nonparametric, and the second requires that the variance function has a dimension reduction structure like the mean function. This is often the case with models with a dimensionality reduction structure, such as generalized linear models. The second case holds in more general semiparametric settings for the central subspace and the central mean subspace, which have the same dimensions; see [18–20].

This work mainly deals with the estimation problem of spatio-temporal heterogeneity. By combining the idea of the local linear method and Nadaraya–Watson smoothing techniques, we propose a method for estimating the variance function in the single index model with heteroscedastic errors, and the resulting estimator based on a fully nonparametric variance function or a dimensionality reduction structure is proved to be consistent. For the parametric part, we use an efficient estimator of the parametric component by applying the iterative generalized least squares method in heteroscedastic generalized linear models and taking into account the estimated heteroscedasticity. We call this model fitting method reweighting estimation, and it is shown that the resulting constant coefficient estimators have smaller asymptotic variances than the rMAVE estimators, which neglect error heteroscedasticity but retain the same biases.

Throughout the rest of the paper, the symbols ' \xrightarrow{P} ' and ' \xrightarrow{D} ' denote convergence in probability and convergence in distribution, respectively. The symbol A^+ indicates the Moore–Penrose inverse of the symmetric matrix A . The notation $\|\cdot\|$ denotes the Euclidean distance.

This article is structured as follows. In Section 2, the estimation process of the rMAVE method is briefly described. In Section 3, two estimators for variance are proposed, namely, a nonparametric estimator of kernel smoothing type and a nonparametric estimator integrated with the dimensionality reduction structure. In Section 4, a reweighting estimation method and its asymptotic properties are given. In Section 5, the effectiveness of the proposed method is verified through simulations. The analyses for Nanjing air quality data are found in Section 6. Conclusions are presented in Section 7. The assumptions required for the theorem and the proof are in Appendices A and B.

2. A Brief Description of the rMAVE

Consider the regression model (1). Let (X, Y) be the random variable in model (1). The parameter β_0 is obtained by minimizing the expectation of the joint distribution (X, Y) ; see [21],

$$\min_{\beta} [E\{Y - E(Y|\beta^T X)\}^2]. \tag{2}$$

For any β , the conditional variance given $\beta^T X$ is:

$$\sigma_{\beta}^2(X) = E[E\{Y - E(Y|\beta^T X)\}^2|\beta^T X]. \tag{3}$$

It follows that:

$$E\{Y - E(Y|\beta^T X)\}^2 = E\{\sigma_{\beta}^2(X)\}.$$

Therefore, minimizing Expression (2) is equivalent to minimizing, with respect to β ,

$$E\{\sigma_{\beta}^2(X)\} \quad \text{subject to} \quad \beta^T \beta = 1. \tag{4}$$

Suppose that $\{(X_{i,t}, Y_{i,t})\}$ is a random sample from Formula (2). Let $g(v) = E(Y|\beta^T X = v)$. For any given $X_{j,\tau}$, a local linear expansion of $E(Y_{i,t}|\beta^T X_{i,t})$ at $X_{j,\tau}$ is:

$$E(Y_{i,t}|\beta^T X_{i,t}) \approx a + b\beta^T X_{ij}, \tag{5}$$

where $a = g(\beta^T X_{j,\tau})$, $b = g'(\beta^T X_{j,\tau})$, and $X_{i,t,j,\tau} = X_{i,t} - X_{j,\tau}$. Following the idea of local linear smoothing estimation, we can estimate $\sigma_{\beta}^2(X_{j,\tau})$ by exploiting the approximation:

$$\sum_{i \in \mathcal{I}_n} \sum_{t \in \mathcal{T}_n} \{Y_{i,t} - E(Y_{i,t}|\beta^T X_{i,t})\}^2 w_{(i,t),(j,\tau)} \approx \sum_{i \in \mathcal{I}_n} \sum_{t \in \mathcal{T}_n} [Y - \{a + b\beta^T X_{(i,t),(j,\tau)}\}]^2 w_{(i,t),(j,\tau)}, \tag{6}$$

where $w_{(i,t),(j,\tau)} = K_{h_n}(\beta^T X_{(i,t),(j,\tau)}) / \sum_{l \in \mathcal{I}_n} \sum_{r \in \mathcal{T}_n} K_{h_n}(\beta^T X_{(l,r),(j,\tau)})$, $K_{h_n}(\cdot) = h_n^{-1}K(\cdot/h_n)$, h_n is a bandwidth, and $K(\cdot)$ is an univariate symmetric density function. Therefore, the estimator of $\sigma_{\beta}^2(\cdot)$ at $\beta^T X_{j,\tau}$ is just the minimum value of expression (6), namely:

$$\hat{\sigma}_{\beta}^2(X_{j,\tau}) = \min_{a,b} \sum_{i \in \mathcal{I}_n} \sum_{t \in \mathcal{T}_n} [Y_{i,t} - \{a + b\beta^T X_{(i,t),(j,\tau)}\}]^2 w_{(i,t),(j,\tau)}. \tag{7}$$

Under some mild conditions, we have $\hat{\sigma}_{\beta}^2(X_{j,\tau}) - \sigma_{\beta}^2(X_{j,\tau}) = o_p(1)$. On the basis of Expressions (2), (4), and (7), we can estimate β by solving the minimization problem:

$$\min_{\beta: \beta^T \beta = 1} \left\{ \sum_{j \in \mathcal{I}_n} \sum_{\tau \in \mathcal{T}_n} \hat{\sigma}_{\beta}^2(X_{j,\tau}) \right\} = \min_{\substack{\beta: \beta^T \beta = 1 \\ a_{j,\tau}, b_{j,\tau}, j \in \mathcal{I}_n \\ \tau \in \mathcal{T}_n, t \in \mathcal{T}_n}} \left(\sum_{j \in \mathcal{I}_n} \sum_{\tau \in \mathcal{T}_n} [Y_{i,t} - \{a_{j,\tau} + b_{j,\tau} \beta^T X_{(i,t),(j,\tau)}\}]^2 w_{(i,t),(j,\tau)} \right). \tag{8}$$

Let $G(\cdot) = (g(\cdot), g'(\cdot))^T$. The estimation algorithm for β and $G(\cdot)$ can be described as follows.

Step 0. Give the calculation of the initial value β of β_0 .

Step 1. Calculate:

$$\hat{f}_{\beta}(\beta^T X_{j,\tau}) = n^{-1} \sum_{\substack{i \in \mathcal{I}_n \\ t \in \mathcal{T}_n}} K_{h_n}(\beta^T X_{(i,t),(j,\tau)})$$

and:

$$\begin{pmatrix} a_{j,\tau}^{\beta} \\ b_{j,\tau}^{\beta} h_n \end{pmatrix} = \left\{ \sum_{i \in \mathcal{I}_n} \sum_{t \in \mathcal{T}_n} K_{h_n}(\beta^T X_{(i,t),(j,\tau)}) \begin{pmatrix} 1 \\ \beta^T X_{(i,t),(j,\tau)} / h_n \end{pmatrix} \begin{pmatrix} 1 \\ \beta^T X_{(i,t),(j,\tau)} / h_n \end{pmatrix}^T \right\}^{-1} \\ \times \sum_{i \in \mathcal{I}_n} \sum_{t \in \mathcal{T}_n} K_{h_n}(\beta^T X_{(i,t),(j,\tau)}) \begin{pmatrix} 1 \\ \beta^T X_{(i,t),(j,\tau)} / h_n \end{pmatrix} Y_{i,t}.$$

Step 2. Calculate:

$$\beta = \left\{ \sum_{j \in \mathcal{I}_n} \sum_{\tau \in T_n} \sum_{i \in \mathcal{I}_n} \sum_{t \in T_n} K_{h_n}(\beta^T X_{(i,\tau),(j,\tau)}) \hat{\rho}_{j,\tau}^\beta (b_{j,\tau}^\beta)^2 X_{(i,\tau),(j,\tau)} X_{(i,\tau),(j,\tau)}^T / \hat{f}_\beta(\beta^T X_{j,\tau}) \right\}^{-1} \\ \times \sum_{j \in \mathcal{I}_n} \sum_{\tau \in T_n} \sum_{i \in \mathcal{I}_n} \sum_{t \in T_n} K_{h_n}(\beta^T X_{(i,\tau),(j,\tau)}) \hat{\rho}_{j,\tau}^\beta b_{j,\tau}^\beta X_{(i,\tau),(j,\tau)} (Y_{i,t} - \hat{a}_{j,\tau}^\beta) / \hat{f}_\beta(\beta^T X_{j,\tau}),$$

where $\hat{\rho}_{j,\tau}^\beta = \rho_n \left(n^{-1} \{ \sum_{i \in \mathcal{I}_n} \sum_{t \in T_n} K_{h_n}(\beta^T X_{(i,\tau),(j,\tau)}) \} \right)$, and $\rho_n(\cdot)$ is a trimming function.

Step 3. Repeat steps 1–3 with $\beta := \beta / \|\beta\|$, where $\|\cdot\|$ denotes the Euclidean distance, until convergence. The vector obtained in the last iteration is defined as the rMAVE estimator of β , denoted by $\hat{\beta}_n$.

Step 4. Put $\hat{\beta}_n$ into step 1 and obtain the estimators of $g(\cdot)$ and $g'(\cdot)$, denoted by $\hat{G}_n(\cdot) = (\hat{g}_n(\cdot), \hat{g}'_n(\cdot))^T$.

Combining $\hat{\beta}_n, \hat{g}_n(\cdot)$ and model (1) leads to the residuals

$$\hat{\varepsilon}_{i,t} = Y_{i,t} - \hat{g}_n(\hat{\beta}_n^T X_{i,t}), \quad i \in \mathcal{I}_n, t \in T_n. \tag{9}$$

Remark 1. The calculation of the initial value $\hat{\beta}_{(1)}$ can refer to [10,13,22]. To deal with the large deviation of the boundary points, a suitable trimming function is introduced; see [11], which is of the form:

$$\rho_n(v) = \begin{cases} 1, & v \geq 2c_0n^{-\epsilon}, \\ \frac{\exp\{(2c_0n^{-\epsilon}-v)^{-1}\}}{\exp\{(2c_0n^{-\epsilon}-v)^{-1}\} + \exp\{(v-c_0n^{-\epsilon})^{-1}\}}, & 2c_0n^{-\epsilon} > v > c_0n^{-\epsilon}, \\ 0, & v \leq c_0n^{-\epsilon}, \end{cases}$$

where $0 < \epsilon < \frac{1}{20}$ and $0 < c_0 < \frac{1}{20}$.

3. Estimation of the Variance Function

In this section, we focus on estimating the variance function in single-index heteroscedastic models. The estimation methods and the convergence properties of their estimations are given respectively.

3.1. Estimation of the Variance Function with Fully Nonparametric Function

For any given point X_{s_0,t_0} and the estimation $\hat{\beta}_n$, an estimation of $E(\varepsilon^2|X)$ at X_{s_0,t_0} is:

$$\hat{\sigma}_{\hat{\beta}_n}^2(X_{s_0,t_0}) = \frac{\sum_{j \in \mathcal{I}_n} \sum_{\tau \in T_n} \hat{\varepsilon}_{j,\tau}^2 L_{l_n}(X_{(j,\tau)(s_0,t_0)})}{\sum_{j \in \mathcal{I}_n} \sum_{\tau \in T_n} L_{l_n}(X_{(j,\tau)(s_0,t_0)}), \tag{10}$$

where $\hat{\varepsilon}_{j,\tau}$ are the residuals calculated by (9), $L_{l_n}(\cdot) = l_n^{-1}L(\cdot/l_n)$, $L(\cdot)$ is a d -dimensional symmetric density function such that $\int L(u)du = 1$ and $\int uL(u)du = 0$, and l_n is a bandwidth. Taking $X_{s_0,t_0} = X_{i,t}$ ($i \in \mathcal{I}_n, t \in T_n$), the estimations at all the designed points $X_{i,t}$ can be obtained, i.e., $\hat{\sigma}_{\hat{\beta}_n}^2(X_{i,t})$ ($i \in \mathcal{I}_n, t \in T_n$).

The following theorems are about the asymptotic result for the estimator $\hat{\sigma}_{\hat{\beta}_n}^2(X_{i,t})$.

Theorem 1. According to the assumptions (C1)–(C4), we have:

$$\sup_{X_{s_0,t_0} \in R^d} \left| \hat{\sigma}_{\hat{\beta}_n}^2(X_{s_0,t_0}) - \sigma^2(X_{s_0,t_0}) \right| = o_P \left\{ l_n^{2d} + \left(\frac{\log(1/l_n^d)}{nl_n^d} \right)^{\frac{1}{2}} \right\}.$$

Theorem 1 shows that the estimator $\hat{\sigma}_{\hat{\beta}_n}^2(X_{s_0,t_0})$ is consistent, and the proof is found in the Appendix B.

3.2. Estimation of the Variance Function with Dimension Reduction Structure

If the dimension of the covariates is large, the core estimates given in the previous section suffer from the curse of dimensionality and, furthermore, the proposed test statistic may not be very informative. If the model contains a dimensionality reduction structure, we should use it to make the test more powerful. Let us assume the following model structure:

$$Y \vdash X | \beta^T X. \tag{11}$$

Formula (11) notes that, given $\beta^T X$, Y and X are conditionally independent. According to [18–20], this is a general dimension reduction structure that includes the model $Y = G(\beta^T X, \varepsilon)$ as its special case.

If the above structure holds for model (1), the mean and variance have the same index β . For any given point X_{s_0,t_0} and the estimated $\hat{\beta}_n$, the variance function $E(\varepsilon^2 | \beta^T X)$ at X_{s_0,t_0} is:

$$\hat{\sigma}_{\hat{\beta}_n}^2(X_{s_0,t_0}) = \hat{\sigma}^2(\hat{\beta}_n^T X_{s_0,t_0}) = \frac{\sum_{j \in \mathcal{I}_n} \sum_{\tau \in T_n} \hat{\varepsilon}_j^2 Q_{b_n}(\hat{\beta}_n^T X_{(j,\tau)(s_0,t_0)})}{\sum_{j \in \mathcal{I}_n} \sum_{\tau \in T_n} Q_{b_n}(\hat{\beta}_n^T X_{(j,\tau)(s_0,t_0)}), \tag{12}$$

where $\hat{\sigma}^2(\hat{\beta}_n^T X_{s_0,t_0})$ can also be denoted by $\hat{\sigma}_{\hat{\beta}_n}^2(X_{s_0,t_0})$, $\hat{\varepsilon}_{j,\tau}$ ($j \in \mathcal{I}_n, \tau \in T_n$) are the residuals of the rMAVE, $Q_{b_n}(\cdot) = b_n^{-1}Q(\cdot/b_n)$, $Q(\cdot)$ is a univariate kernel function, and b_n is a bandwidth. Particularly, taking $X_{s_0,t_0} = X_{i,t}$, ($i \in \mathcal{I}_n, t \in T_n$), we can obtain the estimations of the variance function at $X_{i,t}$, i.e., $\hat{\sigma}^2(\hat{\beta}_n^T X_{i,t})$.

For the estimate $\hat{\sigma}^2(\hat{\beta}_n^T X_{s_0,t_0})$ of the variance function, the consistency can be proven.

Theorem 2. According to the assumptions (C1)–(C3) and (C5), we have:

$$\sup_{X_{s_0,t_0} \in R^d} \left| \hat{\sigma}^2(\hat{\beta}_n^T X_{s_0,t_0}) - \sigma^2(\beta_0^T X_{s_0,t_0}) \right| = o_P \left\{ b_n^2 + \left(\frac{\log(1/b_n)}{nb_n} \right)^{\frac{1}{2}} \right\}.$$

The proof of Theorem 2 is given in the Appendix B.

4. Reweighting Estimation and Asymptotic Properties

In this section, we provide the estimation method for the spatio-temporal single-index model and provide the asymptotic properties of the proposed estimator.

4.1. Reweighting Estimation

If the model exhibits heteroscedasticity, the usual approach can be taken by considering weighted estimates. In particular, we describe a reweighting procedure of the single-index model with heteroscedastic errors with the estimated values of the variance function at all designed points. For model (1), weighted versions of the proposed estimation methods following the idea of [23] can be considered. When calculating, we only need to modify the algorithms in Step 2, by replacing $\hat{\rho}_{j,\tau}^\beta = \rho_n \left(n^{-1} \sum_{j \in \mathcal{I}_n} \sum_{\tau \in T_n} K_{h_n}(\beta^T X_{(i,\tau),(j,\tau)}) \right)$ with

$$\hat{\rho}_{j,\tau}^\beta = \rho_n \left(n^{-1} \sum_{j \in \mathcal{I}_n} \sum_{\tau \in T_n} K_{h_n}(\beta^T X_{(i,\tau),(j,\tau)}) \right) / \hat{\sigma}_{\hat{\beta}_n}^2(X_{j,\tau}),$$

where $\hat{\sigma}_{\hat{\beta}_n}^2(X_{j,\tau})$ is defined in (10) or (12). In terms of the modified algorithms, the reweighted estimators for the parameter vector β and the link function $g(\cdot)$ can be obtained, denoted by $\hat{\beta}_R$ and $\hat{g}_R(\cdot)$, respectively.

4.2. Asymptotic Properties

In this subsection, under the assumption of the α mixing condition, we provide the asymptotic distributions of two type estimators of β and show that the reweighted estimator $\hat{\beta}_R$ has no greater asymptotic variance than the rMAVE estimator $\hat{\beta}_n$. Then, the asymptotic distributions of $\hat{g}_n(\cdot)$ and $\hat{g}_R(\cdot)$ are shown.

There are some notations to determine asymptotic results. Let $\mu_\beta(x) = E(X|\beta^T X = \beta^T x)$, $v_\beta(x) = \mu_\beta(x) - x$, $w_\beta(x) = E(XX^T|\beta^T X = \beta^T x)$, and $W_0(x) = v_{\beta_0}(x)v_{\beta_0}^T(x)$.

Theorem 3. Assuming $n \rightarrow \infty$, according to the assumptions (C1)–(C3), we have:

$$\sqrt{n}(\hat{\beta}_n - \beta_0) \xrightarrow{D} N(0, W_{g_0}^+ \Delta W_{g_0}^+),$$

where $W_{g_0} = E\{g'(\beta_0^T X)^2 W_0(X)\}$ and $\Delta = E\{g'(\beta_0^T X)^2 W_0(X) \sigma_{\beta_0}^2(X)\}$.

Theorem 4. Assuming $n \rightarrow \infty$, according to the assumptions (C1)–(C3) and (C4) or (C5), we have:

$$\sqrt{n}(\hat{\beta}_R - \beta_0) \xrightarrow{D} N(0, \tilde{W}_{g_0}^+),$$

where $\tilde{W}_{g_0} = E\{g'(\beta_0^T X)^2 W_0(X) \sigma_{\beta_0}^{-2}(X)\}$.

