A Semiparametric Spatial Dynamic Model *

Yan Sun
School of Economics
Shanghai University of Finance and Economics, P. R. China

Hongjia Yan
Department of Mathematics
The University of York, UK

Wenyang Zhang †
Department of Mathematics
The University of York, UK

Zudi Lu
School of Mathematical Sciences
The University of Adelaide, Australia

August 26, 2013

Abstract

Stimulated by the Boston house price data, in this paper, we propose a semiparametric spatial dynamic model, which extends the ordinary spatial autoregressive models to accommodate the effects of some covariates associated with the house price. A profile likelihood based estimation procedure is proposed. The asymptotic normality of the proposed estimators are derived. We also investigate how to identify the parametric/nonparametric components in the proposed semiparametric model. We show how many unknown parameters an unknown bivariate function amounts to, and propose an AIC/BIC of nonparametric version for model selection. Simulation

*Supported by National Science Foundation of China (Grant 11271242) and the Singapore National Research Foundation under its Cooperative Basic Research Grant and administered by the Singapore Ministry of Healths National Medical Research Council (Grant No. NMRC/CBRG/0014/2012).

†Correspondent author. Email: wenyang.zhang@york.ac.uk

studies are conducted to examine the performance of the proposed methods. The simulation results show our methods work very well. We finally apply the proposed methods to analyse the Boston house price data, which leads to some interesting findings.

KEY WORDS: AIC/BIC, local linear modelling, profile likelihood, spatial interaction.

SHORT TITLE: SSDM.

1 Introduction

The Boston house price data is frequently used in literature to illustrate some new statistical methods. If we use y_i to denote the median value of owner-occupied homes at location s_i, a spatial autoregressive model for the data would be

$$y_i = \sum_{j \neq i} w_{ij} y_j + \epsilon_i, \quad i = 1, \cdots, n, \tag{1.1}$$

where w_{ij} is the impact of y_j on y_i. Apparently, (1.1) is not adequate to address what affect the house price and how. It is better to incorporate the effects of some important covariates, such as crime rate and accessibility to radial highways, into the model. Let X_i, a p dimensional vector, be the vector of the covariates associated with y_i, a reasonable model to fit the data would be

$$y_i = \sum_{j \neq i} w_{ij} y_j + X_i^{\mathrm{T}} \beta + \epsilon_i, \quad i = 1, \cdots, n. \tag{1.2}$$

where w_{ij} and β are unknown. However, there are two problems with model (1.2): first, there are too many unknown parameters; second, the model has not taken into account the location effects of the impacts of the covariates – the impacts of some covariates may vary over location. To control the number of unknown parameters and take the location effects into account, we propose the following model to fit the data

$$y_i = \alpha \sum_{j \neq i} w_{ij} y_j + X_i^{\mathrm{T}} \beta(s_i) + \epsilon_i, \quad i = 1, \cdots, n, \tag{1.3}$$

where w_{ij} is a specified certain physical or economic distance, $\beta(\cdot) = (\beta_1(\cdot), \cdots, \beta_p(\cdot))^{\mathrm{T}}$, ϵ_i, $i = 1, \cdots, n$, are i.i.d., and follow $N(0, \sigma^2)$, $\{X_i, i = 1, \cdots, n\}$ is independent of $\{\epsilon_i, i = 1, \cdots, n\}$. α, σ^2 and $\beta(\cdot)$ are unknown to be estimated. Model (1.3) is the model this paper is going to address. From now on, y_i is of course not necessarily the house price, it is a generic response variable.

In model (1.3), the spatial neighbouring effect of y_j, $j \neq i$, on y_i is formulated through αw_{ij}, where w_{ij} is a specified certain physical or economic distance. Such method to define spatial neighbouring effect is common, see Ord (1975), Anselin (1988), Su and Jin (2010).

Model (1.3) is an useful extension of spatial autoregressive models (Gao *et al.*, 2006; Kelejian and Prucha, 2010; Ord, 1975; Su and Jin, 2010) and varying coefficient models (Cheng *et al.*, 2009; Fan and Zhang, 1999, 2000; Li and Zhang, 2011; Sun *et al.*, 2007; Zhang *et al.*, 2002; Zhang *et al.*, 2009; Wang and Xia, 2009; and Tao and Xia, 2011). One characteristic of model (1.3) is

$$E(\epsilon_i | y_1, \cdots, y_{i-1}, y_{i+1}, \cdots, y_n) \neq 0$$

although $E(\epsilon_i) = 0$, the standard least squares estimation will not work for (1.3). In this paper, based on the local linear modelling and profile likelihood idea, we will propose a local likelihood based estimation procedure for the unknown parameters and functions in (1.3) and derive the asymptotic properties of the obtained estimators.

In reality, some components of $\beta(\cdot)$ in model (1.3) may be constant, and we don't know which components are functional, which are constant. Methodologically speaking, if mistakenly treating a constant component as functional, we would pay a price on the variance side of the obtained estimator, on the other hand, if mistakenly treating a functional component as constant, we would pay a price on the bias side of the obtained estimator. The identification of constant/functional components in $\beta(\cdot)$ is imperative. From practical point of view, the identification of constant components is also of importance. For the data set we study in this paper, $\beta(\cdot)$ can be interpreted as the vector of the impacts of the covariates concerned on the house price. The identification will reveal which covariates have location varying impacts on the house price, which do not. This is apparently something of great interest. In this paper, we will show how many unknown parameters an unknown bivariate function amounts to, and propose an AIC/BIC of nonparametric version to identify the constant components of $\beta(\cdot)$ in model (1.3).

The paper is organised as follows. We begin in Section 2 with a description of the estimation procedure for the proposed model (1.3). In Section 3 we show how many unknown parameters an uknown bivariate function amounts to, and propose an AIC/BIC of nonparametric version for model selection. Asymptotic properties of the proposed estimators are presented in Section 4. The performance of the proposed methods, including both estimation and model selction methods, is assessed by a simulation study in Section 5. In Section 6, we explore how the covariates, which are commonly found to be associated with house price, affect the median value of owner-occupied homes in Boston, and how the impacts of these covariates change over location based on the proposed model and estimation procedure.

Throughout this paper, $\mathbf{0}_k$ is a k dimensional vector with each component being 0, I_k is an identity matrix of size k, $U[0,\ 1]^2$ is a two dimensional uniform distribution on $[0,\ 1] \times [0,\ 1]$.

2 Estimation procedure

Let $w_{ii} = 0$, $W = (w_{ij})$, $Y = (y_1,\ \cdots,\ y_n)^{\mathrm{T}}$, $A = I_n - \alpha W$, and $\mathbf{m} = \left(X_1^{\mathrm{T}}\beta(s_1),\ \cdots,\ X_n^{\mathrm{T}}\beta(s_n)\right)^{\mathrm{T}}$. By simple calculation, we have that the conditional density function of Y given $\mathbf{m}$ is $N\left(A^{-1}\mathbf{m},\ (A^{\mathrm{T}}A)^{-1}\sigma^2\right)$, which leads to the following log likelihood function

$$-\frac{n}{2}\log(2\pi) - n\log(\sigma) + \log(|A|) - \frac{1}{2\sigma^2}(AY - \mathbf{m})^{\mathrm{T}}(AY - \mathbf{m}). \qquad (2.1)$$

Our estimation is profile likelihood based. We first construct the estimator $\tilde{\beta}(\cdot;\ \alpha)$ of $\beta(\cdot)$ pretending α is known, then let $(\hat{\alpha},\ \hat{\sigma}^2)$ maximise (2.1) with $\beta(\cdot)$ being replaced by $\tilde{\beta}(\cdot;\ \alpha)$. $\hat{\alpha}$ and $\hat{\sigma}^2$ are our estimators of α and σ^2, respectively. After the estimator of α is obtained, the estimator of $\beta(\cdot)$ is taken to be $\tilde{\beta}(\cdot;\ \alpha)$ with α and the bandwidth used being replaced by $\hat{\alpha}$ and a slightly larger bandwidth, respectively. The details are as follows.

For any $s = (u,\ v)^{\mathrm{T}}$, we denote $(\partial\beta(s)/\partial u,\ \partial\beta(s)/\partial v)$ by $\dot{\beta}(s)$, where $\partial\beta(s)/\partial u = (\partial\beta_1(s)/\partial u,\ \cdots,\ \partial\beta_p(s)/\partial u)^{\mathrm{T}}$.