Theorem 5. In addition to the assumptions (C1)–(C3) and (C4) or (C5), if ε is independent of X , then we have,

$$\tilde{W}_{g_0}^+ \leq W_{g_0}^+ \Delta W_{g_0}^+.$$

Remark 2. From the result of Theorem 5, a conclusion can be obtained that $\hat{\beta}_R$ is asymptotically more efficient than $\hat{\beta}_n$ in terms of asymptotic variance.

For the rMAVE estimator $\hat{g}_n(\cdot)$ and reweighted estimator $\hat{g}_R(\cdot)$ of the link function $g(\cdot)$ at $X = X_{s_0, t_0}$, their asymptotic distributions are also derived. These are the following theorems.

Theorem 6. According to the assumptions (C1)–(C3), for any $v = \beta_0^T x$, as $n \rightarrow \infty$,

$$\sqrt{nh_n} \left\{ \hat{g}_n(v) - g(v) - \frac{h_n^2}{2} g''(v) \right\} \xrightarrow{D} N \left(0, \frac{\sigma_{\beta_0}^2(x)}{f_{\beta_0}(v)} \right).$$

Theorem 7. According to the assumptions (C1)–(C3), for any $v = \beta_0^T x$, as $n \rightarrow \infty$,

$$\sqrt{nh_n} \left\{ \hat{g}_R(v) - g(v) - \frac{h_n^2}{2} g''(v) \right\} \xrightarrow{D} N \left(0, \frac{\sigma_{\beta_0}^2(x)}{f_{\beta_0}(v)} \right).$$

Remark 3. By comparing Theorems 6 and 7, the asymptotic distributions obtained by rMAVE and reweighting methods are the same, which reflects the characteristic of local regression in nonparametric models.

5. Monte Carlo Study

We use the spectral method to simulate the spatio-temporal process:

$$X_{i,j,t} = (2/\Omega)^{1/2} \sum_{k=1}^{\Omega} \cos(w(1,k) * i + w(2,k) * j + q(k) * t + r(k)),$$

where $w(i, k), i = 1, 2, q(k), k = 1, \dots, \Omega$ are i.i.d. from a standard normal distribution, independent of $r(k), k = 1, \dots, \Omega$, i.i.d. uniform random variables on $[-\pi, \pi]$. As $n \rightarrow \infty, X_{i,j,t}$ converges to a Gaussian ergodic process (see [24]). Additionally, the $\varepsilon_{i,j,t}, i = 1, \dots, n_1, j = 1, \dots, n_2, t = 1, \dots, n_3$ are i.i.d. from a standard normal distribution. We conducted 100 simulation studies with $\Omega = 1000$ and a sample size of $8 \times 8 \times 8 (n_1 = 8, n_2 = 8, n_3 = 8), 10 \times 10 \times 10$. Comparisons were made between the rMAVE estimator and two reweighted estimators. For the convenience of expression, we use RWEFN and RWEDR respectively to represent reweighting estimation of a fully nonparametric function and reweighting estimation by the dimensionality reduction structure.

For simplicity, the rule of thumb [25] is used to select the bandwidth for the rMAVE and reweighted methods. To verify the performance of the proposed methods, we design the following two examples. We take the Gaussian kernel functions $K(u) = Q(u) = 1/(\sqrt{2\pi})e^{-\frac{u^2}{2}}$ and $L(u_1, \dots, u_p) = l(u_1) \times \dots \times l(u_p)$ with $l(u) = 1/(\sqrt{2\pi})e^{-\frac{u^2}{2}}$. The trimming function with $\varepsilon = \frac{1}{10}$ and $c_0 = 0.01$ is considered. In the example, we also take $\beta_0 = (1, 2, 3, 0)^T / \sqrt{14}$, and the samples $X_{i,j,t} = (X_{i,j,t(1)}, X_{i,j,t(2)}, X_{i,j,t(3)}, X_{i,j,t(4)})$ are simulated by the spectral method on the 4-dimensional cube $[-1, 1]^4$.

Example 1. We consider the following model:

$$Y_{i,t} = 2 + 2(\beta_0^T X_{i,t}) + (\theta_0 |\beta_0^T X_{i,t}| + 0.5)\varepsilon_{i,t}.$$

Example 2. We consider the following example:

$$Y_{i,t} = 1 + 2(\beta_0^T X_{i,t} + (\theta_0 |X_{i,t}(1) + X_{i,t}(2)| + 1)\varepsilon).$$

The simulated results of Example 1 are reported in Tables 1 and 2, and the results of Example 2 are shown in Tables 3 and 4.

We assume the value of $\theta_0 = 1, 1.5, 2$ to evaluate the influence of error heteroscedasticity on coefficient estimates. To show the performance of the estimators $\hat{\beta}_n, \hat{\beta}_R^{FN}$ (RWEFN), and $\hat{\beta}_R^{DR}$ (RWEDR), two indices are defined: the sampling standard deviation (SSD) and the relative sampling efficiency (SRE), respectively. In particular, for rMAVE, RWEFN, or RWEDR, let $\hat{\beta}_{j(1)}, \dots, \hat{\beta}_{j(M)}$ be estimates of β_j in M replicates. The SSD for β_j is defined by:

$$SSD(\beta_j) = \left[\frac{1}{M} \sum_{k=1}^M (\hat{\beta}_{j(k)} - \bar{\beta}_j)^2 \right]^{\frac{1}{2}},$$

and the SRE for β_j is defined by:

$$SRE(\beta_j) = \frac{\frac{1}{M} \sum_{k=1}^M (\hat{\beta}_{j(k)} - \beta_j)^2}{\frac{1}{M} \sum_{k=1}^M (\tilde{\beta}_{j(k)} - \beta_j)^2},$$

where $\bar{\beta}_j = \frac{1}{M} \sum_{k=1}^M \hat{\beta}_{j(k)}$, and $\tilde{\beta}$ has the same form as that of $\hat{\beta}_R^{FN}$ or $\hat{\beta}_R^{DR}$, except that the weighting uses the true variance function. The calculation in different sample size results of $\hat{\beta}_n, \hat{\beta}_R^{FN}$ and $\hat{\beta}_R^{DR}$, in terms of the sample mean, sample standard deviation, and relative sampling efficiency, are shown in Tables 1 and 2.

Table 1. Sample means, standard deviations, and relative efficiencies in Example 1 when $n = 8 \times 8 \times 8$.

Item	Estimator	$\theta_0 = 1.0$	$\theta_0 = 1.5$	$\theta_0 = 2.0$
Mean	$\hat{\beta}_{1,n}$	0.2793	0.2745	0.2615
	$\hat{\beta}_{1,R}^{FN}$	0.2761	0.2772	0.2578
	$\hat{\beta}_{1,R}^{DR}$	0.2779	0.2793	0.2600
	$\hat{\beta}_{2,n}$	0.5487	0.5582	0.5291
	$\hat{\beta}_{2,R}^{FN}$	0.5358	0.5390	0.5289
	$\hat{\beta}_{2,R}^{DR}$	0.5420	0.5438	0.5277
	$\hat{\beta}_{3,n}$	0.7858	0.7784	0.7933
	$\hat{\beta}_{3,R}^{FN}$	0.7963	0.7941	0.7947
	$\hat{\beta}_{3,R}^{DR}$	0.7916	0.7892	0.7937
	$\hat{\beta}_{4,n}$	−0.007	−0.0070	0.0015
	$\hat{\beta}_{4,R}^{FN}$	−0.0004	−0.0055	0.0009
	$\hat{\beta}_{4,R}^{DR}$	−0.0031	−0.0036	0.0017
SSD	$\hat{\beta}_{1,n}$	0.0279	0.0362	0.2145
	$\hat{\beta}_{1,R}^{FN}$	0.0318	0.0334	0.1653
	$\hat{\beta}_{1,R}^{DR}$	0.0349	0.0421	0.1597
	$\hat{\beta}_{2,n}$	0.0430	0.0626	0.2183
	$\hat{\beta}_{2,R}^{FN}$	0.0235	0.0319	0.1642
	$\hat{\beta}_{2,R}^{DR}$	0.0228	0.0283	0.1583
	$\hat{\beta}_{3,n}$	0.0255	0.0386	0.2253
	$\hat{\beta}_{3,R}^{FN}$	0.0125	0.0178	0.1633
	$\hat{\beta}_{3,R}^{DR}$	0.0151	0.0190	0.1558
	$\hat{\beta}_{4,n}$	0.0144	0.0189	0.2306
	$\hat{\beta}_{4,R}^{FN}$	0.0063	0.0113	0.1668
	$\hat{\beta}_{4,R}^{DR}$	0.0064	0.0081	0.1572
SRE	$\hat{\beta}_{1,n}$	0.6758	0.6974	2.5971
	$\hat{\beta}_{1,R}^{FN}$	0.7974	0.6210	1.4813
	$\hat{\beta}_{1,R}^{DR}$	0.9713	0.9779	1.3349
	$\hat{\beta}_{2,n}$	4.1839	6.5845	2.6142
	$\hat{\beta}_{2,R}^{FN}$	1.1334	1.5277	1.4925
	$\hat{\beta}_{2,R}^{DR}$	1.1822	1.3079	1.3617
	$\hat{\beta}_{3,n}$	3.0887	4.7931	2.7372
	$\hat{\beta}_{3,R}^{FN}$	0.6304	0.8887	1.6809
	$\hat{\beta}_{3,R}^{DR}$	1.1264	1.2255	1.6126
	$\hat{\beta}_{4,n}$	4.0594	4.0605	2.6925
	$\hat{\beta}_{4,R}^{FN}$	0.6396	1.5820	1.6784
	$\hat{\beta}_{4,R}^{DR}$	0.8161	0.8023	1.6117

Table 2. Sample means, standard deviations, and relative efficiencies in Example 1 when $n = 10 \times 10 \times 10$.

Item	Estimator	$\theta_0 = 1.0$	$\theta_0 = 1.5$	$\theta_0 = 2.0$
Mean	$\hat{\beta}_{1,n}$	0.2602	0.2558	0.2642
	$\hat{\beta}_{1,R}^{FN}$	0.2563	0.2587	0.2577
	$\hat{\beta}_{1,R}^{DR}$	0.2516	0.2548	0.2644
	$\hat{\beta}_{2,n}$	0.5345	0.5501	0.5096
	$\hat{\beta}_{2,R}^{FN}$	0.5319	0.5423	0.5038
	$\hat{\beta}_{2,R}^{DR}$	0.5391	0.5427	0.5312
	$\hat{\beta}_{3,n}$	0.8029	0.7912	0.8169
	$\hat{\beta}_{3,R}^{FN}$	0.8052	0.7998	0.8227
	$\hat{\beta}_{3,R}^{DR}$	0.8032	0.7993	0.8044
	$\hat{\beta}_{4,n}$	−0.0113	0.0493	−0.0018
	$\hat{\beta}_{4,R}^{FN}$	−0.0059	0.0161	−0.0094
	$\hat{\beta}_{4,R}^{DR}$	−0.0039	0.0229	−0.0028
SSD	$\hat{\beta}_{1,n}$	0.0306	0.0446	0.0256
	$\hat{\beta}_{1,R}^{FN}$	0.0238	0.0186	0.0346
	$\hat{\beta}_{1,R}^{DR}$	0.0185	0.0125	0.0138
	$\hat{\beta}_{2,n}$	0.0260	0.0179	0.0323
	$\hat{\beta}_{2,R}^{FN}$	0.0142	0.0133	0.0278
	$\hat{\beta}_{2,R}^{DR}$	0.0148	0.0141	0.0198
	$\hat{\beta}_{3,n}$	0.0158	0.0214	0.0226
	$\hat{\beta}_{3,R}^{FN}$	0.0041	0.0090	0.0257
	$\hat{\beta}_{3,R}^{DR}$	0.0088	0.0120	0.0135
	$\hat{\beta}_{4,n}$	0.0210	0.0305	0.0219
	$\hat{\beta}_{4,R}^{FN}$	0.0207	0.0252	0.0214
	$\hat{\beta}_{4,R}^{DR}$	0.0224	0.0212	0.0203
SRE	$\hat{\beta}_{1,n}$	1.6501	9.6640	3.6479
	$\hat{\beta}_{1,R}^{FN}$	1.1514	1.9198	7.0692
	$\hat{\beta}_{1,R}^{DR}$	0.9846	1.4218	1.0852
	$\hat{\beta}_{2,n}$	3.3675	2.855	4.9831
	$\hat{\beta}_{2,R}^{FN}$	1.0456	1.2121	5.1506
	$\hat{\beta}_{2,R}^{DR}$	1.1884	1.3611	1.2101
	$\hat{\beta}_{3,n}$	3.9555	5.358	4.8345
	$\hat{\beta}_{3,R}^{FN}$	0.4494	0.7964	7.1304
	$\hat{\beta}_{3,R}^{DR}$	1.2435	1.4159	1.2323
	$\hat{\beta}_{4,n}$	1.0402	4.8466	1.1845
	$\hat{\beta}_{4,R}^{FN}$	0.8493	1.2902	1.3420
	$\hat{\beta}_{4,R}^{DR}$	0.9447	1.4113	1.0247

Table 3. Sample means, standard deviations, and relative efficiencies in Example 2 when $n = 8 \times 8 \times 8$.

Item	Estimator	$\theta_0 = 1.0$	$\theta_0 = 1.5$	$\theta_0 = 2.0$
Mean	$\hat{\beta}_{1,n}$	0.3180	0.3203	0.3091
	$\hat{\beta}_{1,R}^{FN}$	0.3026	0.3060	0.3146
	$\hat{\beta}_{1,R}^{DR}$	0.3013	0.2936	0.3070
	$\hat{\beta}_{2,n}$	0.5422	0.5131	0.5647
	$\hat{\beta}_{2,R}^{FN}$	0.5417	0.5147	0.5581
	$\hat{\beta}_{2,R}^{DR}$	0.5412	0.5152	0.5587
	$\hat{\beta}_{3,n}$	0.7760	0.7945	0.7602
	$\hat{\beta}_{3,R}^{FN}$	0.7824	0.7973	0.7681
	$\hat{\beta}_{3,R}^{DR}$	0.7842	0.8022	0.7635
	$\hat{\beta}_{4,n}$	0.0114	−0.0101	0.0185
	$\hat{\beta}_{4,R}^{FN}$	0.0083	0.0015	0.0301
	$\hat{\beta}_{4,R}^{DR}$	0.0076	0.0014	0.0045
SSD	$\hat{\beta}_{1,n}$	0.0239	0.0599	0.1064
	$\hat{\beta}_{1,R}^{FN}$	0.0228	0.0576	0.1007
	$\hat{\beta}_{1,R}^{DR}$	0.0205	0.0317	0.0672
	$\hat{\beta}_{2,n}$	0.0127	0.0278	0.0515
	$\hat{\beta}_{2,R}^{FN}$	0.0095	0.0235	0.0421
	$\hat{\beta}_{2,R}^{DR}$	0.0082	0.0202	0.0115
	$\hat{\beta}_{3,n}$	0.0081	0.0373	0.0762
	$\hat{\beta}_{3,R}^{FN}$	0.0098	0.0350	0.0632
	$\hat{\beta}_{3,R}^{DR}$	0.0069	0.0308	0.0360
	$\hat{\beta}_{4,n}$	0.0509	0.0200	0.0337
	$\hat{\beta}_{4,R}^{FN}$	0.0357	0.0133	0.0337
	$\hat{\beta}_{4,R}^{DR}$	0.0344	0.0055	0.0089
SRE	$\hat{\beta}_{1,n}$	1.8127	0.8880	0.7741
	$\hat{\beta}_{1,R}^{FN}$	1.0206	1.1188	1.2239
	$\hat{\beta}_{1,R}^{DR}$	0.9112	0.9941	0.9501
	$\hat{\beta}_{2,n}$	0.6660	1.5348	0.8369
	$\hat{\beta}_{2,R}^{FN}$	1.2903	1.1565	1.0343
	$\hat{\beta}_{2,R}^{DR}$	0.9032	0.9960	0.9485
	$\hat{\beta}_{3,n}$	1.7353	0.8257	1.5072
	$\hat{\beta}_{3,R}^{FN}$	1.0690	1.1678	1.3741
	$\hat{\beta}_{3,R}^{DR}$	0.9846	1.0108	0.9238
	$\hat{\beta}_{4,n}$	1.8545	4.6052	3.8478
	$\hat{\beta}_{4,R}^{FN}$	0.8656	4.2492	1.9594
	$\hat{\beta}_{4,R}^{DR}$	0.9830	0.9644	1.3002

Table 4. Sample means, standard deviations, and relative efficiencies in Example 2 when $n = 10 \times 10 \times 10$.

Item	Estimator	$\theta_0 = 1.0$	$\theta_0 = 1.5$	$\theta_0 = 2.0$
Mean	$\hat{\beta}_{1,n}$	0.3033	0.3071	0.3112
	$\hat{\beta}_{1,R}^{FN}$	0.3022	0.3011	0.3018
	$\hat{\beta}_{1,R}^{DR}$	0.2986	0.2960	0.2952
	$\hat{\beta}_{2,n}$	0.5532	0.5267	0.5474
	$\hat{\beta}_{2,R}^{FN}$	0.5492	0.5288	0.5448
	$\hat{\beta}_{2,R}^{DR}$	0.5449	0.5298	0.5380
	$\hat{\beta}_{3,n}$	0.7700	0.8020	0.7742
	$\hat{\beta}_{3,R}^{FN}$	0.7764	0.8017	0.7799
	$\hat{\beta}_{3,R}^{DR}$	0.7781	0.8016	0.7810
	$\hat{\beta}_{4,n}$	0.0552	−0.0040	−0.012
	$\hat{\beta}_{4,R}^{FN}$	0.0468	−0.0033	−0.0089
	$\hat{\beta}_{4,R}^{DR}$	0.0463	−0.0031	−0.0063
SSD	$\hat{\beta}_{1,n}$	0.0296	0.0154	0.0150
	$\hat{\beta}_{1,R}^{FN}$	0.0238	0.0146	0.0117
	$\hat{\beta}_{1,R}^{DR}$	0.0225	0.0100	0.0098
	$\hat{\beta}_{2,n}$	0.0209	0.0162	0.0132
	$\hat{\beta}_{2,R}^{FN}$	0.0188	0.0125	0.0098
	$\hat{\beta}_{2,R}^{DR}$	0.0146	0.0102	0.0026
	$\hat{\beta}_{3,n}$	0.0088	0.0256	0.0087
	$\hat{\beta}_{3,R}^{FN}$	0.0073	0.0230	0.0033
	$\hat{\beta}_{3,R}^{DR}$	0.0023	0.0160	0.0010
	$\hat{\beta}_{4,n}$	0.0582	0.0062	0.0920
	$\hat{\beta}_{4,R}^{FN}$	0.0566	0.0017	0.0832
	$\hat{\beta}_{4,R}^{DR}$	0.0508	0.0009	0.0708
SRE	$\hat{\beta}_{1,n}$	0.9646	1.5420	1.7484
	$\hat{\beta}_{1,R}^{FN}$	0.9738	0.9465	0.8725
	$\hat{\beta}_{1,R}^{DR}$	0.9994	1.0162	1.1460
	$\hat{\beta}_{2,n}$	1.1077	1.8686	1.9866
	$\hat{\beta}_{2,R}^{FN}$	1.0915	1.4366	0.9473
	$\hat{\beta}_{2,R}^{DR}$	1.0115	0.7647	0.9885
	$\hat{\beta}_{3,n}$	1.3524	1.3291	1.5706
	$\hat{\beta}_{3,R}^{FN}$	0.8750	1.2429	0.8803
	$\hat{\beta}_{3,R}^{DR}$	0.9261	1.0108	1.1370
	$\hat{\beta}_{4,n}$	1.0374	1.4991	0.7267
	$\hat{\beta}_{4,R}^{FN}$	1.0193	0.8973	0.9798
	$\hat{\beta}_{4,R}^{DR}$	0.9941	0.9644	1.1907

The following conclusions can be obtained from Tables 1–4. First, in all cases, the values of the sample mean are reasonably close to their true values, suggesting that $\hat{\beta}_n$, $\hat{\beta}_R^{FN}$, and $\hat{\beta}_R^{DR}$ are asymptotically unbiased. Second, for $\theta_0 \neq 0$, the values of SSD and SRE for the RWEFN and RWEDR methods are smaller than those for the rMAVE method. Third, from tables, the values of $\hat{\beta}_R^{DR}$ slightly outperform $\hat{\beta}_R^{FN}$ when error heteroscedasticity exists. This is also explainable because the example has a dimensionality reduction structure. Fourth, $\hat{\beta}_R^{FN}$ and $\hat{\beta}_R^{DR}$ have better performance when the error heteroscedasticity is larger, which implies that the improvement in $\hat{\beta}_R^{FN}$ and $\hat{\beta}_R^{DR}$ over $\hat{\beta}_n$ becomes obvious with the growth of errors. Overall, $\hat{\beta}_R^{FN}$ and $\hat{\beta}_R^{DR}$ work much better than $\hat{\beta}_n$, and $\hat{\beta}_R^{DR}$ is the best. Finally, because the composition of Example 2 is more complex, the average value of the obtained estimators is worse than that of Example 1.

6. Real Data Analysis

Based on the reweighting method proposed in this article, the air pollution data of Nanjing were studied, with the data coming from the China Meteorological Data Network and the real air data from the Environment Big Data Center. The data contain the air quality index (AQI) $PM_{2.5}$ ($\mu g/m^3$), PM_{10} ($\mu g/m^3$), CO (mg/m^3), SO_2 ($\mu g/m^3$), NO_2 ($\mu g/m^3$), and ozone eight-hour O_3_{8h} ($\mu g/m^3$) of Nanjing from 23 October 2020 to 23 October 2022, including the AQI in response to a variable Y and the rest of the variables as covariate X_1, X_2, \dots, X_6 .

When the data were estimated using the rMAVE method, the value of the estimated parameter was $\hat{\beta}_n = (0.6606, -0.1529, 0.8679, 11.1681, 0.3236, 0.4541)^T$. The results obtained by reweighting the data are as follows: first, $\hat{\beta}_R^{DR} = (0.04167, -0.0102, 0.0637, 0.9964, 0.0202, 0.02915)^T$ was obtained by adopting a dimensional-reduction structure. The other difference is estimated by taking the kernel regression, resulting in $\hat{\beta}_R^{FN} = (0.04238, -0.0108, 0.0675, 0.9961, 0.0200, 0.0293)^T$. The fitting result is shown in Figure 1, which contains the true values as well as the values fitted by the three methods, where $R^2_{Rmave} = 0.7586$, $R^2_{DR} = 0.7606$, $R^2_{FN} = 0.7606$. Due to the four lines being too close to distinguish, the right side of Figure 1 is the enlarged part. The estimators obtained by the reweighting method have a better fitting effect.

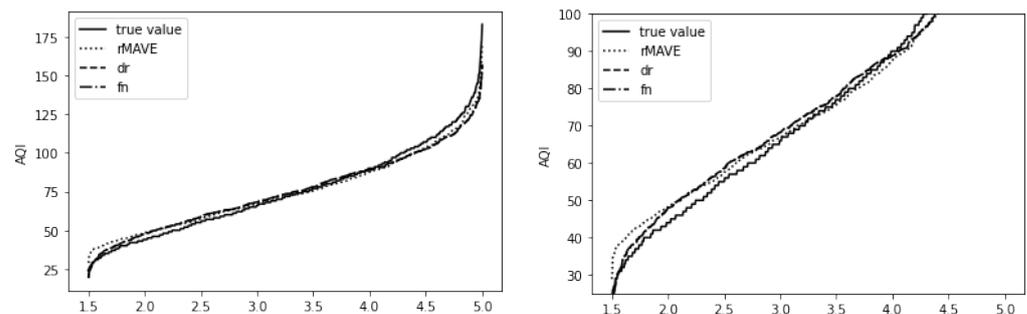


Figure 1. Air pollution data fitting.

7. Conclusions

This work considers an estimation problem in single-index models with spatio-temporal correlation and heterogeneity. We propose a parametric component reweighting estimated method based on the variance function of the error. Theoretical results show that the proposed reweighting estimators have smaller asymptotic variances while maintaining the same biases. Numerical simulations show that the estimators revealed that heterogeneity is closer to the true value. A real data analysis was conducted to illustrate the proposed methods.

Author Contributions: Methodology, H.H.; Formal analysis, Z.Z.; Writing—original draft, H.W.; Writing—review & editing, C.H. All authors have read and agreed to the published version of the manuscript.

Funding: This research was supported by the National Social Science Fund of China under Grant No. 22BTJ021.

Data Availability Statement: The data presented in this study are available on request from the corresponding author.

Conflicts of Interest: The authors declare no conflict of interest.

Appendix A. Assumptions

In order to obtain asymptotic results, we will assume throughout the paper that at any fixed moment t' , $\{\varepsilon_{i,t'}, i \in \mathcal{I}_n\}$ satisfies the following mixing condition: there exists a function $\varphi_1(t) \searrow 0$ as $t \rightarrow \infty$, $\varphi_1(0) = 1$, such that whenever $E, E' \subset \mathcal{I}_n$ with finite cardinals

$$\begin{aligned} \alpha(\beta(E), \beta(E')) &= \sup\{P(AB) - P(A)P(B) | A \in \beta(E), B \in \beta(E')\} \\ &\leq \psi_1(\text{card}(E), \text{card}(E'))\varphi_1(d(E, E')), \end{aligned} \tag{A1}$$

where $\beta(E)$ (resp, $\beta(E')$) denotes the Borel σ -field generated by $\{\varepsilon_{i,t'}\}_{i \in E}$ (resp, $\{\varepsilon_{i,t'}\}_{i \in E'}$), $\text{Card } E$ (resp, $\text{Card } E'$) is the cardinality of E (resp, E') and $d(E, E')$ is the ordinary Euclidean distance. $\psi_1 : \mathbb{N}^2 \rightarrow \mathbb{R}^+$ is a symmetric positive function that is nondecreasing in each variable.

Similarly, we assume that at any fixed location i_0 , $\{\varepsilon_{i_0,t}, t \in T_n\}$ satisfies the following mixing condition: there exists a function $\varphi_2(t) \searrow 0$ as $t \rightarrow \infty$, $\varphi_2(0) = 1$, such that whenever $G, G' \subset T_n$ with finite cardinals

$$\begin{aligned} \alpha(\beta(G), \beta(G')) &= \sup\{P(AB) - P(A)P(B) | A \in \beta(G), B \in \beta(G')\} \\ &\leq \psi_2(\text{card}(G), \text{card}(G'))\varphi_2(d(G, G')), \end{aligned} \tag{A2}$$

where $\beta(G)$ (resp, $\beta(G')$) denote the Borel σ -field generated by $\{\varepsilon_{i_0,t}\}_{t \in G}$ (resp, $\{\varepsilon_{i_0,t}\}_{t \in G'}$), $\psi_2 : \mathbb{N}^2 \rightarrow \mathbb{R}^+$ is a symmetric positive function that is nondecreasing in each variable. Furthermore, the random field $\{(X_{i,t}, Y_{i,t}), i \in \mathcal{I}_n, t \in T_n\}$ also satisfies the above assumption.

The following conditions are imposed to obtain the asymptotic properties of the resulting estimators. They are not the weakest, but they are introduced to make the evidence easier.

- (C1) The density function $f_\beta(v)$ of $\beta^T X$ and its derivatives up to the third order are bounded on R for all $\beta: \|\beta - \beta_0\| < \delta$ where $\delta > 0$ is a constant, $E\|X\|^6 < \infty$, and $E|Y|^3 < \infty$.
- (C2) The conditional mean $g_\beta(v) = E(Y | \beta^T X = v)$ and its derivatives up to the third order are bounded for all $\beta: \|\beta - \beta_0\| < \delta$ where $\delta > 0$.
- (C3) $K(\cdot)$ is a symmetric univariate density function with finite moments of all orders and a bounded derivative. Bandwidth $h_n \propto n^{-1/5}$ and $\lim_{n \rightarrow \infty} \log(1/h_n)/nh_n^2 = 0$.
- (C4) $L(\cdot)$ is a symmetric multivariate density function with finite moments and bounded derivatives. Bandwidth $l_n \propto n^{-d/5}$ and $\lim_{n \rightarrow \infty} \log(1/l_n^d)/nl_n^{2d} = 0$. $\exists s > 2, \tilde{\delta} < 1 - s^{-1}$, such that $\lim_{n \rightarrow \infty} n^{2\tilde{\delta}-1}l_n = +\infty$.
- (C5) $Q(\cdot)$ is a symmetric univariate density function with finite moments and bounded derivatives. Bandwidth $b_n \propto n^{-1/5}$ and $\lim_{n \rightarrow \infty} \log(1/b_n)/nb_n^2 = 0$. $\exists s > 2, \tilde{\rho} < 1 - s^{-1}$, such that $\lim_{n \rightarrow \infty} n^{2\tilde{\rho}-1}b_n = +\infty$.

According to (C1), covariate X can have discrete components provided that $\beta^T X$ is continuous for β in a small neighbor of β_0 ; see also [23]. In order to be able to use the optimal bandwidth, there needs to be a moment of order higher than the second order moment for the variable Y . The smoothness requirement of the link function in hypothesis (C2) can be relaxed to the existence of a bounded second-order derivative but requires

a more complex proof process and a smaller bandwidth. The hypotheses (C3)–(C5) are common hypotheses in kernel regression; see [13].

Let $\xi_n = (nh_n / \log n)^{-1/4}$, $\kappa_n = h_n^2 + \xi_n$, and $\delta_\beta = \|\beta - \beta_0\|$. For any k , e_k denotes a k -dimension vector with all elements equal to 1, and E_k denotes a $p \times p$ matrix with all elements equal to 1. For any vector $V(v)$ of functions of v , we define $(V(v))' = dV(v)/dv$. Recall that $K(\cdot)$ is a symmetric density function. Thus, $\mu_0 = \int K(v)dv = 1$ and $\mu_1 = \int vK(v)dv = 0$. For ease of exposition, we assume that $\mu_2 = \int v^2K(v)dv = 1$. Otherwise, $K(v) = \mu_2^{1/2}K(\mu_2^{1/2}v)$. Let $\mathcal{D}_n = \{x : \|x\| < n^c, f_\beta(\beta^T x) > n^{-\epsilon}, \beta \in \mathcal{B}_n\}$, where $\mathcal{B}_n = \{\beta : \|\beta - \beta_0\| < n^{-1+2c_0}\}$, $c \geq 1$. In order to prove the result of the theorems, we first present the following lemmas.

Appendix B. Proof

Lemma A1. Suppose $m_n(\chi, Z)$, $n = 1, 2, \dots$, are measurable functions of Z with index $\chi \in R^d$, where d is any integer number, such that (i) $|m_n(\chi, Z)| \leq a_n M(Z)$ with $E(M(Z)^r) < \infty$ for some $r > 2$, and a_n increases with n such that $a_n < c_0 n^{1-2/r}$; (ii) $E(m_n(\chi, Z)) \leq a_n m_0^2(\chi)$ with $|m_0(\chi) - m_0(\chi')| \leq c\|\chi - \chi'\|^{\alpha_1}$, where $\alpha_1 > 0$ and $c > 0$ are two constants (without loss of generality, we assume $m_0(\chi) \geq 1$); and that (iii) $|m_n(\chi, Z) - m_n(\chi', Z)| \leq |\chi - \chi'|^{\alpha_1} n^{\alpha_2} G(Z)$ with some $\alpha_2 > 0$, and $EG^2(Z)$ exists. Suppose $\{Z_i, i = 1, \dots, n\}$ is a random sample from Z . Then, for any positive α_0 , we have:

$$\sup_{|\chi| \leq n^{\alpha_0}} |\{m_0(\chi)\}^{-1} n^{-1} \sum_{\substack{i \in \mathcal{I}_n \\ t \in \mathcal{T}_n}} \{m_n(\chi, Z_{i,t}) - Em_n(\chi, Z_{i,t})\}| = O_P\{(a_n \log n / n)^{1/4}\}.$$

Proof of Lemma A1. For the “continuity argument” approach, see [13,25]. For simplicity, let $m_n(\chi, Z)$ denote $m_n(\chi, Z)/m_0(\chi)$. Let $\mathcal{D}_n \stackrel{\text{def}}{=} \{|\chi| < n^{\alpha_0}\}$ be bounded and its Borel measure be less than $n^{d+\alpha_0}$. There are n^{α_4} ($\alpha_4 > d + \alpha_0$) balls B_{n_k} centered at χ_{n_k} , $1 \leq k \leq n^{\alpha_4}$ with a diameter smaller than $cn^{2(2+\alpha_2)/\alpha_1}$, so that $\mathcal{D}_n \subset \cup_{1 \leq k \leq n^{\alpha_4}} B_{n_k}$. Then,

$$\begin{aligned} & \sup_{\chi \in \mathcal{D}_n} \left| \frac{1}{n} \sum_{\substack{i \in \mathcal{I}_n \\ t \in \mathcal{T}_n}} \{m_n(\chi, Z_{i,t}) - Em_n(\chi, Z_{i,t})\} \right| \\ & \leq \max_{1 \leq k \leq n^{\alpha_4}} \left| \frac{1}{n} \sum_{\substack{i \in \mathcal{I}_n \\ t \in \mathcal{T}_n}} \{m_n(\chi_{n_k}, Z_{i,t}) - Em_n(\chi_{n_k}, Z_{i,t})\} \right| \\ & + \max_{1 \leq k \leq n^{\alpha_4}} \sup_{\chi \in B_{n_k}} \left| \frac{1}{n} \sum_{\substack{i \in \mathcal{I}_n \\ t \in \mathcal{T}_n}} [\{m_n(\chi, Z_{i,t}) - m_n(\chi_{n_k}, Z_{i,t})\} - E\{m_n(\chi, Z_{i,t}) - m_n(\chi_{n_k}, Z_{i,t})\}] \right| \\ & \stackrel{\text{def}}{=} \max_{1 \leq k \leq n^{\alpha_4}} |R_{n,k,1}| + \max_{1 \leq k \leq n^{\alpha_4}} \sup_{\chi \in B_{n_k}} |R_{n,2}|. \end{aligned} \tag{A3}$$

By (ii), we have:

$$\begin{aligned} \max_{1 \leq k \leq n^{\alpha_4}} \sup_{\chi \in B_{n_k}} |m(\chi, Z_{i,t}) - m(\chi_{n_k}, Z_{i,t})| & \leq \max_{1 \leq k \leq n^{\alpha_4}} \sup_{\chi \in B_{n_k}} n^{\alpha_2} |\chi - \chi_{n_k}|^{\alpha_1} G(Z_{i,t}) \\ & \leq cn^{-2} G(Z_{i,t}). \end{aligned}$$

By assumption (iii),

$$\max_{1 \leq k \leq n^{\alpha_4}} \sup_{\chi \in B_{n_k}} |R_{n,2}| \leq cn^{-2} \sum_{\substack{i \in \mathcal{I}_n \\ t \in \mathcal{T}_n}} G(Z_{i,t}) = O(n^{-1}). \tag{A4}$$

Let $L_n = n \log(n) a_n^{-\frac{8}{9}}$, $m_n^L(\chi, Z_{i,t}) = m_n(\chi_{n_k}, Z_{i,t}) I\{|M(Z_{i,t})| \geq L_n\}$, and $m_n^O(\chi, Z_{i,t}) = m_n(\chi_{n_k}, Z_{i,t}) - m_n^L(\chi, Z_{i,t})$; then, we have:

$$R_{n,k,1} = \frac{1}{n} \sum_{\substack{i \in \mathcal{I}_n \\ t \in \mathcal{T}_n}} \left[m_n^L(\chi_{n_k}, Z_{i,t}) - E\{m_n^L(\chi_{n_k}, Z_{i,t})\} \right] + \frac{1}{n} \sum_{\substack{i \in \mathcal{I}_n \\ t \in \mathcal{T}_n}} \zeta_{n_k,i,t}, \tag{A5}$$

where $\zeta_{n_k,i,t} = m_n^O(\chi_{n_k}, Z_{i,t}) - E\{m_n^O(\chi_{n_k}, Z_{i,t})\}$. Because of $P(|M(Z)| > L_n) \leq L_n^{-r} E|M(Z)|^r$ and $\sum_{n=1}^{\infty} T_n^{-r} < \infty$, for all sufficiently large n , with probability 1, $|M(Z_{i,t})| \leq L_n$ for (i, t) . This implies that the first term on the right-hand side of (A5) is eventually 0. By assumption (ii), we have:

$$\max_{1 \leq k \leq n^{\alpha_4}} \text{Var} \left(\sum_{\substack{i \in \mathcal{I}_n \\ t \in \mathcal{T}_n}} n \zeta_{n_k,i,t} \right) \leq c_1 n a_n \stackrel{\text{def}}{=} S_1. \tag{A6}$$

By (i), we can obtain:

$$\max_{1 \leq k \leq n^{\alpha_4}} |\zeta_{n_k,i,t}| \leq c_0 a_n T_n \leq c \{n a_n \log n\}^{1/9} \stackrel{\text{def}}{=} S_2. \tag{A7}$$