For any given s, by the Taylor's expansion, we have

$$\beta(s_i) \approx \beta(s) + \dot{\beta}(s)(s_i - s)$$

when s_i is in a small neighbourhood of s, which leads to the following objective function for estimating $\beta(s)$

$$\sum_{i=1}^{n} \left(y_i^* - X_i^{\mathrm{T}}\mathbf{a} - X_i^{\mathrm{T}}\mathbf{B}(s_i - s)\right)^2 K_h(\|s_i - s\|), \tag{2.2}$$

where y_i^* is the ith component of AY, $K_h(\cdot) = K(\cdot/h)/h^2$, $K(\cdot)$ is a kernel function, and h is a bandwidth. Let $(\hat{\mathbf{a}},\ \hat{\mathbf{B}})$ minimise (2.2), the 'estimator' $\tilde{\beta}(s;\ \alpha)$ of $\beta(s)$ is taken to be $\hat{\mathbf{a}}$. By simple calculations, we have

$$\tilde{\beta}(s;\ \alpha) = \hat{\mathbf{a}} = (I_p,\ 0_{p\times 2p})\left(\mathcal{X}^{\mathrm{T}}\mathcal{W}\mathcal{X}\right)^{-1}\mathcal{X}^{\mathrm{T}}\mathcal{W}AY, \tag{2.3}$$

where $0_{p\times q}$ is a matrix of size $p \times q$ with each entry being 0, and

$$\mathcal{X} = \begin{pmatrix} X_1 & \cdots & X_n \\ X_1 \otimes (s_1 - s) & \cdots & X_n \otimes (s_n - s) \end{pmatrix}^{\mathrm{T}},$$

$$\mathcal{W} = \mathrm{diag}\left(K_h(\|s_1 - s\|),\ \cdots,\ K_h(\|s_n - s\|)\right).$$

Replacing $\beta(s_i)$ in (2.1) by $\tilde{\beta}(s_i;\ \alpha)$ and ignoring the constant term, we have the objective function for estimating α and σ^2

$$-n\log(\sigma) + \log(|A|) - \frac{1}{2\sigma^2}(AY - \tilde{\mathbf{m}})^{\mathrm{T}}(AY - \tilde{\mathbf{m}}), \tag{2.4}$$

where $\tilde{\mathbf{m}}$ is $\mathbf{m}$ with $\beta(s_i)$ being replaced by $\tilde{\beta}(s_i;\ \alpha)$. Let α_i, $i = 1,\ \cdots,\ n$, be the eigenvalues of W,

$$\tilde{\sigma}^2 = \frac{1}{n}(AY - \tilde{\mathbf{m}})^{\mathrm{T}}(AY - \tilde{\mathbf{m}}),$$ with $\tilde{m}=0$, this is (24) of Ord, 1975

and $(\hat{\alpha},\ \hat{\sigma}^2)$ maximise (2.4). Noticing that $|A| = \prod_{i=1}^{n}(1 - \alpha\alpha_i)$, by simple calculation, we have $\hat{\alpha}$ is the maximiser of

$$-n\log(\tilde{\sigma}) + \sum_{i=1}^{n}\log(|1 - \alpha\alpha_i|), \tag{2.5}$$

↑ eq. 2.7 in Ord (1975)

and $\hat{\sigma}^2$ is $\tilde{\sigma}^2$ with α being replaced by $\hat{\alpha}$.

To maximise (2.5) is not a big deal as it is an one dimensional optimation problem, we can use grid point method to solve it.

The estimator $\hat{\beta}(\cdot)\ \left(= (\hat{\beta}_1(\cdot),\ \cdots,\ \hat{\beta}_p(\cdot))^{\mathrm{T}}\right)$ is $\tilde{\beta}(\cdot;\ \alpha)$ with α being replaced by $\hat{\alpha}$ and the bandwidth h by a slightly larger bandwidth h_1. The reason to replace the bandwidth h by a slightly larger one is that h is for the estimation of constant parameters such as α and σ^2, it is usually smaller than the one for estimation of functional parameters, this is because undersmooth is needed for the estimators of constant parameters to achieve the optimal convergence rate.

In reality, some components of $\beta(\cdot)$ may be constant. If a component of $\beta(\cdot)$ is a constant, say $\beta_1(\cdot) = \beta_1$, we use the average of $\hat{\beta}_1(s_i)$, $i = 1,\ \cdots,\ n$, to estimate the constant β_1, that is

$$\hat{\beta}_1 = \frac{1}{n}\sum_{i=1}^{n}\hat{\beta}_1(s_i).$$

How to identify the constant components of $\beta(\cdot)$ will be addressed in next section.

3 Identification of constant components

3.1 Criterion for identification

As we mentioned before, some components of $\beta(\cdot)$ in model (1.3) may be constant in reality, and to identify such constant components is of importance. In this paper, we appeal the AIC or BIC to identify the constant components. The AIC for (1.3), in which some components of $\beta(\cdot)$ may be constant, is defined as follows

$$\mathrm{AIC} = n\log(\hat{\sigma}) - \log(|\hat{A}|) + \frac{1}{2\hat{\sigma}^2}(\hat{A}Y - \hat{\mathbf{m}})^{\mathrm{T}}(\hat{A}Y - \hat{\mathbf{m}}) + \mathcal{K}, \qquad (3.1)$$

where $\hat{A}$ and $\hat{\mathbf{m}}$ are A and $\mathbf{m}$ with the unknown parameters and functions being replaced by their estimators, $\mathcal{K}$ is the number of unknown parameters in model (1.3). The BIC can be defined in a similar way.

Because there are unknown functions in model (1.3), the first hurdle in the calculation of AIC of model (1.3) is to find how many unknown constants an unknown bivariate function amounts to. In the following, based on the residual sum of squares of standard bivariate nonparametric regression model, we propose an ad hoc way to solve this problem.

Suppose we have the following standard bivariate nonparametric regression model,

$$\eta_i = g(s_i) + e_i, \quad i = 1, \cdots, n, \tag{3.2}$$

where $E(e_i) = 0$ and $\mathrm{var}(e_i) = \sigma_e^2$. The residual sum of squares of (3.2) is

$$\mathrm{RSS} = \sum_{i=1}^{n} \{\eta_i - \hat{g}(s_i)\}^2$$

where $\hat{g}(\cdot)$ is the local linear estimator of $g(\cdot)$. On the other hand,

$E(\mathrm{RSS}/\sigma_e^2) = n-$ the number of unknown parameters in the regression function

So, the number $\mathcal{T}$ of unknown constants the unknown function $g(\cdot)$ amounts to can be reasonably viewed as

$$\mathcal{T} = n - E(\mathrm{RSS}/\sigma_e^2) = n - \sigma_e^{-2} E\left[\sum_{i=1}^{n} \{\eta_i - \hat{g}(s_i)\}^2\right].$$

To make $\mathcal{T}$ more convenient to use, we derive the asymptotic form of $\mathcal{T}$. Let

$$\mathbf{S}_i = \begin{pmatrix} 1 & s_1^{\mathrm{T}} - s_i^{\mathrm{T}} \\ \vdots & \vdots \\ 1 & s_n^{\mathrm{T}} - s_i^{\mathrm{T}} \end{pmatrix}, \quad \boldsymbol{\eta} = \begin{pmatrix} \eta_1 \\ \vdots \\ \eta_n \end{pmatrix}, \quad \mathbf{e} = \begin{pmatrix} e_1 \\ \vdots \\ e_n \end{pmatrix},$$

and

$$\mathcal{W}_i = \mathrm{diag}\left(K_h(u_1 - u_i)K_h(v_1 - v_i), \cdots, K_h(u_n - u_i)K_h(v_n - v_i)\right),$$

we have

$$\hat{g}(s_i) = (1, 0, 0)\left(\mathbf{S}_i^{\mathrm{T}} \mathcal{W}_i \mathbf{S}_i\right)^{-1} \mathbf{S}_i^{\mathrm{T}} \mathcal{W}_i \boldsymbol{\eta}$$

By the standard argument in Fan and Gijbels (1996) and the Lemma 1 in Fan and Zhang (1999), we have

$$\mathcal{T} = \left(2K^2(0) - \nu_*^2\right) h^{-2} + o(h^{-2})$$

when $h = o(n^{-1/6})$ and $nh^2 \longrightarrow \infty$, where $\nu_* = \int K^2(t)dt$.

We conclude that an unknown bivariate function amounts to $(2K^2(0) - \nu_*^2)\, h^{-2}$ unknown constants. Based on this conclusion, if the number of constant components in $\beta(\cdot)$ is q, the $\mathcal{K}$ in (3.1) will be $q + (p-q)\,(2K^2(0) - \nu_*^2)\, h^{-2}$.

To identify the constant components in $\beta(\cdot)$ in (1.3) is basically a model selection problem. Theoretically speaking, we go for the model with the smallest AIC (or BIC). However, in practice, it is almost computationally impossible to compute the AICs for all possible models. We have to use some algorithm to reduce the computational burden. In the following, we are going to introduce two algorithms for the model selection.

3.2 Computational algorithms

In this section, we use AIC as an example to demonstrate the introduced algorithms. The model in which $\beta(\cdot)$ has its i_1th, i_2th, $\cdots$, and i_kth components being constant is denoted by $\{i_1, \cdots, i_k\}$. When $k = 0$ we define the model as the model in which all components of $\beta(\cdot)$ are functional, and denote it by $\{\}$.

Backward elimination

The first algorithm we introduce is the backward elimination. Details are as follows.

(1) We start with the full model, $\{1, \cdots, p\}$, and compute its AIC by (3.1). Denote the full model by $\mathcal{M}_p$, its AIC by AIC_p.

(2) For any integer k, suppose the current model is $\mathcal{M}_k = \{i_1, \cdots, i_k\}$ with AIC given by AIC_k. Take $\mathcal{M}_{k-1}$ to be the model with the largest maximum of log likelihood function among the models $\{i_1, \cdots, i_{j-1}, i_{j+1}, \cdots, i_k\}$, $j = 1, \cdots, k$. If $\text{AIC}_k < \text{AIC}_{k-1}$, the chosen model is $\mathcal{M}_k$, and the model selection is ended; otherwise, continue to compute $\mathcal{M}_l$ and AIC_l until either $\text{AIC}_l < \text{AIC}_{l-1}$ or $l = 0$.