Let $S_3 = c \{n a_n \log n\}^{\frac{3}{4}}$, $N = \frac{S_3}{c S_2}$. According to Theorem 2.1 of [26],

$$\begin{aligned} P \left(\left| \sum_{\substack{i \in \mathcal{I}_n \\ t \in \mathcal{T}_n}} \zeta_{n_k,i,t} \right| > S_3 \right) &\leq 4 \exp \left\{ -S_3^2 \left(64 \frac{n}{N} S_1 + \frac{8}{3} S_3 N S_2 \right)^{-1} \right\} + 4 \frac{n}{N} \alpha_N \\ &\leq \exp c a_n \log n n^{\frac{1}{8}} + 4 \frac{n}{N} \alpha_N. \end{aligned} \tag{A8}$$

According to the α mixing condition, $\alpha_N \rightarrow 0$ when N is large enough. Hence,

$$\sum_{n=1}^{\infty} \Pr \left(\max_{1 \leq k \leq n^{\alpha_4}} \left| \sum_{\substack{i \in \mathcal{I}_n \\ t \in \mathcal{T}_n}} \zeta_{n_k,i,t} \right| \geq S_3 \right) < \infty.$$

By the Borel-Cantelli lemma, we have:

$$\max_{1 \leq k \leq n^{\alpha_4}} \left| \sum_{\substack{i \in \mathcal{I}_n \\ t \in \mathcal{T}_n}} \zeta_{n_k,i,t} \right| = O(S_3). \tag{A9}$$

Combining (A4) and (A9), we have:

$$\max_{1 \leq k \leq n^{\alpha_4}} |R_{n,k,1}| = O(S_3). \tag{A10}$$

Therefore, Lemma A1 follows. \square

Lemma A2 ([11]). Assume that $\varphi(\beta)$ and $\phi(\beta)$ are two measurable functions of (X, Y) such that $\sup_{\beta, \theta} |\varphi(\beta) - \varphi(\theta)| < |\beta - \theta| \zeta$ a.s., $\sup_{\beta, \theta} |\phi(\beta) - \phi(\theta)| < |\beta - \theta| \zeta$ a.s., with $E \zeta^r < c$ for some $r > 2$, and $E\{\phi(\beta) | \beta^T X\} = 0$ for all $\beta \in \mathcal{B}$. Suppose that $(X_{i,t}, Y_{i,t})$, $\varphi_i(\beta)$, and $\phi_i(\beta)$, $i = 1, \dots, n$ are random copies of (X, Y) , $\varphi(\beta)$ and $\phi(\beta)$, respectively. If (C1), (C3), and (C4) hold, then:

$$\sup_{\beta \in \mathcal{B}} \left| \frac{1}{2} \sum_{i=1}^n \sum_{j=1}^n [K_{h_n}(\beta^T X_{ij}) \varphi_j(\beta) - E_j\{K_{h_n}(\beta^T X_{ij}) \varphi_j(\beta)\}] \phi_i(\beta) \right| = O(\zeta_n^2).$$

Lemma A3. Suppose $E(Z|\beta^T X = \beta^T x) = m(\beta^T x)$ and its derivatives up to the second order are bounded for all $\beta \in \mathcal{B}$, and suppose that $E\|Z\|^r$ exists for some $r > 3$. Let $(X_{i,t}, Z_{i,t})$ be a sample from (X, Z) and $X_{(i,t)x} = X_{i,t} - x$. If conditions (C1), (C3), and (C4) hold, then:

$$n^{-1} \sum_{\substack{i \in \mathcal{I}_n \\ t \in \mathcal{T}_n}} K_{h_n}(\beta^T X_{(i,t)x}) (\beta^T X_{(i,t)x} / h_n)^p Z_i = f_\beta(\beta^T x) m(\beta^T x) \mu_p + \{f_\beta(\beta^T x) m(\beta^T x)\}' \mu_{p+1} h_n + O_P(\kappa_n),$$

where $\mu_p = \int K(v) v^p dv, p = 0, 1, \dots$.

Proof of Lemma A3. Since:

$$E\{K_{h_n}(\beta^T X_{(i,t)x}) (\beta^T X_{(i,t)x} / h_n)^p m(\beta^T X_{i,t})\} = f_\beta(\beta^T x) m(\beta^T x) \mu_p + \{f_\beta(\beta^T x) m(\beta^T x)\}' \mu_{p+1} h_n + O_P(h_n^2),$$

where $\mu_p = \int v^p K(v) dv$, combining it with Lemma A1, we complete the proof of Lemma A3. \square

Lemma A4. Let

$$\Sigma_n^\beta(x) = n^{-1} \sum_{\substack{i \in \mathcal{I}_n \\ t \in \mathcal{T}_n}} K_{h_n}(\beta^T X_{(i,t)x}) \begin{pmatrix} 1 \\ \beta^T X_{(i,t)x} / h_n \end{pmatrix} \begin{pmatrix} 1 \\ \beta^T X_{(i,t)x} / h_n \end{pmatrix}^T,$$

and

$$\begin{pmatrix} a_x^\beta \\ b_x^\beta h_n \end{pmatrix} = \{n \Sigma_n^\beta(x)\}^{-1} \sum_{i=1}^n K_{h_n}(\beta^T X_{(i,t)x}) \begin{pmatrix} 1 \\ \beta^T X_{(i,t)x} / h_n \end{pmatrix} Y_{i,t}.$$

Under conditions (C1)–(C4), we have:

$$a_x^\beta = g(\beta_0^T x) + g'(\beta_0^T x) v_{\beta_0}^T(x) (\beta_0 - \beta) + \frac{1}{2} g''(\beta_0^T x) h_n^2 + \Theta_{n,1}(x) + O_P\{(h_n \kappa_n + \delta_{\beta 2}) n^\epsilon\} (1 + \|x\|^4),$$

$$b_x^\beta h_n = g'(\beta_0^T x) h_n + \Theta_{n,2}(x) + O_P\{(h_n \kappa_n + \delta_{\beta 2}) n^\epsilon e_p\} (1 + \|x\|^4),$$

where:

$$\Theta_{n,1}(x) = \{n f_\beta(\beta^T x)\}^{-1} \sum_{\substack{i \in \mathcal{I}_n \\ t \in \mathcal{T}_n}} K_{h_n}(\beta^T X_{(i,t)x}) \varepsilon_{i,t},$$

and

$$\Theta_{n,2}(x) = \{n h_n f_\beta(\beta^T x)\}^{-1} \sum_{\substack{i \in \mathcal{I}_n \\ t \in \mathcal{T}_n}} K_{h_n}(\beta^T X_{(i,t)x}) \beta^T X_{(i,t)x} \varepsilon_{i,t}.$$

Proof of Lemma A4. By the Taylor expansion of $g(\beta_0^T X_{i,t})$ at $\beta_0^T x$ and $\beta_0 = \beta + (\beta_0 - \beta)$, we have:

$$\begin{aligned} Y_{i,t} &= g(\beta_0^T x) + g'(\beta_0^T x) \beta_0^T + \frac{1}{2} g''(\beta_0^T x) (\beta_0^T X_{(i,t)x})^2 + \varepsilon_{i,t} + O(|\beta_0^T X_{(i,t)x}|^3) \\ &= g(\beta_0^T x) + g'(\beta_0^T x) \beta_0^T X_{(i,t)x} + \frac{1}{2} g''(\beta_0^T x) (\beta_0^T X_{(i,t)x})^2 + \varepsilon_{i,t} + \Omega_n(x, X_{i,t}, \beta) \\ &= g(\beta_0^T x) + g'(\beta_0^T x) \beta^T X_{(i,t),x} + \frac{1}{2} g''(\beta_0^T x) (\beta_0^T X_{(i,t),x})^2 + g'(\beta_0^T x) (\beta_0 - \beta)^T X_{(i,t),x} \\ &\quad + \varepsilon_{i,t} + \Omega_n(x, X_{i,t}, \beta), \end{aligned} \tag{A11}$$

where $\Omega_n(x, X_i, \beta) = O(|\beta^T X_{(i,t),x}|^3 + |\beta^T X_{(i,t),x}| \cdot \|X_{(i,t),x}\| \delta_\beta + \|X_{(i,t),x}\|^2 \delta_\beta^2)$. By Equation (A11), it follows that:

$$\begin{aligned}
 \begin{pmatrix} a_x^\beta \\ b_x^\beta h_n \end{pmatrix} &= \{n\Sigma_n^\beta(x)\}^{-1} \sum_{\substack{i \in \mathcal{I}_n \\ t \in \mathcal{T}_n}} K_{h_n}(\beta^T X_{(i,t),x}) \begin{pmatrix} 1 \\ \beta^T X_{(i,t),x}/h_n \end{pmatrix} \\
 &\cdot \{g(\beta_0^T x) + g'(\beta_0^T x)\beta^T X_{(i,t),x}\} \\
 &+ \{n\Sigma_n^\beta(x)\}^{-1} \sum_{\substack{i \in \mathcal{I}_n \\ t \in \mathcal{T}_n}} K_{h_n}(\beta^T X_{(i,t),x}) \begin{pmatrix} 1 \\ \beta^T X_{(i,t),x}/h_n \end{pmatrix} \frac{1}{2} g''(\beta_0^T x) (\beta_0^T X_{(i,t),x})^2 \\
 &+ \{n\Sigma_n^\beta(x)\}^{-1} \sum_{\substack{i \in \mathcal{I}_n \\ t \in \mathcal{T}_n}} K_{h_n}(\beta^T X_{(i,t),x}) \begin{pmatrix} 1 \\ \beta^T X_{(i,t),x}/h_n \end{pmatrix} g'(\beta_0^T x) (\beta_0 - \beta)^T X_{(i,t),x} \\
 &+ \{n\Sigma_n^\beta(x)\}^{-1} \sum_{\substack{i \in \mathcal{I}_n \\ t \in \mathcal{T}_n}} K_{h_n}(\beta^T X_{(i,t),x}) \begin{pmatrix} 1 \\ \beta^T X_{(i,t),x}/h_n \end{pmatrix} \varepsilon_{i,t} \\
 &+ \{n\Sigma_n^\beta(x)\}^{-1} \sum_{\substack{i \in \mathcal{I}_n \\ t \in \mathcal{T}_n}} K_{h_n}(\beta^T X_{(i,t),x}) \begin{pmatrix} 1 \\ \beta^T X_{(i,t),x}/h_n \end{pmatrix} \Omega_n(x, X_{i,t}, \beta) \\
 &:= A_1 + A_2 + A_3 + A_4 + A_5. \tag{A12}
 \end{aligned}$$

According to Lemma A3, we have:

$$\Sigma_n^\beta(x) = \begin{pmatrix} f_\beta(\beta^T x) & f'_\beta(\beta^T x)h_n \\ f'_\beta(\beta^T x)h_n & f_\beta(\beta^T x) \end{pmatrix} + O_p(\kappa_n).$$

If $f_\beta(\beta^T x) > n^\epsilon$, we have:

$$\{\Sigma_n^\beta(x)\}^{-1} = f_\beta^{-1}(\beta^T x) \left\{ I - f_\beta^{-1} \begin{pmatrix} 0 & f'_\beta(\beta^T x)h_n \\ f'_\beta(\beta^T x)h_n & 0 \end{pmatrix} \right\} + O_p(\kappa_n n^\epsilon).$$

It is easy to check that

$$A_1 = \begin{pmatrix} g(\beta_0^T x) \\ g'(\beta_0^T x)h_n \end{pmatrix}. \tag{A13}$$

According to Lemma A3 and Equation (A11), we have:

$$A_2 = \begin{pmatrix} g''(\beta_0^T x)h_n^2 \\ g''(\beta_0^T x)f'_\beta(\beta^T x)h_n^3 \end{pmatrix} + O_p(h_n \kappa_n n^\epsilon e_{d+1}). \tag{A14}$$

By Lemma A3, we have:

$$A_3 = \begin{pmatrix} g'(\beta_0^T x)v_\beta^T(x)(\beta_0 - \beta) \\ O(h_n \delta_\beta)(1 + \|x\|) \end{pmatrix} + O_p(\kappa_n \delta_\beta n^\epsilon e_{d+1})(1 + \|x\|). \tag{A15}$$

For the noise term, we have:

$$A_4 = \{nf_\beta(\beta^T x)\}^{-1} \sum_{i=1}^n K_{h_n}(\beta^T X_{(i,t),x}) \begin{pmatrix} 1 \\ \beta^T X_{(i,t),x}/h_n \end{pmatrix} \varepsilon_{i,t} + O_p(\kappa_n \delta_\beta n^\epsilon e_{d+1}). \tag{A16}$$

It follows from Lemma A2 that:

$$n^{-1} \sum_{i=1}^n K_{h_n}(\beta^T X_{(i,t),x}) \begin{pmatrix} 1 \\ \beta^T X_{(i,t),x}/h_n \end{pmatrix} |\beta^T X_{(i,t),x}|^k \|X_{(i,t),x}\|^l = O_p(h_n^k)(1 + \|x\|^l).$$

Then, we have:

$$A_5 = O_p\{(h_n^3 + \delta_\beta^2)n^\epsilon e_{d+1}\}(1 + \|x\|^3). \tag{A17}$$

Thus, Lemma A4 follows from (A13)–(A17). □

Lemma A5. Under conditions (C1)–(C4), we have:

$$\begin{aligned} & \left\{ n^{-2} \sum_{j \in \mathcal{I}_n} \sum_{\tau \in T_n} \sum_{i \in \mathcal{I}_n} \sum_{t \in T_n} \hat{\rho}_{j,\tau}^\beta (b_{j,\tau}^\beta)^2 K_{h_n}(\beta^T X_{(i,t),(j,\tau)}) X_{(i,t),(j,\tau)} X_{(i,t),(j,\tau)}^T / \hat{f}_\beta(\beta^T X_{j,\tau}) \right\}^{-1} \\ = & \beta_0 \beta_0^T [E\{g'(\beta_0^T X)^2\}]^{-1} h_n^{-2} - \frac{1}{2} [E\{g'(\beta_0^T X)^2\}]^{-1} (\beta_0 F^T W_{g_0}^+ + W_{g_0}^+ F \beta_0^T) + \frac{1}{2} W_{g_0}^+ \\ & + O_p\{(\kappa_n h_n n^\epsilon + \delta_\beta / h_n^2) E_d\}, \end{aligned}$$

where $W_{g_0} = E\{g'(\beta_0^T X)^2 v_{\beta_0}(X) v_{\beta_0}^T(X)\}$ and $F = E\left\{ \frac{[g'(\beta_0^T X)]^2 (f_{\beta_0}(\beta_0^T X) v_{\beta_0}(X))'}{f_{\beta_0}(\beta_0^T X)} \right\}$.

Proof of Lemma A5. According to condition (C1), we have:

$$\begin{aligned} \sum_{i \in \mathcal{I}_n} \sum_{t \in T_n} P\left(\bigcup_{\substack{i \in \mathcal{I}_n \\ t \in T_n}} \{X_{i,t} \notin \mathcal{D}_n\}\right) & \leq \sum_{i \in \mathcal{I}_n} \sum_{t \in T_n} nP(X_{i,t} \notin \mathcal{D}_n) \leq \sum_{i \in \mathcal{I}_n} \sum_{t \in T_n} nP(|X_{i,t}| \geq n^c) \\ & < \sum_{i \in \mathcal{I}_n} \sum_{t \in T_n} nn^{-6c} E|X|^6 < \infty. \end{aligned}$$

By the Borel–Cantelli lemma, we have:

$$P\left(\bigcup_{\substack{i \in \mathcal{I}_n \\ t \in T_n}} \{X_{i,t} \notin \mathcal{D}_n\}, i.o.\right) = 0. \tag{A18}$$

Let $\tilde{\mathcal{D}}_n = \{x : f_\beta(\beta^T x) > 2n^{-\epsilon}\}$. Similarly, we have:

$$P\left(\bigcup_{\substack{i \in \mathcal{I}_n \\ t \in T_n}} \{X_{i,t} \notin \tilde{\mathcal{D}}_n\}, i.o.\right) = 0. \tag{A19}$$

Hence, we can exchange summations over $\{X_{j,\tau} : j \in \mathcal{I}_n, \tau \in T_n\}$, $\{X_{j,\tau} : X_{j,\tau} \in \mathcal{D}_n, j \in \mathcal{I}_n, \tau \in T_n\}$, $\{X_{j,\tau} : X_{j,\tau} \in \tilde{\mathcal{D}}_n, j \in \mathcal{I}_n, \tau \in T_n\}$ in the sense of almost sure consistently. By Lemma A3 and condition (C1), we have:

$$\hat{f}_\beta(\beta^T x) = f_\beta(\beta^T x) + O(\kappa_n).$$

By Equation (A19) and the smoothness of $\rho_n(\cdot)$ with $|\rho'_n(\cdot)| < cn^\epsilon$ for some $c > 0$, we have:

$$\rho_n(\hat{f}_\beta(\beta^T x)) = \rho_n(f_\beta(\beta^T x)) + O(\kappa_n n^\epsilon) = 1 + O(\kappa_n n^\epsilon). \tag{A20}$$