Curvature-to-Average ratio (CTAR) based method

A more aggressive way to reduce the computational burden involved in the model selection procedure is based on the ratio of the curvature of the estimated function to its average. Explicitly, we first treat all $\beta_j(\cdot)$, $j = 1, \cdots, p$, as functional. For each j, $j = 1, \cdots, p$, we compute the curvature-to-average ratio (CTAR) R_j of the estimated function $\hat{\beta}_j(\cdot)$:

$$R_j = \frac{1}{\bar{\beta}_j^2} \sum_{i=1}^{n} \left\{ \hat{\beta}_j(s_i) - \bar{\beta}_j \right\}^2, \quad \bar{\beta}_j = \frac{1}{n} \sum_{i=1}^{n} \hat{\beta}_j(s_i), \quad j = 1, \cdots, p.$$

We sort R_j, $j = 1, \cdots, p$, in an increasing order, say $R_{i_1} \leq \cdots \leq R_{i_p}$, then compute the AICs for the models $\{i_1, \cdots, i_k\}$ from $k = 0$ to the turning point k_0 where the AIC starts to increase. The chosen model is $\{i_1, \cdots, i_{k_0}\}$.

The algorithm based on the CTAR is much faster than the backward elimination based algorithm, however, from simulations, we find it less accurate although it still works reasonably well.

4 Asymptotic properties

In this section, we are going to present the asymptotic properties of the proposed estimators. We will, in this section, only present the asymptotic results, and leave the theoretical proofs in the Appendix.

Although we assume ϵ_i in (1.3) follows normal distribution in our model assumption, we do not need this assumption when deriving the asymptotic properties of the proposed estimators. So, in this section, we do not assume ϵ_i follows normal distribution unless otherwise stated.

In this section, for w_{ij} in (1.3), we assume that there exists a sequence $\rho_n > 0$ such that $w_{ij} = O(1/\rho_n)$ uniformly with respect to i, j and the matrices W and A^{-1} are uniformly bounded in both row and column sums.

We now introduce some notations needed in the presentation of the asymptotic properties of the proposed estimators: Let $\mu_j = E\epsilon_1^j$,

$$\kappa_0 = \int_{R^2} K(\|s\|)ds, \quad \kappa_2 = \int_{R^2} [(1,0)s]^2 K(\|s\|)ds = \int_{R^2} [(0,1)s]^2 K(\|s\|)ds,$$

$$\nu_0 = \int_{R^2} K^2(\|s\|)ds, \quad \nu_2 = \int_{R^2} [(1,0)s]^2 K^2(\|s\|)ds = \int_{R^2} [(0,1)s]^2 K^2(\|s\|)ds,$$

$$G = (g_{ij}) = WA^{-1}, \quad \Psi = E(X_1 X_1^{\mathrm{T}}), \ \Gamma = EX_1,$$

$$Z_1(s) = \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} g_{ii} \beta(s_i) K_h(\|s_i - s\|),$$

$$Z_2(s) = \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} \sum_{j\neq i}^{n} g_{ij} \beta(s_j) K_h(\|s_i - s\|), \quad Z(s) = Z_1(s) + \Psi^{-1}\Gamma\Gamma^{\mathrm{T}} Z_2(s),$$

$$Z = \kappa_0^{-1}\Big(f^{-1}(s_1)X_1^{\mathrm{T}} Z(s_1), \cdots, f^{-1}(s_n)X_n^{\mathrm{T}} Z(s_n)\Big)^{\mathrm{T}},$$

$$\pi_1 = \lim_{n\to\infty} \frac{\mathrm{tr}((G+G^{\mathrm{T}})G)}{n}, \quad \pi_2 = \lim_{n\to\infty} \frac{\mathrm{tr}(G)}{n}, \quad \pi_3 = \lim_{n\to\infty} \frac{1}{n} \sum_{i=1}^{n} g_{ii}^2,$$

$$\lambda_1 = \lim_{n\to\infty} \frac{1}{n} E[(G\mathbf{m} - Z)^{\mathrm{T}} (G\mathbf{m} - Z)],$$

$$\lambda_2 = \lim_{n\to\infty} \frac{1}{n} E[(G\mathbf{m} - Z)^{\mathrm{T}} G_c], \quad \lambda_3 = \lim_{n\to\infty} \frac{1}{n} E[(G\mathbf{m} - Z)^{\mathrm{T}} \mathbf{1}_n]$$

where $G_c = (g_{11}, \cdots, g_{nn})^{\mathrm{T}}$ and $\mathbf{1}_n$ is an n dimensional vector with each component being 1. Further, let

$$\Omega = \begin{pmatrix} \frac{1}{\sigma^2}\lambda_1 + \pi_1 & \frac{1}{\sigma^2}\pi_2 \\ \frac{1}{\sigma^2}\pi_2 & \frac{1}{2\sigma^4} \end{pmatrix}, \quad \Sigma = \begin{pmatrix} \frac{\mu_4 - 3\sigma^4}{\sigma^4}\pi_3 + \frac{2\mu_3}{\sigma^4}\lambda_2 & \frac{\mu_3}{2\sigma^6}\lambda_3 + \frac{\mu_4 - 3\sigma^4}{2\sigma^6}\pi_2 \\ \frac{\mu_3}{2\sigma^6}\lambda_3 + \frac{\mu_4 - 3\sigma^4}{2\sigma^6}\pi_2 & \frac{\mu_4 - 3\sigma^4}{4\sigma^8} \end{pmatrix},$$

$$s = (u, v)^{\mathrm{T}}, \quad \beta_{uu}(s) = \left(\frac{\partial^2 \beta_1(s)}{\partial u^2}, \cdots, \frac{\partial^2 \beta_p(s)}{\partial u^2}\right)^{\mathrm{T}},$$

$$\beta_{vv}(s) = \left(\frac{\partial^2 \beta_1(s)}{\partial v^2}, \cdots, \frac{\partial^2 \beta_p(s)}{\partial v^2}\right)^{\mathrm{T}}$$

and

$$S = \begin{pmatrix} (X_1^{\mathrm{T}}, \mathbf{0}_{1\times 2p})\Big(\mathcal{X}_{(1)}^{\mathrm{T}} \mathcal{W}_{(1)} \mathcal{X}_{(1)}\Big)^{-1} \mathcal{X}_{(1)}^{\mathrm{T}} \mathcal{W}_{(1)} \\ \vdots \\ (X_n^{\mathrm{T}}, \mathbf{0}_{1\times 2p})\Big(\mathcal{X}_{(n)}^{\mathrm{T}} \mathcal{W}_{(n)} \mathcal{X}_{(n)}\Big)^{-1} \mathcal{X}_{(n)}^{\mathrm{T}} \mathcal{W}_{(n)} \end{pmatrix}$$

where $\mathcal{X}_{(i)}$ and $\mathcal{W}_{(i)}$ are $\mathcal{X}$ and $\mathcal{W}$ respectively with s being replaced by s_i, $i = 1, \cdots, n$.

By some simple calculations, we can see the matrix Ω defined above is the limit of the Fisher information matrix of α and σ^2. As the singularity of matrix Ω may have serious implication on the convergence rate of the proposed estimators, we present the asymptotic properties for the case where Ω is nonsingular and the case where Ω is singular separately. We present the nonsingular case in Theorems 1 - 3, and singular case in Theorems 4 - 7.

Theorem 1. *Under the Conditions (1)-(7) or Conditions (1)-(6), $(\tilde{7})$ and (8) in Appendix A, Ω is nonsingular, and when $n^{1/2}h^2/\log^2 n \to \infty$ and $nh^8 \to 0$, $\hat{\alpha}$ and $\hat{\sigma}^2$ are consistent estimators of α and σ^2, respectively.*

Theorem 1 shows the conditions under which Ω is nonsingular and the consistency of $\hat{\alpha}$ and $\hat{\sigma}^2$ under such conditions. Based on Theorem 1, we can derive the asymptotic nomality of $\hat{\alpha}$ and $\hat{\sigma}^2$.

Theorem 2. *Under the assumptions of Theorem 1, if the second partial derivative of $\beta(s)$ is Lipschitz continuous and $nh^6 \to 0$,*

$$\sqrt{n}\left(\hat{\alpha}-\alpha,\ \hat{\sigma}^2-\sigma^2\right)^T \xrightarrow{D} N(\mathbf{0},\ \Omega^{-1}+\Omega^{-1}\Sigma\Omega^{-1}).$$

Further, if ϵ_i is normally distributed,

$$\sqrt{n}\left(\hat{\alpha}-\alpha,\ \hat{\sigma}^2-\sigma^2\right)^T \xrightarrow{D} N(\mathbf{0},\ \Omega^{-1}).$$

Theorem 2 implies that the convergence rate of $\hat{\alpha}$ is of order $n^{-1/2}$ when Ω is nonsingular, which is the optimal rate for parametric estimation. We will see, in Theorem 5, this rate can not be achieved by $\hat{\alpha}$ when Ω is singular.

Theorem 3. *Under the assumptions of Theorem 1, if $nh_1^6 = O(1)$ and $h/h_1 \to 0$,*

$$\sqrt{nh_1^2 f(s)}\Big(\hat{\beta}(s)-\beta(s)-2^{-1}\kappa_0^{-1}\kappa_2 h_1^2\{\beta_{uu}(s)+\beta_{vv}(s)\}\Big) \xrightarrow{D} N\Big(\mathbf{0},\ \kappa_0^{-2}\nu_0\sigma^2\Psi^{-1}\Big).$$

for any given s.

Theorem 3 shows $\hat{\beta}(\cdot)$ is asymptotic normal and achieves the convergence rate of order $n^{-1/6}$, which is the optimal rate for bivariate nonparametric estimation.

We now turn to the case where Ω is singular.

Theorem 4. *Under the Conditions (1)-(6) and (9) in Appendix A, Ω is singular, and if $nh^8 \to 0$, $n^{1/2}h^2/\log^2 n \to \infty$, $\rho_n \to \infty$, $\rho_n h^4 \to 0$ and $nh^2/\rho_n \to \infty$, $\hat{\alpha}$ is a consistent estimator of α.*

Theorem 5. *Under the assumptions of Theorem 4, if the second partial derivative of $\boldsymbol{\beta}(s)$ is Lipschitz continuous and $nh^6 \to 0$,*

$$\sqrt{n/\rho_n}(\hat{\alpha} - \alpha) \xrightarrow{D} N(0,\ \sigma^2\lambda_4^{-1}),$$

where

$$\lambda_4 = \lim_{n\to\infty} \frac{\rho_n}{n} E[(G\mathbf{m}) - SG\mathbf{m})^T(G\mathbf{m} - SG\mathbf{m})].$$

Theorem 5 shows the convergence rate of $\hat{\alpha}$ is of order $(n/\rho_n)^{-1/2}$ which is slower than $n^{-1/2}$ when $\rho_n \longrightarrow \infty$. However, we will see, from Theorem 7, this has no effect on the asymptotic properties of $\hat{\boldsymbol{\beta}}(\cdot)$.

Theorem 6. *Under the assumptions of Theorem 5,*

$$\sqrt{n}(\hat{\sigma}^2 - \sigma^2) \xrightarrow{D} N(0,\ \mu_4 - \sigma^4).$$

Theorem 6 shows that although the asymptotic variance of $\hat{\sigma}^2$ is different to that when Ω is nonsingular, $\hat{\sigma}^2$ still enjoys convergence rate of $n^{-1/2}$.

Theorem 7. *Under the assumptions of Theorem 4, if $nh_1^6 = O(1)$ and $h/h_1 \to 0$,*

$$\sqrt{nh_1^2 f(s)}\Big(\hat{\boldsymbol{\beta}}(s) - \boldsymbol{\beta}(s) - 2^{-1}\kappa_0^{-1}\kappa_2 h_1^2\{\boldsymbol{\beta}_{uu}(s) + \boldsymbol{\beta}_{vv}(s)\}\Big) \xrightarrow{D} N\Big(\mathbf{0}, \kappa_0^{-2}\nu_0\sigma^2\Psi^{-1}\Big)$$

for any given s.

From Theorem 3 and Theorem 7, we can see the singularity of Ω has no effect on the asymptotic distribution of $\hat{\boldsymbol{\beta}}(\cdot)$.

5 Simulation studies

In this section, we will use simulated examples to examine the performances of the proposed estimation and model selection procedure. In all simulated examples and the real data analysis later on, we set w_{ij} to be

$$w_{ij} = \exp(-\|s_i - s_j\|)/\sum_{k\neq i}\exp(-\|s_i - s_k\|), \quad \|s_i\| = (s_i^{\mathrm{T}} s_i)^{1/2}. \qquad (5.1)$$

We first examine the performance of the proposed estimation procedure, then the model selection procedure.

5.1 Performance of the estimation procedure

Example 1. In model (1.3), we set $p = 2$, $\sigma^2 = 1$,

$$\alpha = 0.5, \quad \beta_1(s) = \sin(\|s\|^2\pi), \quad \beta_2(s) = \cos(\|s\|^2\pi),$$

and independently generate X_i from $N(\mathbf{0}_2,\ I_2)$, s_i from $U[0,\ 1]^2$, ϵ_i from $N(0,\ \sigma^2)$, $i = 1,\ \cdots,\ n$. y_i, $i = 1,\ \cdots,\ n$, are generated through model (1.3). We are going to apply the proposed estimation method based on the generated $(s_i,\ X_i^{\mathrm{T}},\ y_i)$, $i = 1,\ \cdots,\ n$, to estimate $\beta_1(\cdot)$, $\beta_2(\cdot)$, α and σ^2, and examine the accuracy of the proposed estimation procedure.

We use the Epanechnikov kernel $K(t) = 0.75(1-t^2)_+$ as the kernel function in the estimation procedure. The bandwidth used in the estimation is 0.45

We use mean squared error (MSE) to assess the accuracy of an estimator of an unknown constant parameter, mean integrated squared error (MISE) to assess the accuracy of an estimator of an unknown function.

For each given sample size n, we do 200 simulations. We compute the MSEs of the estimators of the unknown constants and the MISEs of the estimators of the unknown functions for sample size $n = 400$, $n = 500$ and $n = 600$. The obtained results are presented in Table 1. Table 1 shows the proposed estimation procedure works very well. To have a more visible idea about the performance of the proposed estimation procedure, we set sample size $n = 500$ and do 200 simulations. We single out the one with median performance among the 200 simulations. The estimate of α coming from this simulation is 0.42, the estimate of σ^2 is 0.98. The estimated unknown functions from this simulation are presented in Figures 1 and 2, and are superimposed with the true functions. All these show our estimation procedure works very well.

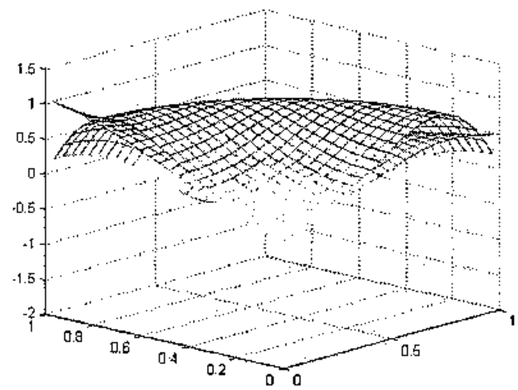

Figure 1: $\beta_1(s) = \sin(\|s\|^2\pi)$

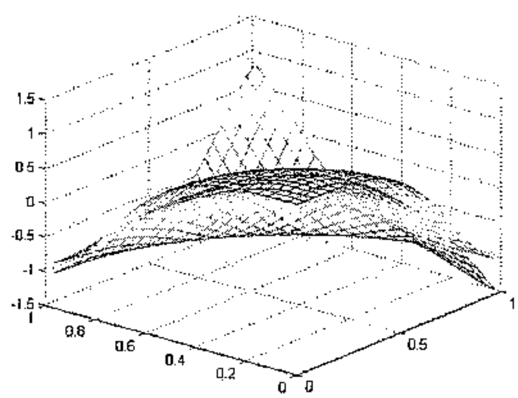

Figure 2: $\beta_2(s) = \cos(\|s\|^2\pi)$

5.2 Performance of the model selection procedure

Example 2. In model (1.3), we set $p = 3$, $\beta_1(\cdot)$ and $\beta_2(\cdot)$ the same as that in Example 1, $\beta_3(\cdot) = \beta_3 = 1$. We generate X_i, s_i, ϵ_i, y_i $i = 1, \cdots, n$, in the

Table 1: **The MISEs and MSEs**

	$\hat{\beta}_1(\cdot)$	$\hat{\beta}_2(\cdot)$	$\hat{\alpha}$	$\hat{\sigma}^2$
n=400	0.0497	0.0462	0.0108	0.0051
n=500	0.0418	0.0364	0.0078	0.0042
n=600	0.0369	0.0327	0.0055	0.0035

The column corresponding to the estimator of an unknown function is the MISEs of the estimator for $n = 400$, $n = 500$ and $n = 600$, corresponding to the estimator of an unknown constant is the MSEs of the estimator.

same way as that in Example 1, except that X_i is from $N(\mathbf{0}_3,\ I_3)$. Based on the generated data, we are going to apply the proposed AIC or BIC to select the correct model, and examine the performances of the proposed AIC, BIC and the two algorithms in identifying the constant components in model (1.3).

We still use the Epanechnikov kernel as the kernel function in the model selection, however, the bandwidth used is 0.25 for AIC and 0.35 for BIC, which is smaller than that for estimation. In general, the bandwidth used for model selection should be smaller than that for estimation. In fact, we have tried different bandwidths, it turned out any bandwidth in a reasonable range such as $[0.2,\ 0.3]$ for AIC, $[0.3,\ 0.4]$ for BIC would do the job very well.

Due to the very expensive computation involved, for any given sample size n, we only do 200 simulations, and in each simulation, we apply either AIC or BIC coupled with either of the two proposed algorithms to select model. For each candidate model, the ratios of picking up this model in the 200 simulations are computed for different cases. The results are presented in Table 2. We can see, from Table 2, the proposed BIC with Backward elimination performs best, and the others are doing reasonably well too.