Denote by (β, D) an orthogonal matrix. According to Lemma A3, we have:

$$\begin{aligned} n^{-1} \sum_{i \in \mathcal{I}_n} \sum_{t \in T_n} K_{h_n}(\beta^T X_{(i,t),x}) \beta^T X_{(i,t),x} X_{(i,t),x}^T \beta & = f_\beta(\beta^T x) h_n^2 + O_p(h_n^2 \kappa_n), \\ n^{-1} \sum_{i \in \mathcal{I}_n} \sum_{t \in T_n} K_{h_n}(\beta^T X_{(i,t),x}) \beta^T X_{(i,t),x} X_{(i,t),x}^T D & = f'_\beta(\beta^T x) v_\beta^T(x) D h_n^2 + O_p(h_n^2 \kappa_n e_{p-1}), \\ n^{-1} \sum_{i \in \mathcal{I}_n} \sum_{t \in T_n} K_{h_n}(\beta^T X_{(i,t),x}) D^T X_{(i,t),x} X_{(i,t),x}^T D & = f_\beta(\beta^T x) C^\beta(x) + O_p(\kappa_n E_{p-1})(1 + |x|^2), \end{aligned}$$

where $C^\beta(x) = D^T \tilde{w}_\beta(x)D$, $\tilde{w}_\beta(x) = w_\beta(x) - x\mu_\beta^T(x) - \mu_\beta(x)x^T + xx^T$. It follows from (A19) that:

$$\begin{aligned} & \{n\hat{f}_\beta(\beta^T x)\}^{-1} \sum_{i \in \mathcal{I}_n} \sum_{t \in T_n} K_{h_n}(\beta^T X_{(i,t),x}) X_{(i,t),x} X_{(i,t),x}^T \\ &= (\beta, D) \begin{pmatrix} h_n^2 & F_\beta^T(x) Dh_n^2 \\ D^T F_\beta(x) h_n^2 & C^\beta(x) \end{pmatrix} (\beta, D)^T + O_p(h_n \kappa_n n^\epsilon E_d)(1 + \|x\|^2), \end{aligned}$$

where $F_\beta(x) = f'_\beta(\beta^T x)v_\beta(x)/f_\beta(\beta^T x)$. According to (A20), we have: $b_x^\beta = g'(\beta_0^T x) + O(\kappa_n n^{2\epsilon}/h_n)(1 + \|x\|^4)$. Note that $E\{g'(\beta_0^T X_{j,\tau})^2 C^{\beta_0}(X_{j,\tau})\} = 2D_0^T W_{g_0} D_0$. Thus,

$$\begin{aligned} & n^{-2} \sum_{i \in \mathcal{I}_n} \sum_{t \in T_n} \sum_{j \in \mathcal{I}_n} \sum_{\tau \in T_n} \hat{\rho}_{j,\tau}^\beta (b_{j,\tau}^\beta)^2 K_{h_n}(\beta^T X_{(i,t),(j,\tau)}) X_{(i,t),(j,\tau)} X_{(i,t),(j,\tau)}^T / \hat{f}_\beta(\beta^T X_{j,\tau}) \\ &= n^{-1} \sum_{X_{j,\tau} \in \mathcal{D}_n} \rho_n(\hat{f}_\beta(\beta^T X_{j,\tau})) (b_{j,\tau}^\beta)^2 \{n\hat{f}_\beta(\beta^T X_{j,\tau})\}^{-1} \\ & \cdot \sum_{i \in \mathcal{I}_n} \sum_{t \in T_n} K_{h_n}(\beta^T X_{(i,t),x}) X_{(i,t),(j,\tau)} X_{(i,t),(j,\tau)}^T \\ &= (\beta, D) n^{-1} \sum_{\substack{j \in \mathcal{I}_n \\ \tau \in T_n}} \left\{ g'(\beta_0^T X_{j,\tau})^2 \begin{pmatrix} h_n^2 & F_\beta^T(X_{j,\tau}) Dh_n^2 \\ D^T F_\beta(X_{j,\tau}) h_n^2 & C^\beta(X_{j,\tau}) \end{pmatrix} + O(h_n \kappa_n n^\epsilon)(1 + |X_{j,\tau}|^6) \right\} (\beta, D)^T \\ &= (\beta, D) \begin{pmatrix} E\{g'(\beta_0^T X)^2\} h_n^2 & F^T Dh_n^2 \\ D^T F h_n^2 & 2D^T W_{g_0} D \end{pmatrix} (\beta, D)^T + O(h_n \kappa_n n^\epsilon + \delta_\beta E_d), \end{aligned}$$

where $F = E\{[g'(\beta_0^T X)]^2 F_{\beta_0}(X)\}$. Therefore, by the matrix inversion formula in blocks [27], we have:

$$\begin{aligned} & \left\{ n^{-2} \sum_{i \in \mathcal{I}_n} \sum_{t \in T_n} \sum_{j \in \mathcal{I}_n} \sum_{\tau \in T_n} \hat{\rho}_{j,\tau}^\beta (b_{j,\tau}^\beta)^2 K_{h_n}(\beta^T X_{(i,t),(j,\tau)}) X_{(i,t),(j,\tau)} X_{(i,t),(j,\tau)}^T / \hat{f}_\beta(\beta^T X_j) \right\}^{-1} \\ &= \beta \beta^T [E\{g'(\beta_0^T X_{j,\tau})^2\}]^{-1} h_n^{-2} - \frac{1}{2} [E\{g'(\beta_0^T X_{j,\tau})^2\}]^{-1} (DHD^T F \beta^T + \beta F^T D^T HD) \\ & \quad + \frac{1}{2} DHD^T + O\{h_n^{-2}(h_n \kappa_n n^\epsilon + \delta_\beta) E_d\}, \end{aligned}$$

where $H = (D^T W_{g_0} D)^{-1}$. Note that $D = D_0 + \delta_\beta$. Then, $H = (D_0^T W_{g_0} D_0)^{-1} + \delta_\beta$. By the definition of the Moore–Penrose inverse, we have:

$$D_0(D_0^T W_{g_0} D_0)^{-1} D_0^T = W_{g_0}^+.$$

Combining the facts that $DD^T = I - \beta \beta^T = I - \beta_0 \beta_0^T + O(\delta_\beta)$ and $\beta_0^T W_{g_0} = 0$, we complete the proof. \square

Lemma A6. Under conditions (C1)–(C4), we have:

$$\begin{aligned} & n^{-2} \sum_{\substack{j \in \mathcal{I}_n \\ \tau \in T_n}} \sum_{i \in \mathcal{I}_n} \sum_{t \in T_n} \hat{\rho}_{j,\tau}^\beta K_{h_n}(\beta^T X_{(i,t),(j,\tau)}) b_{j,\tau}^\beta X_{(i,t),(j,\tau)} (Y_{i,t} - a_{j,\tau}^\beta - b_{j,\tau}^\beta \beta_0^T X_{(i,t),(j,\tau)}) / \hat{f}_\beta(\beta^T X_{j,\tau}) \\ &= W_{g_0}(\beta - \beta_0) + n^{-1} \sum_{i \in \mathcal{I}_n} \sum_{t \in T_n} g'(\beta_0^T X_{i,t}) v_{\beta_0}(X_{i,t}) \varepsilon_{i,t} + o_p(n^{-1/2} e_d). \end{aligned}$$

Proof of Lemma A6. According to Lemma A4, we have:

$$\begin{aligned}
 & n^{-2} \sum_{\substack{j \in \mathcal{I}_n \\ \tau \in \mathcal{T}_n}} \sum_{i \in \mathcal{I}_n} \hat{\rho}_{j,\tau}^\beta K_{h_n}(\beta^T X_{(i,t),(j,\tau)}) b_{j,\tau}^\beta X_{(i,t),(j,\tau)} (Y_i - a_{j,\tau}^\beta - b_{j,\tau}^\beta \beta_0^T X_{(i,t),(j,\tau)}) / \hat{f}_\beta(\beta^T X_{j,\tau}) \\
 = & n^{-2} \sum_{\substack{j \in \mathcal{I}_n \\ \tau \in \mathcal{T}_n}} \sum_{i \in \mathcal{I}_n} \hat{\rho}_{j,\tau}^\beta \hat{f}_\beta^{-1}(\beta^T X_{j,\tau}) K_{h_n}(\beta^T X_{(i,t),(j,\tau)}) b_{j,\tau}^\beta X_{(i,t),(j,\tau)} \varepsilon_{i,t} \\
 & + n^{-2} \sum_{\substack{j \in \mathcal{I}_n \\ \tau \in \mathcal{T}_n}} \sum_{i \in \mathcal{I}_n} \hat{\rho}_{j,\tau}^\beta \hat{f}_\beta^{-1}(\beta^T X_{j,\tau}) K_{h_n}(\beta^T X_{(i,t),(j,\tau)}) b_{j,\tau}^\beta X_{(i,t),(j,\tau)} g'(\beta_0^T X_{j,\tau}) v_\beta^T(X_{j,\tau})(\beta - \beta_0) \\
 & + n^{-2} \sum_{\substack{j \in \mathcal{I}_n \\ \tau \in \mathcal{T}_n}} \sum_{i \in \mathcal{I}_n} \hat{\rho}_{j,\tau}^\beta \hat{f}_\beta^{-1}(\beta^T X_{j,\tau}) K_{h_n}(\beta^T X_{(i,t),(j,\tau)}) b_{j,\tau}^\beta X_{(i,t),(j,\tau)} \frac{1}{2} g''(\beta_0^T X_{j,\tau}) \\
 & \times \{(\beta^T X_{(i,t),(j,\tau)})^2 - h_n^2\} \\
 & - n^{-2} \sum_{\substack{j \in \mathcal{I}_n \\ \tau \in \mathcal{T}_n}} \sum_{i \in \mathcal{I}_n} \hat{\rho}_{j,\tau}^\beta \hat{f}_\beta^{-1}(\beta^T X_{j,\tau}) K_{h_n}(\beta^T X_{(i,t),(j,\tau)}) \\
 & \times b_{j,\tau}^\beta X_{(i,t),(j,\tau)} \{ \Theta_{n,1}(X_{j,\tau}) + \Theta_{n,2}(X_{j,\tau}) \beta_0^T X_{(i,t),(j,\tau)} / h_n \} \\
 & + n^{-2} \sum_{\substack{j \in \mathcal{I}_n \\ \tau \in \mathcal{T}_n}} \sum_{i \in \mathcal{I}_n} \hat{\rho}_{j,\tau}^\beta \hat{f}_\beta^{-1}(\beta^T X_{j,\tau}) K_{h_n}(\beta^T X_{(i,t),(j,\tau)}) b_{j,\tau}^\beta X_{(i,t),(j,\tau)} \\
 & \times O\{(h_n \kappa_n + \delta_\beta^2) n^\epsilon\} (1 + |X_{j,\tau}|^4) (1 + |\beta_0^T X_{(i,t),(j,\tau)}| / h_n) \\
 := & B_1 + B_2 + B_3 + B_4 + B_5.
 \end{aligned}$$

Denote by $U_{(i,t),(j,\tau)} = \rho_n(f_\beta(\beta^T X_{j,\tau})) K_{h_n}(\beta^T X_{(i,t),(j,\tau)}) b_{j,\tau}^\beta X_{(i,t),(j,\tau)}$. We have $E\{E_{j,\tau}(U_{(i,t),(j,\tau)}) - g'(\beta_0^T X_{j,\tau}) v_\beta(X_{j,\tau})\}^2 = O\{(h_n + \delta_\beta)^2\}$. Note that $\Theta_{n,1} = O(\kappa_n n^\epsilon)$ and $\Theta_{n,2} = O(\kappa_n n^\epsilon)$. By Lemmas A1 and A2, we have:

$$\begin{aligned}
 B_1 &= n^{-2} \sum_{\substack{j \in \mathcal{I}_n \\ \tau \in \mathcal{T}_n}} \sum_{i \in \mathcal{I}_n} \{U_{(i,t),(j,\tau)} - E_j(U_{(i,t),(j,\tau)})\} \varepsilon_{i,t} + n^{-1} \sum_{\substack{i \in \mathcal{I}_n \\ t \in \mathcal{T}_n}} E_j(U_{(i,t),(j,\tau)}) \varepsilon_{i,t} + o_d(n^{-1/2} e_d) \\
 &= n^{-1} \sum_{\substack{i \in \mathcal{I}_n \\ t \in \mathcal{T}_n}} g'(\beta_0^T X_{i,t}) v_\beta(X_{i,t}) \varepsilon_i + o_p(n^{-1/2} e_d). \tag{A21}
 \end{aligned}$$

According to Lemma A3, (A18), and (A19), we have:

$$B_2 = W_{g0}(\beta - \beta_0) + O_p\{(\kappa_n + \delta_\beta) e_d / h_n\}. \tag{A22}$$

By Lemmas A3 and A4, we have:

$$B_3 = O_p\{(h_n^2 \kappa_n n^\epsilon e_d)\}. \tag{A23}$$

Letting $V_{(i,t),(j,\tau)} = K_{h_n}(\beta^T X_{(i,t),(j,\tau)}) X_{(i,t),(j,\tau)}$, then we have $E_i(V_{(i,t),(j,\tau)}) = v_\beta(X_{j,\tau}) + O_p(h_n^2 e_d) (1 + \|X_{j,\tau}\|)$. Note that $E\{v_\beta(X_{j,\tau})\} = 0$. According to Lemma A2, we have:

$$\begin{aligned}
 B_4 &= n^{-2} \sum_{\substack{j \in \mathcal{I}_n \\ \tau \in T_n}} \sum_{i \in \mathcal{I}_n} \rho_n(f_{\beta}(\beta^T X_{j,\tau})) V_{(i,t),(j,\tau)} g'(\beta^T X_{j,\tau}) f_{\beta}^{-1}(\beta^T X_{j,\tau}) \\
 &\times \{ \Theta_{n,1}(X_{j,\tau}) + \Theta_{n,2}(X_{j,\tau}) \beta_0^T X_{(i,t),(j,\tau)} / h_n \} \\
 &\quad + O_p(n^{-1/2} e_d) \\
 &= n^{-1} \sum_{\substack{j \in \mathcal{I}_n \\ \tau \in T_n}} \rho_n(f_{\beta}(\beta^T X_{j,\tau})) v_{\beta}(X_{j,\tau}) g'(\beta^T X_{j,\tau}) f_{\beta}^{-1}(\beta^T X_{j,\tau}) \\
 &\times \{ \Theta_{n,1}(X_{j,\tau}) + \Theta_{n,2}(X_{j,\tau}) \beta_0^T X_{(i,t),(j,\tau)} / h_n \} \\
 &\quad + O_p(n^{-1/2} e_d) \\
 &= O_p(n^{-1/2}) e_d.
 \end{aligned} \tag{A24}$$

By Lemma A1 and Corollary 2 in [28] (p.122), we have:

$$B_5 = O_p\{ (h_n \kappa_n + \delta_{\beta}^2 e_d) n^{\epsilon} \} O(1) n^{-1} \sum_{j \in \mathcal{I}_n} \sum_{\tau \in T_n} |X_{j,\tau}|^5 = O_p\{ (h_n \kappa_n + \delta_{\beta}^2) n^{\epsilon} e_d \}. \tag{A25}$$

Combining (A21)–(A25), we complete the proof of this lemma. \square

Lemma A7. According to Theorem 3 in [29], for $\delta < 1 - s^{-1} \lim_{n \rightarrow +\infty} n^{2\delta-1} \tilde{h} = +\infty$, we have:

$$\sup_{x \in R^d} \left| \frac{1}{n} \sum_{i \in \mathcal{I}_n} \sum_{t \in T_n} [\tilde{K}_{\tilde{h}}(\tilde{X}_{i,t} - x) \tilde{Y}_{i,t} - E(\tilde{K}_{\tilde{h}}(\tilde{X}_{i,t} - x) \tilde{Y}_{i,t})] \right| = O_p \left[\left(\frac{\log(n)}{n \tilde{h}^d} \right) \right].$$

Lemma A8. Under conditions (C1)–(C5), we have:

$$\sup_{x \in R^d} \left| \sigma_{\beta_0}^2(x) - \sum_{i \in \mathcal{I}_n} \sum_{t \in T_n} \tilde{w}_{i,t}(x) \sigma_{\beta_0}^2(x) \right| = O_p \left\{ \tilde{h}^{2d} + \left[\frac{\log(1/\tilde{h}^d)}{n \tilde{h}^d} \right]^{\frac{1}{2}} \right\},$$

where:

$$\tilde{w}_{i,t}(x) = \frac{\tilde{K}_{\tilde{h}}(X_{i,t} - x)}{\sum_{j \in \mathcal{I}_n} \sum_{\tau \in T_n} \tilde{K}_{\tilde{h}}(X_{j,\tau} - x)}, j \in \mathcal{I}_n, \tau \in T_n.$$

Proof of Theorem 1. For any $x \in R^d$, let:

$$\omega_{j,\tau}(x) = \frac{L_n(X_{j,\tau} - x)}{\sum_{i \in \mathcal{I}_n} \sum_{t \in T_n} L_n(X_{i,t} - x)}.$$

According to Equations (9) and (10), by Lemma A4, we have:

$$\begin{aligned}
 & \hat{\sigma}_{\hat{\beta}_n}^2(x) \sum_{\substack{j \in \mathcal{I}_n \\ \tau \in \mathcal{T}_n}} \hat{\varepsilon}_{j,\tau}^2 \omega_{j,\tau}(x) = \sum_{\substack{j \in \mathcal{I}_n \\ \tau \in \mathcal{T}_n}} (Y_{j,\tau} - a_{i,t}^{\hat{\beta}_n})^2 \omega_{j,\tau}(x) \\
 &= \sum_{\substack{j \in \mathcal{I}_n \\ \tau \in \mathcal{T}_n}} \left[Y_{j,\tau} - g(\beta_0^T X_{j,\tau}) - g'(\beta_0^T X_{j,\tau}) v_{\beta_0}^T(X_{j,\tau})(\beta_0 - \hat{\beta}_n) - \frac{1}{2} g''(\beta_0^T X_{j,\tau}) h_n^2 - \Theta_{n,1}(X_{j,\tau}) \right]^2 \\
 &\quad \cdot \omega_{j,\tau}(x) \\
 &= \sum_{\substack{j \in \mathcal{I}_n \\ \tau \in \mathcal{T}_n}} \varepsilon_{j,\tau}^2 \omega_{j,\tau}(x) + \sum_{\substack{j \in \mathcal{I}_n \\ \tau \in \mathcal{T}_n}} g'(\beta_0^T X_{j,\tau})^2 v_{\beta_0}^T(X_{j,\tau})(\beta_0 - \hat{\beta}_n) v_{\beta_0}^T(X_{j,\tau})(\beta_0 - \hat{\beta}_n) \omega_{j,\tau}(x) \\
 &+ \frac{1}{4} \sum_{\substack{j \in \mathcal{I}_n \\ \tau \in \mathcal{T}_n}} g''(\beta_0^T X_{j,\tau})^2 h_n^4 \omega_{j,\tau}(x) + \sum_{\substack{j \in \mathcal{I}_n \\ \tau \in \mathcal{T}_n}} \Theta_{n,1}(X_{j,\tau})^2 \omega_{j,\tau}(x) \\
 &- 2 \sum_{\substack{j \in \mathcal{I}_n \\ \tau \in \mathcal{T}_n}} g'(\beta_0^T X_{j,\tau}) v_{\beta_0}^T(X_{j,\tau})(\beta_0 - \hat{\beta}_n) \varepsilon_{j,\tau} \omega_{j,\tau}(x) - \sum_{\substack{j \in \mathcal{I}_n \\ \tau \in \mathcal{T}_n}} g''(\beta_0^T X_{j,\tau}) h_n^2 \varepsilon_{j,\tau} \omega_{j,\tau}(x) \\
 &- 2 \sum_{\substack{j \in \mathcal{I}_n \\ \tau \in \mathcal{T}_n}} \Theta_{n,1}(X_{j,\tau}) \varepsilon_{j,\tau} \omega_{j,\tau}(x) + \sum_{\substack{j \in \mathcal{I}_n \\ \tau \in \mathcal{T}_n}} g''(\beta_0^T X_{j,\tau}) h_n^2 \Theta_{n,1}(X_{j,\tau}) \omega_{j,\tau}(x) \\
 &+ \sum_{\substack{j \in \mathcal{I}_n \\ \tau \in \mathcal{T}_n}} g'(\beta_0^T X_{j,\tau}) v_{\beta_0}^T(X_{j,\tau})(\beta_0 - \hat{\beta}_n) g''(\beta_0^T X_{j,\tau}) h_n^2 \omega_{j,\tau}(x) \\
 &+ 2 \sum_{\substack{j \in \mathcal{I}_n \\ \tau \in \mathcal{T}_n}} g'(\beta_0^T X_{j,\tau}) v_{\beta_0}^T(X_{j,\tau})(\beta_0 - \hat{\beta}_n) \Theta_{n,1}(X_{j,\tau}) \omega_{j,\tau}(x) \\
 &:= C_1 + C_2 + C_3 + C_4 + C_5 + C_6 + C_7 + C_8 + C_9 + C_{10}. \tag{A26}
 \end{aligned}$$

Thus, we can obtain from the results in Lemmas A7 and A8 that:

$$\begin{aligned}
 & \sup_{x \in R^d} \left| C_1 - \sigma_{\beta_0}^2(x) \right| = \sup_{x \in R^d} \left| C_1 - \sum_{\substack{j \in \mathcal{I}_n \\ \tau \in \mathcal{T}_n}} \omega_{j,\tau}(x) \sigma_{\beta_0}^2(X_{j,\tau}) + \sum_{\substack{j \in \mathcal{I}_n \\ \tau \in \mathcal{T}_n}} \omega_{j,\tau}(x) \sigma_{\beta_0}^2(X_{j,\tau}) - \sigma_{\beta_0}^2(x) \right| \\
 &\leq \sup_{x \in R^d} \left| C_1 - \sum_{\substack{j \in \mathcal{I}_n \\ \tau \in \mathcal{T}_n}} \omega_{j,\tau}(x) \sigma_{\beta_0}^2(X_{j,\tau}) \right| + \sup_{x \in R^d} \left| \sum_{\substack{j \in \mathcal{I}_n \\ \tau \in \mathcal{T}_n}} \omega_{j,\tau}(x) \sigma_{\beta_0}^2(X_{j,\tau}) - \sigma_{\beta_0}^2(x) \right| \\
 &= O_p \left(j_n^{2d} + \left[\frac{\log(1/l_n^d)}{nl_n^d} \right]^{\frac{1}{2}} \right). \tag{A27}
 \end{aligned}$$

Since $\hat{\beta}_n \in \mathcal{B}_n$, applying the result of Lemma A7 obtains:

$$\begin{aligned}
 \sup_{x \in R^d} |C_2| &= \sup_{x \in R^d} \left| \sum_{\substack{j \in \mathcal{I}_n \\ \tau \in \mathcal{T}_n}} g'(\beta_0^T X_{j,\tau})^2 v_{\beta_0}^T(X_{j,\tau})(\beta_0 - \hat{\beta}_n) v_{\beta_0}^T(X_{j,\tau})(\beta_0 - \hat{\beta}_n) \omega_{j,\tau}(x) \right| \\
 &= O_p(n^{-1+2c_0}), \tag{A28}
 \end{aligned}$$

$$\sup_{x \in R^d} |C_3| = \sup_{x \in R^d} \left| \frac{1}{4} \sum_{\substack{j \in \mathcal{I}_n \\ \tau \in \mathcal{T}_n}} g''(\beta_0^T X_{j,\tau})^2 h_n^4 \omega_{j,\tau}(x) \right| = O_p(h_n^4), \tag{A29}$$

$$\sup_{x \in R^d} |C_4| = \sup_{x \in R^d} \left| \sum_{\substack{j \in \mathcal{I}_n \\ \tau \in \mathcal{T}_n}} \Theta_{n,1}(X_{j,\tau})^2 \omega_{j,\tau}(x) \right| = O_p \left(\frac{\log(n)}{nl_n^d} \right), \tag{A30}$$

$$\begin{aligned} \sup_{x \in R^d} |C_5| &= \sup_{x \in R^d} \left| 2 \sum_{\substack{j \in \mathcal{I}_n \\ \tau \in \mathcal{T}_n}} g'(\beta_0^T X_{j,\tau}) v_{\beta_0}^T(X_{j,\tau}) (\beta_0 - \hat{\beta}_n) \varepsilon_{j,\tau} \omega_{j,\tau}(x) \right| \\ &= O_p \left[n^{-1+2c_0} \cdot \left(\frac{\log(n)}{nl_n^d} \right)^{\frac{1}{2}} \right], \end{aligned} \tag{A31}$$

$$\sup_{x \in R^d} |C_6| = \sup_{x \in R^d} \left| - \sum_{\substack{j \in \mathcal{I}_n \\ \tau \in \mathcal{T}_n}} g''(\beta_0^T X_{j,\tau}) h_n^2 \varepsilon_{j,\tau} \omega_{j,\tau}(x) \right| = O_p(h_n^2) \cdot O_p \left[\left(\frac{\log(n)}{nl_n^d} \right)^{\frac{1}{2}} \right], \tag{A32}$$

$$\sup_{x \in R^d} |C_7| = \sup_{x \in R^d} \left| -2 \sum_{\substack{j \in \mathcal{I}_n \\ \tau \in \mathcal{T}_n}} \Theta_{n,1}(X_{j,\tau}) \varepsilon_{j,\tau} \omega_{j,\tau}(x) \right| = O_p \left(\frac{\log(n)}{nl_n^d} \right), \tag{A33}$$

$$\sup_{x \in R^d} |C_8| = \sup_{x \in R^d} \left| \sum_{\substack{j \in \mathcal{I}_n \\ \tau \in \mathcal{T}_n}} g''(\beta_0^T X_{j,\tau}) h_n^2 \Theta_{n,1}(X_{j,\tau}) \omega_{j,\tau}(x) \right| = O_p(h_n^2) \cdot O_p \left[\left(\frac{\log(n)}{nl_n^d} \right)^{\frac{1}{2}} \right], \tag{A34}$$

$$\begin{aligned} \sup_{x \in R^d} |C_9| &= \sup_{x \in R^d} \left| \sum_{\substack{j \in \mathcal{I}_n \\ \tau \in \mathcal{T}_n}} g'(\beta_0^T X_{j,\tau}) v_{\beta_0}^T(X_{j,\tau}) (\beta_0 - \hat{\beta}_n) g''(\beta_0^T X_{j,\tau}) h_n^2 \omega_{j,\tau}(x) \right| \\ &= O_p \left(h_n^2 \cdot n^{-1+2c_0} \right), \end{aligned} \tag{A35}$$

$$\begin{aligned} \sup_{x \in R^d} |C_{10}| &= \sup_{x \in R^d} \left| 2 \sum_{\substack{j \in \mathcal{I}_n \\ \tau \in \mathcal{T}_n}} g'(\beta_0^T X_{j,\tau}) v_{\beta_0}^T(X_{j,\tau}) (\beta_0 - \hat{\beta}_n) \Theta_{n,1}(X_{j,\tau}) \omega_{j,\tau}(x) \right| \\ &= O_p \left(n^{-1+2c_0} \right) \cdot O_p \left[\left(\frac{\log(n)}{nl_n^d} \right)^{\frac{1}{2}} \right]. \end{aligned} \tag{A36}$$

Thus, combining (A26)–(A36), we complete the proof of Theorem 1. □

Proof of Theorem 2. Theorem 2 can be proven in the same way as Theorem 1, so it is omitted here. □

Proof of Theorem 3. After one iteration, by Lemmas A5 and A6, we have that the new $\beta^{(1)}$ is:

$$\begin{aligned} \beta^{(1)} &= \beta_0 + \left\{ \sum_{\substack{j \in \mathcal{I}_n \\ \tau \in \mathcal{T}_n}} \sum_{\substack{i \in \mathcal{I}_n \\ t \in \mathcal{T}_n}} \hat{\rho}_{j,\tau}^\beta (b_{j,\tau}^\beta)^2 K_{h_n}(\beta^T X_{(i,t),(j,\tau)}) X_{(i,t),(j,\tau)} X_{(i,t),(j,\tau)}^T / \hat{f}_\beta(\beta^T X_{j,\tau}) \right\}^{-1} \\ &\times \sum_{\substack{j \in \mathcal{I}_n \\ \tau \in \mathcal{T}_n}} \sum_{\substack{i \in \mathcal{I}_n \\ t \in \mathcal{T}_n}} \hat{\rho}_{j,\tau}^\beta K_{h_n}(\beta^T X_{(i,t),(j,\tau)}) X_{(i,t),(j,\tau)} (Y_{i,t} - a_{j,\tau}^\beta - b_{j,\tau}^\beta \beta_0^T X_{(i,t),(j,\tau)}) / \hat{f}_\beta(\beta^T X_{j,\tau}) \\ &= \beta_0 + h_n^{-2} \{E[g'(\beta_0^T X)]^2\}^{-1} \beta_0 \beta_0^T W_{g_0}(\beta - \beta_0) \\ &\quad + h_n^{-2} \{E[g'(\beta_0^T X)]^2\}^{-1} \frac{1}{n} \sum_{i=1}^n g'(\beta_0^T X_{i,t}) \beta_0 \beta_0^T v_{\beta_0}(X_{i,t}) \varepsilon_{i,t} \\ &\quad - \frac{1}{2} \{E[g'(\beta_0^T X)]^2\}^{-1} \beta_0 F^T W_{g_0}^+ W_{g_0}(\beta - \beta_0) \\ &\quad - \frac{1}{2n} \{E[g'(\beta_0^T X)]^2\}^{-1} \sum_{\substack{i \in \mathcal{I}_n \\ t \in \mathcal{T}_n}} g'(\beta_0^T X_{i,t}) \beta_0 F^T W_{g_0}^+ v_{\beta_0}(X_{i,t}) \varepsilon_{i,t} \\ &\quad - \frac{1}{2} \{E[g'(\beta_0^T X)]^2\}^{-1} W_{g_0}^+ F \beta_0^T W_{g_0}(\beta - \beta_0) \\ &\quad - \frac{1}{2n} \{E[g'(\beta_0^T X)]^2\}^{-1} \sum_{\substack{i \in \mathcal{I}_n \\ t \in \mathcal{T}_n}} g'(\beta_0^T X_{i,t}) W_{g_0}^+ F \beta_0^T v_{\beta_0}(X_{i,t}) \varepsilon_{i,t} \\ &\quad + \frac{1}{2} W_{g_0}^+ W_{g_0}(\beta - \beta_0) + \frac{1}{2n} W_{g_0}^+ \sum_{\substack{i \in \mathcal{I}_n \\ t \in \mathcal{T}_n}} g'(\beta_0^T X_{i,t}) v_{\beta_0}(X_{i,t}) \varepsilon_{i,t} \\ &= \beta_0 + \frac{1}{2} (I - \beta_0 \beta_0^T)(\beta - \beta_0) + \frac{1}{2n} W_{g_0}^+ \sum_{\substack{i \in \mathcal{I}_n \\ t \in \mathcal{T}_n}} g'(\beta_0^T X_{i,t}) v_{\beta_0}(X_{i,t}) \varepsilon_{i,t} + O_p(n^{-\frac{1}{2}} e_d); \end{aligned}$$

the last equation holds because $\beta_0^T v_{\beta_0}(X_{i,t}) = 0$, $\beta_0^T W_{g_0} = 0$, $W_{g_0}^+ W_{g_0} = I - \beta_0 \beta_0^T$, and

$$\frac{1}{2n} F^T W_{g_0}^+ \sum_{\substack{i \in \mathcal{I}_n \\ t \in \mathcal{T}_n}} g'(\beta_0^T X_{i,t}) v_{\beta_0}(X_{i,t}) \varepsilon_{i,t} + \frac{1}{2} F^T W_{g_0}^+ W_{g_0}(\beta - \beta_0) = O_p(e_d).$$

It is easy to check that $\|\beta^{(1)}\| = 1 + o(1)$, so then $\beta^{(1)} / \|\beta^{(1)}\| = \beta^{(1)}$. Let $\beta^{(k)}$ be the value of β after k iterations. We have:

$$\beta^{(k+1)} = \beta_0 + \frac{1}{2} (I - \beta_0 \beta_0^T)(\beta^{(k)} - \beta_0) + \frac{1}{2n} W_{g_0}^+ \sum_{\substack{i \in \mathcal{I}_n \\ t \in \mathcal{T}_n}} g'(\beta_0^T X_{i,t}) v_{\beta_0}(X_{i,t}) \varepsilon_{i,t} + O_p(n^{-\frac{1}{2}} e_d).$$

Recurring the preceding equation, we have, as the iteration $k \rightarrow \infty$:

$$\begin{aligned} \beta^{(k)} &= \beta_0 + \frac{1}{2^{k-1}} (I - \beta_0 \beta_0^T)(\beta^{(1)} - \beta_0) + (1 - \frac{1}{2^{k-1}}) \frac{1}{n} W_{g_0}^+ \sum_{\substack{i \in \mathcal{I}_n \\ t \in \mathcal{T}_n}} g'(\beta_0^T X_{i,t}) v_{\beta_0}(X_{i,t}) \varepsilon_{i,t} \\ &\rightarrow \beta_0 + \frac{1}{n} W_{g_0}^+ \sum_{\substack{i \in \mathcal{I}_n \\ t \in \mathcal{T}_n}} g'(\beta_0^T X_{i,t}) v_{\beta_0}(X_{i,t}) \varepsilon_{i,t}. \end{aligned}$$

The following lemmas are used to prove Theorem 3. \square

Lemma A9. (i) Suppose (A1) holds, and let $\mathcal{L}_1(\mathcal{F})$ denote the class of \mathcal{F} -measurable r.v.'s X satisfying $\|X\|_{l_1} = (E|X|^{l_1})^{\frac{1}{l_1}} < \infty$. Let $X \in \mathcal{L}_1(\mathcal{B}(E))$, $Y \in \mathcal{L}_2(\mathcal{B}(E))$. Suppose $1 \leq l_1, l_2, l_3 < \infty$, and $l_1^{-1} + l_2^{-1} + l_3^{-1} = 1$; then:

$$|E(XY) - EXEY| \leq c\|X\|_{l_1}\|Y\|_{l_2}\{\psi_1(\text{Card}E, \text{Card}E')\varphi_1(d(E, E'))\}^{\frac{1}{3}}. \tag{A37}$$

(ii) If, moreover, $|X|$ and $|Y|$ are $P_{-a.s}$ bounded, the right-hand side of (A37) can be replaced with $c\psi_1(\text{Card}E, \text{Card}E')\varphi_1(d(E, E'))$.

Lemma A10. (i) Suppose (A2) holds, and let $\mathcal{L}_1(\mathcal{F})$ denote the class of \mathcal{F} -measurable r.v.'s X satisfying $\|X\|_{l_1} = (E|X|^{l_1})^{\frac{1}{l_1}} < \infty$. Let $X \in \mathcal{L}_1(\mathcal{B}(G))$, $Y \in \mathcal{L}_2(\mathcal{B}(G))$. Suppose $1 \leq l_1, l_2, l_3 < \infty$, and $l_1^{-1} + l_2^{-1} + l_3^{-1} = 1$; then:

$$|E(XY) - EXEY| \leq c\|X\|_{l_1}\|Y\|_{l_2}\{\psi_2(\text{Card}G, \text{Card}G')\varphi_2(d(G, G'))\}^{\frac{1}{3}}. \tag{A38}$$

(ii) If, moreover, $|X|$ and $|Y|$ are $P_{-a.s}$ bounded, the right-hand side of (A38) can be replaced with $c\psi_2(\text{Card}G, \text{Card}G')\varphi_2(d(G, G'))$.

Lemma A11. Under assumptions (C1) and (C2), one has:

$$E\left[\sqrt{n}\left(\beta^{(k)} - \beta_0\right)\right] = 0. \tag{A39}$$

Proof of Lemma A11.

$$\begin{aligned} E\left[\sqrt{n}\left(\beta^{(k)} - \beta_0\right)\right] &= \sqrt{n}E\left(\beta_0 + \frac{1}{n}W_{g_0}^+ \sum_{\substack{i \in \mathcal{I}_n \\ t \in \mathcal{T}_n}} g'(\beta_0^T X_{i,t})v_{\beta_0}(X_{i,t})\varepsilon_{i,t} - \beta_0\right) \\ &= \frac{1}{\sqrt{n}}EW_{g_0}^+ \sum_{\substack{i \in \mathcal{I}_n \\ t \in \mathcal{T}_n}} g'(\beta_0^T X_{i,t})v_{\beta_0}(X_{i,t})\varepsilon_{i,t} \\ &= 0. \end{aligned}$$

□

Lemma A12. Under assumptions (C1)–(C5), one has:

$$\text{Var}\left[\sqrt{n}\left(\beta^{(k)} - \beta_0\right)\right] = W_{g_0}^+ \Delta W_{g_0}^+, \tag{A40}$$

where $W_{g_0} = E\{g'(\beta_0^T X)^2 W_0(X)\}$ and $\Delta = E\{g'(\beta_0^T X)^2 W_0(X)\sigma_{\beta_0}^2(X)\}$.

Proof of Lemma A12.

$$\begin{aligned} \text{Var}\left[\sqrt{n}\left(\beta^{(k)} - \beta_0\right)\right] &= \frac{1}{n} \sum_{\substack{i \in \mathcal{I}_n \\ t \in \mathcal{T}_n}} \text{Var}\left(W_{g_0}^+ g'(\beta_0^T X_{i,t})v_{\beta_0}(X_{i,t})\varepsilon_{i,t} - \beta_0\right) \\ &\quad + \frac{1}{n} \sum_{\substack{i \neq j \\ t \neq \tau}} EW_{g_0}^+ g'(\beta_0^T X_{i,t})v_{\beta_0}(X_{i,t})\varepsilon_{i,t} W_{g_0}^+ g'(\beta_0^T X_{j,\tau})v_{\beta_0}(X_{j,\tau})\varepsilon_{j,\tau} \\ &= J_1 + J_2, \end{aligned}$$

$$\begin{aligned}
 J_1 &= \frac{1}{n} \sum_{\substack{i \in \mathcal{I}_n \\ t \in T_n}} \text{Var} \left(W_{g_0}^+ g'(\beta_0^T X_{i,t}) v_{\beta_0}(X_{i,t}) \varepsilon_{i,t} \right) \\
 &= \text{Var} \left(W_{g_0}^+ g'(\beta_0^T X_{i,t}) v_{\beta_0}(X_{i,t}) \varepsilon_{i,t} - \beta_0 \right) \\
 &= W_{g_0}^+ E \{ g'(\beta_0^T X)^2 W_0(X) \sigma_{\beta_0}^2(X) \} W_{g_0}^+.
 \end{aligned}$$

By the boundedness of $(W_{g_0}^+ g'(\beta_0^T X_{i,t}) v_{\beta_0}(X_{i,t}))$ for $i \in \mathcal{I}_n, t \in T_n$, and then by the mixing condition, and similar to the proof of $I_{22} \rightarrow 0$ in Lemma 4.4 in [5],

$$\begin{aligned}
 J_2 &= \frac{1}{n} \sum_{\substack{i \neq j \\ t \neq \tau}} E W_{g_0}^+ g'(\beta_0^T X_{i,t}) v_{\beta_0}(X_{i,t}) \varepsilon_{i,t} W_{g_0}^+ g'(\beta_0^T X_{j,\tau}) v_{\beta_0}(X_{j,\tau}) \varepsilon_{j,\tau} \\
 &\leq \frac{c}{n} \sum_{\substack{i \neq j \\ t \neq \tau}} E \varepsilon_{i,t} \varepsilon_{j,\tau} \rightarrow 0.
 \end{aligned}$$

□

Lemma A13. Under assumptions (C1)–(C5), one has:

$$\sqrt{n}(\beta^{(k)} - \beta_0) - E \left[\sqrt{n}(\beta^{(k)} - \beta_0) \right] \xrightarrow{D} N(0, W_{g_0}^+ \Delta W_{g_0}^+). \tag{A41}$$

Proof of Lemma A13. To simplify, let $H_n = W_{g_0}^+ g'(\beta_0^T X_{i,t}) v_{\beta_0}(X_{i,t}) \varepsilon_{i,t}$, $\sqrt{n}(\beta^{(k)} - \beta_0) - E \left[\sqrt{n}(\beta^{(k)} - \beta_0) \right] = \frac{1}{\sqrt{n}} \sum_{\substack{i \in \mathcal{I}_n \\ t \in T_n}} (H_n - EH_n)$. Then, let $\frac{1}{\sqrt{n}} \sum_{\substack{i \in \mathcal{I}_n \\ t \in T_n}} (H_n - EH_n) = \sum_{\substack{i \in \mathcal{I}_n \\ t \in T_n}} \Delta_{i,t}$.