To make the case more convincing, for sample size 500, we do 1000 simulations for each method. The ratio of picking up each candidate model in the 1000 simulations are presented in Table 3 for each method. It is very clear, the results in Table 3 are consistent with that in Table 2. We conclude all of

Table 2: **Ratio of Picking Up Each Candidate Model**

	{1}	{2}	{3}	{1, 2}	{1, 3}	{2, 3}	{1, 2, 3}	{}
n=400	0	0	0.91	0	0.04	0.02	0.03	0
n=500	0	0	0.98	0	0.02	0	0	0
n=600	0	0	1	0	0	0	0	0
n=400	0	0	0.9	0	0.07	0	0.03	0
n=500	0	0	0.94	0	0.06	0	0	0
n=600	0	0	0.96	0	0.03	0	0.01	0
n=400	0	0	0.92	0	0.05	0	0.03	0
n=500	0	0	0.98	0	0.01	0	0.01	0
n=600	0	0	1	0	0	0	0	0
n=400	0	0	0.89	0	0.09	0	0.02	0
n=500	0	0	0.95	0	0.05	0	0	0
n=600	0	0	0.97	0	0.03	0	0	0

The ratios of picking up each candidate model in 200 simulations for different sample sizes. $\{i_1, \cdots, i_k\}$ stands for the model in which $\beta(\cdot)$ has its i_1th, $\cdots$, i_kth components being constant, and the column corresponding to which is the ratios of picking up this model among 200 simulations. Row 2 to row 4 are the ratios obtained based on AIC and Backward elimination when sample size $n = 400$, $n = 500$ and $n = 600$. Row 5 to row 7 are the ratios obtained based on AIC and the CTAR based algorithm, Row 8 to row 10 are the ratios obtained based on BIC and Backward elimination, and Row 11 to row 13 are the ratios obtained based on BIC and the CTAR based algorithm.

the proposed model selection methods work well, and the proposed BIC with Backward elimination performs best.

6 Real data analysis

In this section, we are going to apply the proposed model (1.3) together with the proposed model selection and estimation method to analyse the Boston house price data. Specifically, we are going to explore how some factors such

Table 3: **Ratio of Picking Up Each Candidate Model**

$\{1\}$	$\{2\}$	$\{3\}$	$\{1, 2\}$	$\{1, 3\}$	$\{2, 3\}$	$\{1, 2, 3\}$	$\{\}$
0	0	0.989	0	0.01	0	0.001	0
0	0	0.959	0	0.033	0	0.008	0
0	0	0.992	0	0.005	0	0.003	0
0	0	0.963	0	0.031	0	0.006	0

The ratios of picking up each candidate model in 1000 *simulations for sample size* $n = 500$. $\{i_1, \cdots, i_k\}$ *stands for the model in which* $\beta(\cdot)$ *has its* i_1*th,* $\cdots$, i_k*th components being constant, and the column corresponding to which is the ratios of picking up this model among* 1000 *simulations. Row 2 are ratios obtained based on AIC and Backward elimination when sample size* $n = 500$. *Row 3 are the ratios obtained based on AIC and the CTAR based algorithm, Row 4 are the ratios obtained based on BIC and Backward elimination, and Row 5 are the ratios obtained based on BIC and the CTAR based algorithm.*

as the per capita crime rate by town (denoted by CRIM), average number of rooms per dwelling (denoted by RM), index of accessibility to radial highways (denoted by RAD), full-value property-tax rate per \$10,000 dollar (denoted by TAX), and the percentage of the lower status of the population (denoted by LSTAT) affect the median value of owner-occupied homes in \$1000's (denoted by MEDV), and whether the effects of these factors vary over location.

We use model (1.3) to fit the data with y_i, x_{i1}, x_{i2}, x_{i3}, x_{i4} and x_{i5} being MEDV, CRIM, RM, RAD, TAX and LSTAT, respectively, and $X_i = (x_{i1}, \cdots, x_{i5})^{\mathrm{T}}$. The kernel function used in either estimation procedure or model selection is taken to be the Epanechnikov kernel.

We first try to find which factors have location varying effects on the house price, and which factors do not. This is equivalent to identifying the constant coefficients in the model used to fit the data. We apply the proposed BIC coupled with Backward elimination to do the model selection, and the bandwidth used is chosen to be 17% of the range of the locations. The obtained result shows the coefficients of x_{i3} and x_{i5} are constant, which means all factors, except RAD and LSTAT, have location varying effects on the house price.

We now apply the chosen model

$$y_i = \alpha \sum_{j \neq i} w_{ij} y_j + x_{i1}\beta_1(s_i) + x_{i2}\beta_2(s_i) + x_{i3}\beta_3 + x_{i4}\beta_4(s_i) + x_{i5}\beta_5 + \epsilon_i, \quad (6.1)$$

$i = 1, \cdots, n$, where w_{ij} is defined by (5.1), to fit the data. The proposed estimation procedure is used to estimate the unknown functions and constants, and the bandwidth used in the estimation procedure is taken to be 60% of the range of the locations. The estimates of the unknown constants are presented in Table 4, and the estimates of the unknown functions are presented in Fig 3.

As β_3 and β_5 can be interpreted as the impacts of RAD and LSTAT, respectively, Table 4 shows the index of accessibility to radial highways has positive impact on house price and the percentage of the lower status of the population has negative impact on house price. Apparently, this makes sense. Table 4 also shows that the estimate of α is 0.221, which is an unignorable effect, and indicates the house prices in a neighbourhood do affect each other. This is a true phenomenon in real world.

Table 4: **Estimates of The Unknown Constant Coefficients**

$\hat{\alpha}$	$\hat{\beta}_3$	$\hat{\beta}_5$
0.2210	0.3589	-0.4473

From Fig 3, we can see the impact $\beta_1(\cdot)$ of the per capita crime rate by town on house price is negative and is clearly varying over location. The impact $\beta_2(\cdot)$ of the average number of rooms per dwelling on house price is positive and is also varying over location. It is interesting to see that the impact of the average number of rooms per dwelling is lower in the area where the impact of crime rate is high than the area where the impact of crime rate is low. This implies that the crime rate is a dominate factor on the house price in the area where the impact of crime rate is high. Fig 3 also shows the association between the house price and the full-value property-tax rate is varying over location, and it is generally positive, however, there are some areas where this association is

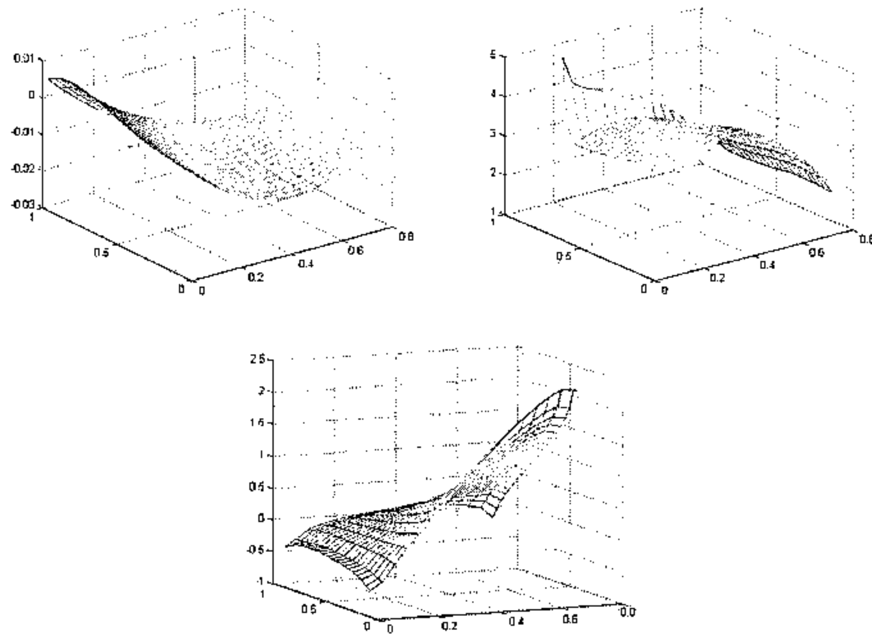

Figure 3: *The 3D plots of* $\hat{\beta}_1(s)$, $\hat{\beta}_2(s)$ *and* $\hat{\beta}_4(s)$. *The left one in the upper panel is* $\hat{\beta}_1(s)$, *right one in the upper panel is* $\hat{\beta}_2(s)$, *and the one in the lower panel is* $\hat{\beta}_4(s)$.

negative. We can also see that the impact of the average number of rooms per dwelling is lower in the area, where the association between the house price and the full-value property-tax rate is strong, than the area where the association is weak.

Appendix A

To avoid confusion of notation, we use α_0 to denote the true value of α in this section. Further, we rewrite $A = I_n - \alpha W$ as $A(\alpha)$ to emphasis its dependence on α and abbreviate $A(\alpha_0)$ as A.

The following regularity conditions are needed to establish the asymptotic properties of the estimators.

Conditions

(1) The kernel function $K(\cdot)$ is a bounded positive, symmetric and Lipshitz continuous function with a compact support on R. And $h \to 0$.

(2) $\{\beta_i(\cdot),\ i = 1, \cdots, p\}$ have continuous second partial derivatives.

(3) $\{X_i\}$ is a sequence of iid. random sample from the population and is independent of ϵ_i, $i = 1, \cdots, n$. Moreover, $E(X_1 X_1^{\mathrm{T}})$ is positive definite, $E\|X_1\|^{2q} < \infty$ and $E|\epsilon_1|^{2q} < \infty$ for some $q > 2$.

(4) $\{s_i\}$ is a sequence of fixed design points on a bounded support $\mathcal{S}$. Further, there exists a positive joint density function $f(\cdot)$ satisfying a Lipshitz condition such that

$$\sup_{s \in \mathcal{S}} \Big| \frac{1}{n} \sum_{i=1}^{n} [r(s_i) K_h(\|s_i - s\|) - \int r(t) K_h(\|t - s\|) f(t) dt \Big| = O(h)$$

for any bounded continuous function $r(\cdot)$ and $K_h(\cdot) = K(\cdot/h)/h^2$ where $K(\cdot)$ satisfies Condition (1). $f(\cdot)$ is bounded away from zero on $\mathcal{S}$.

(5) There exists a sequence $\rho_n > 0$ such that the elements w_{ij} of W are $O(1/\rho_n)$ uniformly in all i, j. As a normalization, $w_{ii} = 0$ for all i. Furthermore, the matrices W and A^{-1} are uniformly bounded in both row and column sums.