Let us now introduce a space–time block decomposition, which has been used by [30]. Fix integers $p_k = O(\log n_k)^{\frac{1}{8(N+1)}}, k = 1, \dots, N + 1, q = O(\log n_{N+1})^{\frac{1}{16(N+1)}}, k = 1, \dots, N + 1$, and assume that, for some integer r ,

$$n_k = r(p_k + q) \text{ and we have } \lim_{n \rightarrow \infty} \frac{p_k}{q} = \infty, k = 1, \dots, N + 1.$$

The random variables $\Delta_{i,t}$ are now set into blocks of different sizes. Let:

$$\begin{aligned}
 U(1, n, j) &= \sum_{\substack{i_k = j_k(p_k + q) + 1 \\ k = 1, \dots, N}}^{j_k(p_k + q) + p_k} \sum_{t = j_{N+1}(p_{N+1} + q) + 1}^{j_{N+1}(p_{N+1} + q) + p_{N+1}} \Delta_{i,t}, \\
 U(2, n, j) &= \sum_{\substack{i_k = j_k(p_k + q) + 1 \\ k = 1, \dots, N}}^{j_k(p_k + q) + p_k} \sum_{t = j_{N+1}(p_{N+1} + q) + p_{N+1} + 1}^{(j_{N+1} + 1)(p_{N+1} + q)} \Delta_{i,t}, \\
 U(3, n, j) &= \sum_{\substack{i_k = j_k(p_k + q) + 1 \\ k = 1, \dots, N-1}}^{j_k(p_k + q) + p_k} \sum_{i_N = j_N(p_N + q) + p_{N+1}}^{(j_N + 1)(p_{N+1} + q)} \sum_{t = j_{N+1}(p_{N+1} + q) + 1}^{j_{N+1}(p_{N+1} + q) + p_{N+1}} \Delta_{i,t}, \\
 U(4, n, j) &= \sum_{\substack{i_k = j_k(p_k + q) + 1 \\ k = 1, \dots, N-1}}^{j_k(p_k + q) + p_k} \sum_{i_N = j_N(p_N + q) + p_{N+1}}^{(j_N + 1)(p_{N+1} + q)} \sum_{t = j_{N+1}(p_{N+1} + q) + p_{N+1} + 1}^{(j_{N+1} + 1)(p_{N+1} + q)} \Delta_{i,t},
 \end{aligned}$$

and so on. Note that:

$$U(2^{N+1} - 1, n, j) = \sum_{\substack{i_k = j_k(p_k + q) + p_k + 1 \\ k = 1, \dots, N}}^{(j_k + 1)(p_k + q)} \sum_{t = j_{N+1}(p_{N+1} + q) + p_{N+1}}^{j_{N+1}(p_{N+1} + q) + p_{N+1}} \Delta_{i,t}.$$

Finally,

$$U(2^{N+1}, \mathbf{n}, \mathbf{j}) = \sum_{\substack{i_k = j_k(p_k+q) + p_k + 1 \\ k=1, \dots, N}}^{(j_k+1)(p_k+q)} \sum_{t=j_{N+1}(p_{N+1}+q) + p_{N+1} + 1}^{(j_{N+1}+1)(p_{N+1}+q)} \Delta_{i,t}.$$

Setting $\mathcal{R} = \{0, \dots, r-1\} \times \dots \times \{0, \dots, r-1\}$, we define for each integer $i = 1, \dots, 2^{N+1}$,

$$T(\mathbf{n}, i) = \sum_{\mathbf{j} \in \mathcal{R}} U(i, \mathbf{n}, \mathbf{j}).$$

Then, with this notation, we obtain the decomposition

$$\frac{1}{\sqrt{n}} \sum_{\substack{i \in \mathcal{I}_n \\ t \in \mathcal{T}_n}} (H_n - EH_n) = \sum_{i=1}^{2^{N+1}} T(\mathbf{n}, i).$$

Note that $T(\mathbf{n}, 1)$ is the sum of the random variables $\Delta_{i,t}$ over “large” blocks, whereas $T(\mathbf{n}, i), 2 \leq i \leq 2^{N+1}$ are sums over “small” blocks. If it is not the case that $n_k = r(p_k + q)$ for r , then an additional term $T(\mathbf{n}, 2^{N+1} + 1)$, say, containing all the $\Delta_{i,t}$ terms that are not included in the big or small blocks, can be considered. This term will not change the proof much.

The main idea is to show that, as $n \rightarrow \infty$:

$$Q_1 = |E \exp[iuT(\mathbf{n}, 1)] - \prod_{\substack{j_k=0 \\ k=1, \dots, N+1}}^{r-1} E \exp[iuU(1, \mathbf{n}, \mathbf{j})]| \rightarrow 0, \tag{A42}$$

$$Q_2 = \left(\sum_{i=2}^{2^{N+1}} T(\mathbf{n}, i) \right)^2 \rightarrow 0, \tag{A43}$$

$$Q_3 = \sum_{\substack{j_k=0 \\ k=1, \dots, N+1}}^{r-1} E[U(1, \mathbf{n}, \mathbf{j})]^2 \rightarrow W_{g0}^+ \Delta W_{g0}^+, \tag{A44}$$

$$Q_4 = \sum_{\substack{j_k=0 \\ k=1, \dots, N+1}}^{r-1} E[(U(1, \mathbf{n}, \mathbf{j}))^2 I\{|U(1, \mathbf{n}, \mathbf{j})| > \varepsilon W_{g0}^+ \Delta W_{g0}^+\}] \rightarrow 0. \tag{A45}$$

□

Proof of (A42). The proof is similar to the proof of (5.42) in [31]. □

Proof of (A43). According to assumption (C1), if for each $2 \leq i \leq 2^{N+1}$, $E(T(\mathbf{n}, i))^2 \rightarrow 0$ exists, (A43) can be proven. Without loss of generality, it suffices to prove that, for $E[T(\mathbf{n}, 2)]^2 \rightarrow 0$. Enumerate the r.v.'s $U(2, \mathbf{n}, \mathbf{j})$ in an arbitrary manner and refer to them as $\hat{U}_1, \dots, \hat{U}_M$, where $M = r^{N+1} = \binom{n_1}{p_1+q} \dots \binom{n_{N+1}}{p_{N+1}+q}$. □

Now,

$$\begin{aligned} E[T(\mathbf{n}, 2)]^2 &= \sum_{j=1}^M \text{Var}(\hat{U}_j) + 2 \sum_{i=1}^M \sum_{\substack{j=1 \\ i \neq j}}^M \text{Cov}(\hat{U}_i, \hat{U}_j) \\ &= I_1 + I_2. \end{aligned} \tag{A46}$$

If $I_1 \rightarrow 0$, and $I_2 \rightarrow 0$ can be shown, $Q_2 \rightarrow 0$ is obvious.

$$\text{Var} \Delta_{i,t} = \text{Var} \left(\frac{1}{\sqrt{n}} (H_n - EH_n) \right) = \frac{1}{n} E (H_n - EH_n)^2 \tag{A47}$$

$$\begin{aligned} &= \frac{1}{n} E W_{g_0}^+ E \{ g'(\beta_0^T X)^2 W_0(X) \sigma_{\beta_0}^2(X) \} W_{g_0}^+ \\ &\leq \frac{c}{n}. \end{aligned} \tag{A48}$$

Similar to (A47), we have:

$$E \Delta_{i,t} \Delta_{j,\tau} = \frac{1}{n} E W_{g_0}^+ g'(\beta_0^T X_{i,t}) v_{\beta_0}(X_{i,t}) \varepsilon_{i,t} W_{g_0}^+ g'(\beta_0^T X_{j,\tau}) v_{\beta_0}(X_{j,\tau}) \varepsilon_{j,\tau} \tag{A49}$$

$$\leq c \frac{1}{n} \varphi(\|i - j\|, |t - \tau|)^{\frac{\delta}{2+\delta}}. \tag{A50}$$

Then, by Equations (A47) and (A49), we have:

$$\begin{aligned} \text{Var}(\hat{U}_i) &= \text{Var} \left(\sum_{k=1, \dots, N}^{p_k} \sum_{t=1}^q \Delta_{i,t} \right) \\ &\leq \sum_{k=1, \dots, N}^{p_k} \sum_{t=1}^q \text{Var} \Delta_{i,t} + \sum_{\substack{j_k=1 \\ k=1, \dots, N \\ i_k \neq j_k \text{ for some } 1 \leq k \leq N \text{ or } t \neq \tau}}^{p_k} \sum_{\tau=1}^q \sum_{i_k=1}^{p_k} \sum_{t=1}^q |E \Delta_{j,\tau} \Delta_{i,t}| \\ &\leq \sum_{k=1, \dots, N}^{p_k} \sum_{t=1}^q \text{Var} \Delta_{i,t} + \sum_{k=1, \dots, N}^{p_k} \sum_{t=1}^q \sum_{\substack{\tau=1 \\ t \neq \tau}}^q |E \Delta_{i,\tau} \Delta_{i,t}| \\ &\quad + \sum_{\substack{i_k=1 \\ \text{for some } k=1, \dots, N}}^{p_k} \sum_{\substack{j_k \neq i_k \\ \text{for some } k=1, \dots, N}}^{p_k} \sum_{t=1}^q \sum_{\tau=1}^q |E \Delta_{j,\tau} \Delta_{i,t}| + \sum_{t=1}^q \sum_{i_k=1}^{p_k} \sum_{\substack{j_k \neq i_k \\ \text{for some } k=1, \dots, N}}^{p_k} |E \Delta_{j,t} \Delta_{i,t}|. \end{aligned}$$

By Lemma A9, the above inequality continues as:

$$\begin{aligned} &\leq \frac{cp_1 \cdots p_N q}{n} + \frac{cp_1 \cdots p_N q}{n} \left[\sum_{t=1}^{\infty} \varphi_2(t)^{\frac{\delta}{2+\delta}} + \sum_{k=1, \dots, N}^{p_k} \varphi_1(\|i\|) \right] + q \sum_{k=1, \dots, N}^{p_k} \varphi_1(\|i\|) \\ &\leq \frac{cp_1 \cdots p_N q}{n} \left[1 + \sum_{t=1}^{\infty} \varphi_2(t)^{\frac{\delta}{2+\delta}} + q \sum_{i=1}^{\infty} i^N \varphi_1(i)^{\frac{\delta}{2+\delta}} \right] \leq \frac{cp_1 \cdots p_N q^2}{n}. \end{aligned}$$

Consequently,

$$I_1 \leq M \frac{cp_1 \cdots p_N q^2}{n^*} = \frac{n^*}{(p_1 + q) \cdots (p_{N+1} + q)} \cdot \frac{cp_1 \cdots p_N q^2}{n^*} \leq \frac{q^2}{p_{N+1}},$$

which tends to zero by the definition of $p_k = O(\log n_k)^{\frac{1}{8(N+1)}}$ and $q = O(\log n_{N+1})^{\frac{1}{16(N+1)}}$.

Let:

$$I(2, n, j) = \{(i, t) : j_k(p_k + q) + 1 \leq i_k \leq j_k(p_k + q) + p_k, 1 \leq k \leq N, j_{N+1}(p_{N+1} + q) + p_{N+1} + 1 \leq t \leq (j_{N+1} + 1)(p_{N+1} + q)\}.$$

Then, $U(2, n, j)$ is the sum of $\Delta_{i,t}$ with sites in $I(2, n, j)$. Since $p_k > q$, if j and j' belong to the two distinct sets $I(2, n, j)$ and $I(2, n, j')$, then $j_k \neq j'_k$ for some $1 \leq k \leq N + 1$, and $\|j - j'\| \geq q$. To simplify, denoting $p = p_1 \cdots p_N$, then we obtain:

$$I_2 \leq cp^2q^2 \left\{ \sum_{t=\tau} \sum_{\|i-j\| \geq q} |E\Delta_{j,\tau}\Delta_{i,t}| + \sum_{i=j} \sum_{|t-\tau| \geq q} |E\Delta_{j,\tau}\Delta_{i,t}| + \sum_{|t-\tau| \geq q} \sum_{\|i-j\| \geq q} |E\Delta_{j,\tau}\Delta_{i,t}| \right\};$$

by Lemma A9, we have the following equation:

$$\begin{aligned} &\leq \frac{cp^2q^2}{n^*} \left\{ n_{N+1} \sum_{\substack{k=1 \\ kq \leq \|i-j\| < (k+1)q}}^{\infty} \varphi_1(\|i-j\|)^{\frac{\delta}{2+\delta}} + n_1 \cdots n_N \right. \\ &\cdot \sum_{\substack{k=1 \\ kq \leq |t-\tau| < (k+1)q}}^{\infty} \varphi_2(|t-\tau|)^{\frac{\delta}{2+\delta}} + n_{N+1}^2 \sum_{\substack{k=1 \\ kq \leq \|i-j\| < (k+1)q}}^{\infty} \varphi_1(\|i-j\|)^{\frac{\delta}{2+\delta}} \left. \right\} \\ &\leq \frac{cp^2q^2}{n^*} \left[n_1 \cdots n_N \sum_{k=1}^{\infty} \varphi_2(kq)^{\frac{\delta}{2+\delta}} + (n_{N+1} + n_{N+1}^2) \sum_{k=1}^{\infty} k^N \varphi_1(kq)^{\frac{\delta}{2+\delta}} \right] \\ &\leq \frac{cp^2q^2}{n_{N+1}}, \end{aligned}$$

which tends to 0 by the definition of p_k and q . Hence, (A43) holds.

Proof of (A44). Let $S_n = \sum_{i=1}^{2^{N+1}} T(n, i)$, $S'_n = T(n, 1)$ and $S''_n = \sum_{i=2}^{2^{N+1}} T(n, i)$. Then, S'_n is the sum of $\Delta_{i,t}$ terms over the “large” blocks, and S''_n is over the “small” ones. Lemmas A11 and A12 imply that $ES_n^2 \rightarrow W_{g_0}^+ \Delta W_{g_0}^+$; this, combined with (A43), entails $ES'_n{}^2 \rightarrow W_{g_0}^+ \Delta W_{g_0}^+$. Now,

$$ES_n{}^2 = Q_3 + \sum_{\substack{j_k=0 \\ k=1, \dots, N+1 \\ i_k \neq j_k \text{ for}}}^{r-1} \sum_{\substack{i_k=0 \\ k=1, \dots, N+1 \\ \text{some } k}}^{r-1} Cov(U(1, \mathbf{n}, j), U(1, \mathbf{n}, i)). \tag{A51}$$

If the last term of Equation (A51) tends to zero as $n \rightarrow \infty$, then Equation (A44) can be obtained. By the same argument used in obtaining $I_2 \rightarrow 0$, the last term of (A51) is bounded by:

$$\begin{aligned} &c(p_1 \cdots p_{N+1})^2 \left\{ \sum_{t=\tau} \sum_{\|i-j\| \geq q} |E\Delta_{j,\tau}\Delta_{i,t}| + \sum_{i=j} \sum_{|t-\tau| \geq q} |E\Delta_{j,\tau}\Delta_{i,t}| \right. \\ &\left. + \sum_{|t-\tau| \geq q} \sum_{\|i-j\| \geq q} |E\Delta_{j,\tau}\Delta_{i,t}| \right\} \\ &\leq \frac{c(p_1 \cdots p_{N+1})^2}{n} \left[n_1 \cdots n_N \sum_{k=1}^{\infty} \varphi_2(kq)^{\frac{\delta}{2+\delta}} + (n_{N+1} + n_{N+1}^2) \sum_{k=1}^{\infty} k^N \varphi_1(kq)^{\frac{\delta}{2+\delta}} \right] \\ &\leq \frac{c(p_1 \cdots p_{N+1})^2}{n_{N+1}}, \end{aligned}$$

which tends to 0 by the assumptions and the definition of p_k and q . □

Proof of (A45). We need a truncation argument, so set $L = L_n = O(n)^{\frac{1}{5}}$, $H_n = W_{g_0}^+ g'(\beta_0^T X_{i,t}) v_{\beta_0}(X_{i,t}) \varepsilon_{i,t}$, $\Delta_{i,t}^L = \frac{1}{\sqrt{n}} [H_n^L - EH_n^L]$, and define $U^L(1, \mathbf{n}, j) = \sum_{\substack{i_k = j_k(p_k+q)+1 \\ k=1, \dots, N}}^{j_k(p_k+q)+p_k}$

$\sum_{t=j_{N+1}(p_{N+1}+q)+1}^{j_{N+1}(p_{N+1}+q)+p_{N+1}} \Delta_{i,t}^L$. Set

$$Q_4^L = \sum_{\substack{j_k=0 \\ k=1, \dots, N+1}}^{r-1} E[(U^L(1, \mathbf{n}, j))^2 I\{|U^L(1, \mathbf{n}, j)| > \varepsilon W_{g_0}^+ \Delta W_{g_0}^+\}].$$