(6) $A^{-1}(\alpha)$ are uniformly bounded in either row or column sums, uniformly in α in a compact support Δ . The true α_0 is an interior point in Δ.

(7) $\lim_{n \to \infty} \frac{1}{n} E[(G\mathbf{m} - Z)^{\mathrm{T}} (G\mathbf{m} - Z)] = \lambda_1 > 0$.

($\tilde{7}$) $\lambda_1 = 0$.

(8) ρ_n is bounded and for any $\alpha \neq \alpha_0$,

$$\lim_{n \to \infty} \Big\{ \frac{1}{n} \log \Big| \sigma^2 A^{-1} (A^{-1})^{\mathrm{T}} \Big| - \frac{1}{n} \log \Big| \sigma_a^2(\alpha) A^{-1}(\alpha) (A^{-1}(\alpha))^{\mathrm{T}} \Big| \Big\} = 0$$

where $\sigma_a^2(\alpha) = \frac{\sigma^2}{n} \mathrm{tr}\{(A(\alpha) A^{-1})^{\mathrm{T}} A(\alpha) A^{-1}\}$.

(9) $\rho_n \to \infty$, the row sums of G have the uniform order $O(1/\sqrt{\rho_n})$, and

$$\lim_{n\to\infty} \frac{\rho_n}{n} E[(G\mathbf{m} - SG\mathbf{m})^{\mathrm{T}}(G\mathbf{m} - SG\mathbf{m})] = \lambda_4 > 0.$$

Remark 1: Condition (1)-(3) are commonly seen in nonparametric estimation. They are not the weakest possible ones, but they are imposed to facilitate the technical proofs. Since the sampling units can be regarded as given, the fixed bounded design Condition (4) is made for technical convenience. Of course as in Linton(1995), Condition (4) does not preclude $\{s_i\}_{i=1}^n$ from being generated by some random mechanism. For example, if s_i's were iid. with joint density $f(\cdot)$, then Condition (4) holds with probability one which can be obtained similarly as Hansen (2008). In this case, we can obtain our results by first conditional on $\{s_i\}_{i=1}^n$ and then go on as usual.

Remark 2: Condition (5)-(8) parallel the corresponding conditions of Lee (2004) and Su and Jin (2010), in which Condition (5)-(6) concern the essential features of the weight matrix for the model. Condition (7) is a sufficient condition which ensures that the likelihood function of α has a unique maximizer. When Condition $(\tilde{7})$ holds and the elements of W are uniformly bounded, the uniqueness of the maximizer can be guaranteed by Condition (8). These two kinds of conditions ensure that Ω which is the limit of the information matrix of the finite dimensional parameters is nonsingular. So they are the crucial conditions for $\sqrt{n}-$ rate of convergence of the finite dimensional parameter estimators.

Remark 3: When $\rho_n \to \infty$, Ω can be nonsingular only if Condition (7) holds. For the situation under Condition $(\tilde{7})$, Ω will become singular. The singularity of the matrix may have implications on the rate of convergence of the estimators. Nevertheless, we follow Lee (2004) and Su and Jin (2010) to consider the situation in which

$$\lim_{n\to\infty} \frac{\rho_n}{n} E[(G\mathbf{m})^{\mathrm{T}}(I_n - S)^{\mathrm{T}}(I_n - S)G\mathbf{m}] = \lambda_4 \in (0, \infty).$$

In this case, it is natural to assume that the elements of $(I_n - S)G\mathbf{m}$ have the uniform order $O_P(1/\sqrt{\rho_n})$ which can be satisfied by the assumption that the row sums of G are of uniform order $O(1/\sqrt{\rho_n})$.

In the following, let H be a diagonal matrix of size $3p$ with its first p elements on the diagonal being 1 and the remaining elements being h, $P = (I_n - S)^{\mathrm{T}}(I_n - S)$ and $\epsilon = (\epsilon_1, \cdots, \epsilon_n)^{\mathrm{T}}$. Moreover, like α_0, we use σ_0^2 to denote the true value of σ^2 to avoid confusion of notation. Since the following notations will be frequently used in the proofs, we list here for easy reference.

$l(\alpha, \sigma^2) = -\frac{n}{2}\log(\sigma^2) + \log(|A(\alpha)|) - \frac{1}{2\sigma^2}(A(\alpha)Y)^{\mathrm{T}}PA(\alpha)Y,$

$l_c(\alpha) = -\frac{n}{2}\log\tilde{\sigma}^2(\alpha) + \log|A(\alpha)|,$

$\tilde{\sigma}^2(\alpha) = \frac{1}{n}(A(\alpha)Y)^{\mathrm{T}}PA(\alpha)Y,$

$\bar{\sigma}^2(\alpha) = \frac{1}{n}E[(A(\alpha)Y)^{\mathrm{T}}PA(\alpha)Y],$

$\sigma_a^2(\alpha) = \frac{\sigma_0^2}{n}\mathrm{tr}\{(A(\alpha)A^{-1})^{\mathrm{T}}A(\alpha)A^{-1}\}.$

To prove the theorems, the following lemmas are needed and their proofs can be founded in Appendix B.

Lemma 1. Let $\{Y_i\}$ be a sequence of independent random variables and $\{s_i\} \in R^2$ are nonrandom vectors. Suppose that for some $q > 2$, $\max_i E|Y_i|^q < \infty$. Then under Condition (1), we have

$$\sup_{s\in\mathcal{S}}\Big|\frac{1}{n}\sum_{i=1}^{n}\Big[K_h(\|s_i - s\|)Y_i - E\{K_h(\|s_i - s\|)Y_i\}\Big]\Big| = O_p\Big(\Big\{\frac{\log n}{nh^2}\Big\}^{1/2}\Big),$$

provided that $n^{1-2/q}h^2/\log^2 n \to \infty$ and $\lim_{n\to\infty}\frac{1}{n}\sum_{i=1}^{n}K_h(\|s_i - s\|) < \infty$ for any $s \in \mathcal{S}$.

Lemma 2. Under the Conditions (1)-(4), then when $n^{1/2}h^2/\log^2 n \to \infty$,

(1) $n^{-1}H^{-1}\mathcal{X}^{\mathrm{T}}\mathcal{W}\mathcal{X}H^{-1} = \begin{pmatrix} \kappa_0 f(s)\Psi & \mathbf{0}_{p\times 2p} \\ \mathbf{0}_{2p\times p} & \kappa_2 f(s)\Psi\otimes I_2 \end{pmatrix} + O_P(c_n\mathbf{1}_{3p}\mathbf{1}_{3p}^{\mathrm{T}})$

holds uniformly in $s \in \mathcal{S}$ where $c_n = h + \{\frac{\log n}{nh^2}\}^{1/2}$,

(2) $\boldsymbol{\beta}(s) - (I_p, \mathbf{0}_{p\times 2p})(\mathcal{X}^{\mathrm{T}}\mathcal{W}\mathcal{X})^{-1}\mathcal{X}^{\mathrm{T}}\mathcal{W}\mathbf{m} = -\frac{\kappa_2 h^2}{2\kappa_0}\Big\{\boldsymbol{\beta}_{uu}(s) + \boldsymbol{\beta}_{vv}(s)\Big\} + o_p(h^2\mathbf{1}_p)$

holds uniformly in $s \in \mathcal{S}$.

Lemma 3. Under the Conditions (1)-(5), then when $n^{1/2}h^2/\log^2 n \to \infty$,

$$n^{-1}H^{-1}\mathcal{X}^{\mathrm{T}}\mathcal{W}G\mathbf{m} - n^{-1}E(H^{-1}\mathcal{X}^{\mathrm{T}}\mathcal{W}G\mathbf{m}) = o_P(1)$$

uniformly in $s \in \mathcal{S}$.

Lemma 4. Under the Conditions (1)(3)(4) and (5), when $n^{1/2}h^2/\log^2 n \to \infty$, we have (1) $\frac{1}{n}E[\mathrm{tr}(P)] = 1 + o(1)$, (2) $\frac{1}{n}E[\mathrm{tr}(G^{\mathrm{T}}P) - \mathrm{tr}(G)] = o(1)$, (3) $\frac{1}{n}E[\mathrm{tr}(G^{\mathrm{T}}PG) - \mathrm{tr}(G^{\mathrm{T}}G)] = o(1)$. Further, when $nh^2/\rho_n \to \infty$, then (4) $\frac{\rho_n}{n}E[\mathrm{tr}(P) - n] = o(1)$, (5) $\frac{\rho_n}{n}E[\mathrm{tr}(G^{\mathrm{T}}P) - \mathrm{tr}(G)] = o(1)$, (6) $\frac{\rho_n}{n}E[\mathrm{tr}(G^{\mathrm{T}}PG) - \mathrm{tr}(G^{\mathrm{T}}G)] = o(1)$.

Lemma 5. Under the Conditions (1)-(5), then when $n^{1/2}h^2/\log^2 n \to \infty$, (1) $(G\mathbf{m})^{\mathrm{T}}P\mathbf{m} = o_P(nh^2)$. Moreover, under the assumption that the second partial derivatives of $\beta(s)$ are all Lipschitz continuous, we have (2) $(G\mathbf{m})^{\mathrm{T}}P\mathbf{m} = O_P(nh^3 + \{nh^2\log n\}^{1/2})$.