Clearly, $|\Delta_{i,t}|^L \leq \frac{cL}{(n)^{\frac{1}{2}}}$; therefore, $|U^L(1, \mathbf{n}, \mathbf{j})| \leq \frac{cL(p_1 \cdots p_{N+1})}{(n)^{\frac{1}{2}}}$. Hence,

$$Q_4^L \leq \frac{cL^2(p_1 \cdots p_{N+1})^2}{n} \sum_{\substack{j_k=0 \\ k=1, \dots, N+1}}^{r-1} P(|U^L(1, \mathbf{n}, \mathbf{j})| > \varepsilon \sigma^2 W_{g_0}^+ \Delta W_{g_0}^+).$$

Now, $|U^L(1, \mathbf{n}, \mathbf{j})| \leq \frac{cL(p_1 \cdots p_{N+1})}{(n)^{\frac{1}{2}}} \rightarrow 0$ by the definition of $p_k = O(\log n_k)^{\frac{1}{8(N+1)}}$. Thus, $P(|U^L(1, \mathbf{n}, \mathbf{j})| > \varepsilon \sigma^2 W_{g_0}^+ \Delta W_{g_0}^+) = 0$ at all \mathbf{j} for sufficiently large n , so $Q_4^L = 0$ for large n . Hence,

$$\sum \Delta_{i,t}^L \xrightarrow{\mathcal{L}} \mathcal{N}(\mathbf{0}, \sigma^2 C^T \Sigma C). \tag{A52}$$

Define $S^* = \sum(\Delta_{i,t} - \Delta_{i,t}^L)$, and we know that $\frac{1}{\sqrt{n}}(H_n - EH_n) = S^* + \sum \Delta_{i,t}^L$. By Equation (A52), to prove the lemma, it suffices to prove $ES^{*2} \rightarrow 0$. Similar to Lemma A12, we can show that $ES^{*2} \rightarrow 0$. According to the above lemmas, we have:

$$\sqrt{n}(\boldsymbol{\beta}^{(k)} - \boldsymbol{\beta}_0) - E[\sqrt{n}(\boldsymbol{\beta}^{(k)} - \boldsymbol{\beta}_0)] \xrightarrow{D} N(0, W_{g_0}^+ \Delta W_{g_0}^+). \tag{A53}$$

Additionally, by Lemma A11, we have $E[\sqrt{n}(\boldsymbol{\beta}^{(k)} - \boldsymbol{\beta}_0)] \xrightarrow{P} 0$, and then we obtain $\sqrt{n}(\boldsymbol{\beta}^{(k)} - \boldsymbol{\beta}_0) \xrightarrow{D} N(0, W_{g_0}^+ \Delta W_{g_0}^+)$. \square

Proof of Theorem 4. By Theorems 1 and 2, similar to Lemma A5, under conditions (C1)–(C4), we have:

$$\begin{aligned} & \left\{ n^{-2} \sum_{\substack{j \in \mathcal{I}_n \\ \tau \in \mathcal{T}_n}} \sum_{\substack{i \in \mathcal{I}_n \\ t \in \mathcal{T}_n}} \tilde{\rho}_{j,\tau}^\beta (b_{j,\tau}^\beta)^2 K_{h_n}(\boldsymbol{\beta}^T X_{(i,t),(j,\tau)}) X_{(i,t),(j,\tau)} X_{(i,t),(j,\tau)}^T / \hat{f}_\beta(\boldsymbol{\beta}^T X_{j,\tau}) \right\}^{-1} \\ = & \boldsymbol{\beta}_0 \boldsymbol{\beta}_0^T [E\{g'(\boldsymbol{\beta}_0^T X)^2 / \sigma_{\beta_0}^2(X)\}]^{-1} h_n^{-2} - \frac{1}{2} [E\{g'(\boldsymbol{\beta}_0^T X)^2 / \sigma_{\beta_0}^2(X)\}]^{-1} (\boldsymbol{\beta}_0 \tilde{F}^T \tilde{W}_{g_0}^+ + \tilde{W}_{g_0}^+ \tilde{F} \boldsymbol{\beta}_0^T) \\ & + \frac{1}{2} \tilde{W}_{g_0}^+ + O_P\{(\kappa_n h_n n^\epsilon + \delta_\beta / h_n^2) E_d\}, \end{aligned}$$

where $\tilde{F} = E\{[g'(\boldsymbol{\beta}_0^T X)]^2 / \sigma_{\beta_0}^2(X) (f_{\beta_0}(\boldsymbol{\beta}_0^T X) v_{\beta_0}(X))' / f_{\beta_0}(\boldsymbol{\beta}_0^T X)\}$. Similar to Lemma A6, under conditions (C1)–(C4), we have:

$$\begin{aligned} & n^{-2} \sum_{\substack{j \in \mathcal{I}_n \\ \tau \in \mathcal{T}_n}} \sum_{\substack{i \in \mathcal{I}_n \\ t \in \mathcal{T}_n}} \tilde{\rho}_{j,\tau}^\beta K_{h_n}(\boldsymbol{\beta}^T X_{(i,t),(j,\tau)}) b_{j,\tau}^\beta X_{(i,t),(j,\tau)} (Y_{i,t} - a_{j,\tau}^\beta - b_{j,\tau}^\beta \boldsymbol{\beta}_0^T X_{(i,t),(j,\tau)}) / \hat{f}_\beta(\boldsymbol{\beta}^T X_{j,\tau}) \\ = & \tilde{W}_{g_0}(\boldsymbol{\beta} - \boldsymbol{\beta}_0) + n^{-1} \sum_{\substack{i \in \mathcal{I}_n \\ t \in \mathcal{T}_n}} g'(\boldsymbol{\beta}_0^T X_{i,t}) v_{\beta_0}(X_{i,t}) \varepsilon_{i,t} / \sigma_{\beta_0}^2(X_{i,t}) + o_P(n^{-1/2}). \end{aligned}$$

Similar to Theorem 3, $\boldsymbol{\beta}_0^T \tilde{W}_{g_0} = 0$, $\tilde{W}_{g_0}^+ \tilde{W}_{g_0} = I - \boldsymbol{\beta}_0 \boldsymbol{\beta}_0^T$, recursing the preceding equation, and we have, as the iteration $k \rightarrow \infty$:

$$\begin{aligned} \boldsymbol{\beta}^{(k)} &= \boldsymbol{\beta}_0 + \frac{1}{2^{k-1}} (I - \boldsymbol{\beta}_0 \boldsymbol{\beta}_0^T) (\boldsymbol{\beta}^{(1)} - \boldsymbol{\beta}_0) + (1 - \frac{1}{2^{k-1}}) \frac{1}{n} \tilde{W}_{g_0}^+ \\ &\cdot \sum_{\substack{i \in \mathcal{I}_n \\ t \in \mathcal{T}_n}} g'(\boldsymbol{\beta}_0^T X_{i,t}) v_{\beta_0}(X_{i,t}) \varepsilon_{i,t} / \sigma_{\beta_0}(X_{i,t}) \\ &\rightarrow \boldsymbol{\beta}_0 + \frac{1}{n} \tilde{W}_{g_0}^+ \sum_{\substack{i \in \mathcal{I}_n \\ t \in \mathcal{T}_n}} g'(\boldsymbol{\beta}_0^T X_{i,t}) v_{\beta_0}(X_{i,t}) / \sigma_{\beta_0}(X_{i,t}) \cdot \varepsilon_{i,t} / \sigma_{\beta_0}(X_{i,t}). \end{aligned}$$

Similar to the proof of Theorem 3, as $n \rightarrow \infty$, Theorem 4 holds. \square

Proof of Theorem 5. Set $\eta_{i,t} = g'(\beta_0^T X_{i,t})v_{\beta_0}(X_{i,t})$, $\Sigma = \text{diag}\{\sigma_{\beta_0}^2(X_1), \dots, \sigma_{\beta_0}^2(X_n)\}$, $\eta = (\eta_1^T, \eta_2^T, \dots, \eta_n^T)$. According to the definitions of W_{g0} , Δ , and \tilde{W}_{g0} , we have:

$$\tilde{W}_{g0}^+ = \lim_{n \rightarrow +\infty} \left(\frac{1}{n} \sum_{\substack{i \in \mathcal{I}_n \\ t \in \mathcal{T}_n}} \eta_{i,t} \eta_{i,t}^T \sigma_{\beta_0}^{-2}(X_n) \right)^+ = \lim_{n \rightarrow +\infty} n (\eta^T \Sigma^{-1} \eta)^+,$$

and

$$\begin{aligned} W_{g0}^+ \Delta W_{g0}^+ &= \lim_{n \rightarrow +\infty} \left(\frac{1}{n} \sum_{\substack{i \in \mathcal{I}_n \\ t \in \mathcal{T}_n}} \eta_{i,t} \eta_{i,t}^T \right)^+ \lim_{n \rightarrow +\infty} \left(\frac{1}{n} \sum_{\substack{i \in \mathcal{I}_n \\ t \in \mathcal{T}_n}} \eta_{i,t} \eta_{i,t}^T \sigma_{\beta_0}^2(X_n) \right) \lim_{n \rightarrow +\infty} \left(\frac{1}{n} \sum_{\substack{i \in \mathcal{I}_n \\ t \in \mathcal{T}_n}} \eta_{i,t} \eta_{i,t}^T \right)^+ \\ &= n (\eta^T \eta)^+ (\eta^T \Sigma \eta) (\eta^T \eta)^+. \end{aligned}$$

For any d -dimensional vector $c \in R^d$, let $\|c\|_{\Sigma} = c^T \Sigma c$. Since Σ is a positive matrix, then $\|\cdot\|_{\Sigma}$ is a norm of space R^p . Therefore, we have:

$$\begin{aligned} c^T (\eta^T \eta)^+ (\eta^T \Sigma \eta) (\eta^T \eta)^+ c &= \|\eta (\eta^T \eta)^+ c\|_{\Sigma} \\ &= \|\eta (\eta^T \eta)^+ c - \Sigma^{-1} \eta (\eta^T \Sigma^{-1} \eta)^+ c + \Sigma^{-1} \eta (\eta^T \Sigma^{-1} \eta)^+ c\|_{\Sigma} \\ &= \|\eta (\eta^T \eta)^+ c - \Sigma^{-1} \eta (\eta^T \Sigma^{-1} \eta)^+ c\|_{\Sigma} + \|\Sigma^{-1} \eta (\eta^T \Sigma^{-1} \eta)^+ c\|_{\Sigma} \\ &\quad + 2c^T (\eta^T \Sigma^{-1} \eta)^+ \eta^T \Sigma^{-1} \Sigma \left[\eta (\eta^T \eta)^+ c - \Sigma^{-1} \eta (\eta^T \Sigma^{-1} \eta)^+ c \right] \\ &= \|\eta (\eta^T \eta)^+ c - \Sigma^{-1} \eta (\eta^T \Sigma^{-1} \eta)^+ c\|_{\Sigma} + \|\Sigma^{-1} \eta (\eta^T \Sigma^{-1} \eta)^+ c\|_{\Sigma} \\ &\geq \|\Sigma^{-1} \eta (\eta^T \Sigma^{-1} \eta)^+ c\|_{\Sigma} = c^T (\eta^T \Sigma^{-1} \eta)^+ c. \end{aligned}$$

By the universality of the column vector c , we can obtain Theorem 5. \square

Proof of Theorem 6. For simplicity, set $v = \beta_0^T x$, by Lemma A4 ; then, we have:

$$\hat{g}_n(v) = a_x^{\beta_0} = g(v) + \frac{1}{2} g''(v) h_n^2 + \Theta_{n,1}(x) + O_P\{(h_n \kappa_n + \delta_{\beta}^2) n^{\epsilon}\} (1 + \|x\|^4).$$

Thus,

$$\sqrt{nh_n} \left\{ \hat{g}_n(v) - g(v) - \frac{h_n^2}{2} g''(v) + O_p(h_n^2) \right\} = \sqrt{h_n} (\sqrt{nf_{\beta_0(v)}})^{-1} \sum_{\substack{i \in \mathcal{I}_n \\ t \in \mathcal{T}_n}} K_{h_n}(\beta_0^T X_{(i,t)x}) \varepsilon_{i,t}.$$

After some simple computations, by Lemma A1, we have:

$$E[\sqrt{h_n} (\sqrt{nf_{\beta_0(v)}})^{-1} \sum_{\substack{i \in \mathcal{I}_n \\ t \in \mathcal{T}_n}} K_{h_n}(\beta_0^T X_{(i,t)x}) \varepsilon_{i,t}] = 0,$$

$$\text{Var}[\sqrt{h_n} (\sqrt{nf_{\beta_0(v)}})^{-1} \sum_{\substack{i \in \mathcal{I}_n \\ t \in \mathcal{T}_n}} K_{h_n}(\beta_0^T X_{(i,t)x}) \varepsilon_{i,t}] = \frac{\sigma_{\beta_0}^2(x)}{f_{\beta_0(v)}} + o(1).$$

Similar to the proof of Theorem 3, as $n \rightarrow \infty$, Theorem 6 holds. \square

References

1. Fotheringham, A.S.; Yang, W.; Kang, W. Multiscale geographically weighted regression (MGWR). *Ann. Am. Assoc. Geogr.* **2017**, *107*, 1247–1265. [CrossRef]
2. Wu, C.; Ren, F.; Hu, W.; Du, Q. Multiscale geographically and temporally weighted regression: Exploring the spatio-temporal determinants of housing prices. *Int. J. Geogr. Inf. Sci.* **2019**, *33*, 489–511. [CrossRef]

3. Chu, H.J.; Huang, B.; Lin, C.Y. Modeling the spatio-temporal heterogeneity in the PM_{10} - $PM_{2.5}$ relationship. *Atmos. Environ.* **2015**, *102*, 176–182. [[CrossRef](#)]
4. Yuan, Z.; Zhou, X.; Yang, T. Hetero-ConvLSTM: A deep learning approach to traffic accident prediction on heterogeneous spatio-temporal data. In Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, London, UK, 19–23 August 2018; pp. 984–992.
5. Wang, H.; Wang, J.; Huang, B. Prediction for spatio-temporal models with autoregression in errors. *J. Nonparametric Stat.* **2012**, *24*, 1–28. [[CrossRef](#)]
6. Wang, H.; Zhao, Z.; Wu, Y.; Luo, X. B-spline method for spatio-temporal inverse model. *J. Syst. Sci. Complex.* **2022**, *35*, 2336–2360. [[CrossRef](#)]
7. Římalová, V.; Fišerová, E.; Menafogliob, A.; Pinic, A. Inference for spatial regression models with functional response using a permutational approach. *J. Multivar. Anal.* **2022**, *189*, 104893. [[CrossRef](#)]
8. Lin, W.; Kulasekera, K.B. Identifiability of single-index models and additive-index models. *Biometrika* **2007**, *94*, 496–501. [[CrossRef](#)]
9. Stoker, T.M. Consistent estimation of scaled coefficients. *Econometrica* **1986**, *54*, 1461–1481. [[CrossRef](#)]
10. Powell, J.L.; Stock, J.H.; Stoker, T.M. Semiparametric estimation of index coefficients. *Econometrica* **1989**, *57*, 1403–1430. [[CrossRef](#)]
11. Xia, Y. Asymptotic distributions for two estimators of the single-index model. *Econom. Theory* **2006**, *22*, 1112–1137. [[CrossRef](#)]
12. Xia, Y. A constructive approach to the estimation of dimension reduction directions. *Ann. Stat.* **2007**, *35*, 2654–2690. [[CrossRef](#)]
13. Härdle, W.; Hall, H.; Ichimura, H. Optimal smoothing in single-index models. *Ann. Stat.* **1993**, *21*, 157–178. [[CrossRef](#)]
14. Li, G.; Peng, H.; Dong, K.; Tong, T. Simultaneous confidence bands and hypothesis testing for single-index models. *Stat. Sin.* **2014**, *24*, 937–955. [[CrossRef](#)]
15. Fan, Y.; James, G.M.; Radchenko, P. Functional additive regression. *Ann. Stat.* **2015**, *43*, 2296–2325. [[CrossRef](#)]
16. Xue, L. Estimation and empirical likelihood for single-index models with missing data in the covariates. *Comput. Stat. Data Anal.* **2013**, *52*, 1458–1476. [[CrossRef](#)]
17. Xue, L.; Zhu, L. Empirical likelihood for single-index model. *J. Multivar. Anal.* **2006**, *97*, 1295–1312. [[CrossRef](#)]
18. Cook, R.D.; Li, B. Dimension reduction for conditional mean in regression. *Ann. Stat.* **2002**, *30*, 455–474. [[CrossRef](#)]
19. Zhu, L.; Ferré, L.; Wang, T. Sufficient dimension reduction through discretization-expectation estimation. *Biometrika* **2010**, *97*, 295–304. [[CrossRef](#)]
20. Ma, Y.; Zhu, L. A review on dimension reduction. *Int. Stat. Rev.* **2013**, *81*, 134–150. [[CrossRef](#)]
21. Zhao, Y.; Li, J.; Wang, H.; Zhao, H.; Chen, X. Efficient estimation in heteroscedastic single-index models. *J. Nonparametric Stat.* **2021**, *33*, 273–298. [[CrossRef](#)]
22. Horowitz, J.L.; Härdle, W. Direct semiparametric estimation of single-index models with discrete covariates. *J. Am. Stat. Assoc.* **1996**, *91*, 1632–1640. [[CrossRef](#)]
23. Ichimura, H. Semiparametric least squares (SLS) and weighted SLS estimation of single-index models. *J. Econom.* **1993**, *58*, 71–120. [[CrossRef](#)]
24. Cressie, N.A.C. *Statistics for Spatial Data*; John Wiley & Sons: New York, NY, USA, 1993.
25. Mack, Y.P.; Silverman, B.W. Weak and strong uniform consistency of kernel regression estimates. *Zeitschrift für Wahrscheinlichkeitstheorie Und Verwandte Geb.* **1982**, *61*, 405–415. [[CrossRef](#)]
26. Liebscher, E. Strong convergence of sums of α -mixing random variables with applications to density estimation. *Stoch. Process. Their Appl.* **1996**, *65*, 69–80. [[CrossRef](#)]
27. Schott, J.R. *Matrix Analysis for Statistics*; Wiley: Hoboken, NJ, USA, 1997.
28. Chow, Y.S.; Teicher, H. *Probability Theory: Independence, Interchangeability, Martingales*; Springer: Berlin/Heidelberg, Germany, 1978; p. 122.
29. Lu, Z.; Arvid, L.; Dag, T.; Yao, Q. Exploring spatial nonlinearity using additive approximation. *Bernoulli* **2007**, *13*, 447–472. [[CrossRef](#)]
30. Wang, H.; Wang, J. Estimation of the trend function for spatio-temporal models. *J. Nonparametric Stat.* **2009**, *21*, 567–588. [[CrossRef](#)]
31. Lu, Z.; Linton, O. Local linear fitting under near epoch dependence. *Econ. Theory* **2007**, *23*, 37–70. [[CrossRef](#)]

Disclaimer/Publisher’s Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.