Lemma 6. Under the Conditions (1)-(5), when $n^{1/2}h^2/\log^2 n \to \infty$ and $nh^8 \to 0$, we have (1) $n^{-1/2}L^{\mathrm{T}}P\mathbf{m} = o_P(1)$ for $L = \mathbf{m}, \epsilon$ and $G\epsilon$, (2) $n^{-1}L^{\mathrm{T}}PG\mathbf{m} = o_P(1)$ for $L = \mathbf{m}$, ϵ and $G\epsilon$.

Lemma 7. Under the Conditions (1)-(5), when $n^{1/2}h^2/\log^2 n \to \infty$, we have (1) $n^{-1}\Big\{(G\mathbf{m})^{\mathrm{T}}PG\mathbf{m} - E[(G\mathbf{m})^{\mathrm{T}}PG\mathbf{m}]\Big\} = o_P(1)$, (2) $n^{-1}E[(G\mathbf{m})^{\mathrm{T}}PG\mathbf{m}] = n^{-1}E[(G\mathbf{m} - Z)^{\mathrm{T}}(G\mathbf{m} - Z)] + o(1)$.

Lemma 8. Under the Conditions (1)-(5), when $n^{1/2}h^2/\log^2 n \to \infty$, we have (1) $n^{-1/2}\{\epsilon^{\mathrm{T}}P\epsilon - \epsilon^{\mathrm{T}}\epsilon\} = o_P(1)$, (2) $n^{-1/2}\{\epsilon^{\mathrm{T}}G^{\mathrm{T}}P\epsilon - \epsilon^{\mathrm{T}}G^{\mathrm{T}}\epsilon\} = o_P(1)$, (3) $n^{-1/2}\{\epsilon^{\mathrm{T}}G^{\mathrm{T}}PG\epsilon - \epsilon^{\mathrm{T}}G^{\mathrm{T}}G\epsilon\} = o_P(1)$, (4) $n^{-1/2}\{(G\mathbf{m})^{\mathrm{T}}P\epsilon - (G\mathbf{m} - SG\mathbf{m})^{\mathrm{T}}\epsilon\} = o_P(1)$.

Lemma 9. Suppose that $B = (b_{ij})_{1\le i,j\le n}$ is a sequence of symmetric matrices with row and column sums uniformly bounded and its elements are also uniformly bounded. Let $\sigma^2_{Q_n}$ be the variance of Q_n where $Q_n = (G\mathbf{m} - SG\mathbf{m})^{\mathrm{T}}\epsilon + \epsilon^{\mathrm{T}}B\epsilon - \sigma_0^2\mathrm{tr}(B)$. Assume that the variance $\sigma^2_{Q_n}$ is $O(n)$ with $\{\frac{\sigma^2_{Q_n}}{n}\}$ bounded away from zero, then we have under Conditions (3)- (5) that $\frac{Q_n}{\sigma_{Q_n}} \xrightarrow{D} N(0,1)$.

Lemma 10. Under the Conditions (1)-(5), the row sums of matrix G having the uniform order $O(1/\sqrt{\rho_n})$ and $n^{1/2}h^2/\log^2 n \to \infty$, we have (1) $(G\mathbf{m})^{\mathrm{T}}P\mathbf{m} = o_P(\rho_n^{-1/2}nh^2)$. Moreover, if the second partial derivatives of $\beta(s)$ are all Lipschitz continuous, then (2) $(G\mathbf{m})^{\mathrm{T}}P\mathbf{m} = O_P(\rho_n^{-1/2}nh^3 + \{nh^2\log n/\rho_n\}^{1/2})$.

Lemma 11. Under the Conditions (1)-(5) and the row sums of matrix G having the uniform order $O(1/\sqrt{\rho_n})$, then when $n^{1/2}h^2/\log^2 n \to \infty$, $\rho_n \to \infty$, $\rho_n h^4 \to 0$ and $nh^2/\rho_n \to \infty$, we have (1) $\frac{\rho_n}{n}\mathbf{m}^{\mathrm{T}}P\mathbf{m} = o_P(1)$, (2) $\frac{\rho_n}{n}L^{\mathrm{T}}PG\mathbf{m} = o_P(1)$ for $L = \mathbf{m}, \epsilon$ and $G\epsilon$, (3) $\sqrt{\frac{\rho_n}{n}}(G\epsilon)^{\mathrm{T}}P\mathbf{m} = o_P(1)$, (4) $\frac{\rho_n}{n}\Big\{(G\mathbf{m})^{\mathrm{T}}PG\mathbf{m} - E[(G\mathbf{m})^{\mathrm{T}}PG\mathbf{m}]\Big\} = o_P(1)$, (5) $\sqrt{\frac{\rho_n}{n}}\{\epsilon^{\mathrm{T}}G^{\mathrm{T}}P\epsilon - \epsilon^{\mathrm{T}}G^{\mathrm{T}}\epsilon\} = o_P(1)$, (6) $\sqrt{\frac{\rho_n}{n}}\{\epsilon^{\mathrm{T}}G^{\mathrm{T}}PG\epsilon - \epsilon^{\mathrm{T}}G^{\mathrm{T}}G\epsilon\} = o_P(1)$, (7) $\sqrt{\frac{\rho_n}{n}}\{(G\mathbf{m})^{\mathrm{T}}P\epsilon - (G\mathbf{m} - SG\mathbf{m})^{\mathrm{T}}\epsilon\} = o_P(1)$.

Lemma 12. Suppose that $B = (b_{ij})_{1\le i,j\le n}$ is a sequence of symmetric matrices with row and column sums uniformly bounded. Let $\sigma^2_{Q_n}$ be the variance of Q_n where $Q_n = (G\mathbf{m} - SG\mathbf{m})^{\mathrm{T}}\epsilon + \epsilon^{\mathrm{T}}B\epsilon - \sigma_0^2\mathrm{tr}(B)$. Assume that the variance $\sigma^2_{Q_n}$ is $O(n/\rho_n)$ with $\{\frac{\rho_n}{n}\sigma^2_{Q_n}\}$ bounded away from zero, the elements of B are of uniform order $O(1/\rho_n)$ and the row sums of G of uniform order $O(1/\sqrt{\rho_n})$. Then we have under $\rho_n \to \infty$ and Conditions (3)-(5) that $\frac{Q_n}{\sigma_{Q_n}} \xrightarrow{D} N(0,1)$.

In the proofs of the theorems, we will use the facts that for constant matrices B and D, $\mathrm{var}(\epsilon^{\mathrm{T}}B\epsilon) = (\mu_4 - 3\sigma_0^4)\sum_{i=1}^{n} b_{ii}^2 + \sigma_0^4[\mathrm{tr}(BB^{\mathrm{T}}) + \mathrm{tr}(B^2)]$ and

$$E(\epsilon^{\mathrm{T}}B\epsilon\epsilon^{\mathrm{T}}D\epsilon) = (\mu_4 - 3\sigma_0^4)\sum_{i=1}^{n} b_{ii}d_{ii} + \sigma_0^4[\mathrm{tr}(B)\mathrm{tr}(D) + \mathrm{tr}(BD) + \mathrm{tr}(BD^{\mathrm{T}})].$$

Moreover, we will frequently use the following facts by Condition (5) (see Lee, 2004) without clearly pointed out:

(1) the elements of $G = WA^{-1}$ are $O(1/\rho_n)$ uniformly in all i, j.

(2) The matrix $G = WA^{-1}$ is uniformly bounded in both row and column sums.

Proof of Theorem 1: First we will show that Ω is nonsingular. Let $\mathbf{d} = (d_1, d_2)^{\mathrm{T}}$ be a constant vector such that $\Omega\mathbf{d} = \mathbf{0}_2$. Then it is sufficient to show that $\mathbf{d} = \mathbf{0}_2$. From the second equation of $\Omega\mathbf{d} = \mathbf{0}_2$ we have that $d_2 = -2\sigma_0^2 \lim_{n\to\infty}\frac{1}{n}\mathrm{tr}(G)d_1$. Plug d_2 into the first equation of $\Omega\mathbf{d} = \mathbf{0}_2$ and we get that

$$d_1\Big\{\frac{1}{\sigma_0^2}\lambda_1 + \lim_{n\to\infty}\Big[\frac{1}{n}\mathrm{tr}((G+G^{\mathrm{T}})G) - \frac{2}{n^2}\mathrm{tr}^2(G)\Big]\Big\} = 0.$$

It follows by Condition (7) that $\lambda_1 > 0$. Moreover, $\mathrm{tr}\{(G+G^{\mathrm{T}})G\} - \frac{2}{n}\mathrm{tr}^2(G) = \frac{1}{2}\mathrm{tr}\{(\tilde{G}^{\mathrm{T}} + \tilde{G})(\tilde{G}^{\mathrm{T}} + \tilde{G})^{\mathrm{T}}\} \ge 0$ where $\tilde{G} = G - \frac{1}{n}\mathrm{tr}(G)I_n$. As we have by Condition (5) that the elements of $\tilde{G}$ are uniformly $O(1/\rho_n)$ and its row and column sums are also uniformly bounded, then it can be easily shown that $\mathrm{tr}\{(\tilde{G}^{\mathrm{T}}+\tilde{G})(\tilde{G}^{\mathrm{T}}+\tilde{G})^{\mathrm{T}}\} = O(\frac{n}{\rho_n})$. Therefore, if Condition ($\tilde{7}$) holds, Condition (8) implies that the limit of $\frac{1}{n}\mathrm{tr}((G+G^{\mathrm{T}})G) - \frac{2}{n^2}\mathrm{tr}^2(G) = \frac{1}{2n}\mathrm{tr}\{(\tilde{G}^{\mathrm{T}}+\tilde{G})(\tilde{G}^{\mathrm{T}}+\tilde{G})^{\mathrm{T}}\} > 0$. Therefore, $d_1 = 0$ and $d_2 = 0$.

Next we will follow the idea of Lee (2004) to show the consistency of $\hat{\alpha}$. Define $Q(\alpha)$ to be $\max_{\sigma^2} E\Big[l(\alpha,\sigma^2)\Big]$ by ignoring the constant term. The optimal solutions of this maximization problem are $\bar{\sigma}^2(\alpha) = \frac{1}{n}E[(A(\alpha)Y)^{\mathrm{T}}PA(\alpha)Y]$. Consequently,

$$Q(\alpha) = -n/2 \cdot \log\bar{\sigma}^2(\alpha) + \log|A(\alpha)|.$$

According to White (1994, Theorem 3.4), it suffices to show the uniform convergence of $n^{-1}\{l_c(\alpha) - Q(\alpha)\}$ to zero in probability on Δ and the unique maximizer condition that

$$\limsup_{n\to\infty} \max_{\alpha\in N^c(\alpha_0,\delta)} n^{-1}|Q(\alpha) - Q(\alpha_0)| < 0 \quad \text{for any } \delta > 0 \tag{A.1}$$

where $N^c(\alpha_0,\delta)$ is the complement of an open neighborhood of α_0 in Δ with diameter δ.

Note that $\frac{1}{n}l_c(\alpha) - \frac{1}{n}Q(\alpha) = -\frac{1}{2}\{\log\tilde{\sigma}^2(\alpha) - \log\bar{\sigma}^2(\alpha)\}$, then to show the uniform convergence, it is sufficient to show that $\tilde{\sigma}^2(\alpha) - \bar{\sigma}^2(\alpha) = o_P(1)$ uniformly on Δ and $\bar{\sigma}^2(\alpha)$ is uniformly bounded away from zero on Δ. Since

$$\begin{aligned}
& \tilde{\sigma}^2(\alpha) - \bar{\sigma}^2(\alpha) \\
= \; & n^{-1}\Big\{(A(\alpha)A^{-1}\mathbf{m})^{\mathrm{T}}PA(\alpha)A^{-1}\mathbf{m} - E[(A(\alpha)A^{-1}\mathbf{m})^{\mathrm{T}}PA(\alpha)A^{-1}\mathbf{m}]\Big\} \\
& + n^{-1}\Big\{(A(\alpha)A^{-1}\epsilon)^{\mathrm{T}}PA(\alpha)A^{-1}\epsilon - \sigma_0^2E[\mathrm{tr}\{(A(\alpha)A^{-1})^{\mathrm{T}}PA(\alpha)A^{-1}\}]\Big\} \\
& + 2n^{-1}(A(\alpha)A^{-1}\mathbf{m})^{\mathrm{T}}PA(\alpha)A^{-1}\epsilon,
\end{aligned}$$

and $A(\alpha)A^{-1} = I_n + (\alpha_0 - \alpha)G$ by $WA^{-1} = G$, it follows from Lemma 6 and Lemma 7(1) that

$$n^{-1}\Big\{(A(\alpha)A^{-1}\mathbf{m})^{\mathrm{T}}PA(\alpha)A^{-1}\mathbf{m} - E[(A(\alpha)A^{-1}\mathbf{m})^{\mathrm{T}}PA(\alpha)A^{-1}\mathbf{m}]\Big\} = o_P(1)$$

and

$$n^{-1}(A(\alpha)A^{-1}\mathbf{m})^{\mathrm{T}}PA(\alpha)A^{-1}\epsilon = o_P(1).$$

Next, we have by Lemma 4(1)-(3), Lemma 8(1)-(3) and Chebyshev inequality that

$$n^{-1}\Big\{(A(\alpha)A^{-1}\epsilon)^{\mathrm{T}}PA(\alpha)A^{-1}\epsilon-\sigma_0^2E[\mathrm{tr}\{(A(\alpha)A^{-1})^{\mathrm{T}}PA(\alpha)A^{-1}\}]\Big\}=o_P(1).$$

Therefore, $\tilde{\sigma}^2(\alpha)-\bar{\sigma}^2(\alpha)=o_P(1)$ uniformly on Δ.

Now we will show that $\bar{\sigma}^2(\alpha)$ is bounded away from zero uniformly on Δ. As we know by simple calculation and Lemma 4(1)-(3) that

$$\begin{aligned}\bar{\sigma}^2(\alpha) &\geq \sigma_0^2n^{-1}E\Big[\mathrm{tr}\{(A(\alpha)A^{-1})^{\mathrm{T}}PA(\alpha)A^{-1}\}\Big]\\ &= \sigma_0^2n^{-1}\mathrm{tr}\{(A(\alpha)A^{-1})^{\mathrm{T}}A(\alpha)A^{-1}\}+o(1),\end{aligned}\tag{A.2}$$

it suffices to show that $\sigma_a^2(\alpha)=\frac{\sigma_0^2}{n}\mathrm{tr}\{(A(\alpha)A^{-1})^{\mathrm{T}}A(\alpha)A^{-1}\}$ is uniformly bounded away from zero on Δ. To do so, we define an auxiliary spatial autoregressive (SAR) process: $Y=\alpha_0WY+\epsilon$ with $\epsilon\sim N(0,\sigma_0^2I_n)$. Then its log likelihood function without the constant term is

$$l_a(\alpha,\sigma^2)=-\frac{n}{2}\log\sigma^2+\log|A(\alpha)|-\frac{1}{2\sigma^2}(A(\alpha)Y)^{\mathrm{T}}A(\alpha)Y.$$

Set $Q_a(\alpha)$ to be $\max\limits_{\sigma^2}E_a[l_a(\alpha,\sigma^2)]$ by ignoring the constant term, where E_a is the expectation under this SAR process. It can be easily shown that

$$Q_a(\alpha)=-n/2\cdot\log\sigma_a^2(\alpha)+\log|A(\alpha)|,$$

By Jensen inequality, for all $\alpha\in\Delta$, $\max\limits_{\sigma^2}E_a[l_a(\alpha,\sigma^2)]\leq E_a[l_a(\alpha_0,\sigma_0^2)]$, thus $Q_a(\alpha)\leq Q_a(\alpha_0)$. As

$$\frac{1}{n}[Q_a(\alpha)-Q_a(\alpha_0)]=-\frac{1}{2}\log\sigma_a^2(\alpha)+\frac{1}{2}\log\sigma_0^2+\frac{1}{n}\Big(\log|A(\alpha)|-\log|A(\alpha_0)|\Big)$$

uniformly on Δ, then it follows that

$$-\frac{1}{2}\log\sigma_a^2(\alpha)\leq-\frac{1}{2}\log\sigma_0^2+\frac{1}{n}\Big(\log|A(\alpha_0)|-\log|A(\alpha)|\Big).$$

If we can show that

$$n^{-1}\{\log|A(\alpha_2)|-\log|A(\alpha_1)|\}=O(1)\ \text{ uniformly in }\alpha_1\text{ and }\alpha_2\text{ on }\Delta\tag{A.3}$$

then $-\frac{1}{2}\log\sigma_a^2(\alpha)$ is bounded from above for any $\alpha \in \Delta$. Therefore, the statement that $\sigma_a^2(\alpha)$ is uniformly bounded away from zero on Δ can be established by a counter argument.

Now we will verify (A.3), it follows by the mean value theorem and Condition (5)-(6) that

$$\begin{aligned} n^{-1}\{\log|A(\alpha_2)| - \log|A(\alpha_1)|\} &= -n^{-1}\mathrm{tr}\{WA^{-1}(\tilde{\alpha})\}(\alpha_2 - \alpha_1) \\ &= O(\rho_n^{-1})(\alpha_2 - \alpha_1) \qquad (A.4) \end{aligned}$$

where $\tilde{\alpha}$ lies between α_1 and α_2. (A.3) is then established by Δ being a bounded set.

To show the uniqueness condition (A.1), write

$$\begin{aligned} n^{-1}[Q(\alpha) - Q(\alpha_0)] &= n^{-1}[Q_a(\alpha) - Q_a(\alpha_0)] \\ &\quad +2^{-1}[\log\sigma_a^2(\alpha) - \log\bar{\sigma}^2(\alpha)] + 2^{-1}[\log\bar{\sigma}^2(\alpha_0) - \log\sigma_0^2], \end{aligned}$$

it follows by Lemma 4(1) and Lemma 6(1) that $\bar{\sigma}^2(\alpha_0) - \sigma_0^2 = \frac{1}{n}E[\mathbf{m}^{\mathrm{T}}P\mathbf{m}] + \sigma_0^2\frac{1}{n}E[\mathrm{tr}(P)] - \sigma_0^2 = o(1)$. Hence, $\log\bar{\sigma}^2(\alpha_0) - \log\sigma_0^2 = o(1)$ as $\bar{\sigma}^2(\alpha_0)$ and σ_0^2 are both bounded away from zero. Moreover, we have already shown in (A.2) that $\lim_{n\to\infty}[\sigma_a^2(\alpha) - \bar{\sigma}^2(\alpha)] \le 0$, hence,

$$\limsup_{n\to\infty} \max_{\alpha\in N^c(\alpha_0,\delta)} n^{-1}[Q(\alpha) - Q(\alpha_0)] \le 0 \text{ for any } \delta > 0.$$

Now we will show that the above inequality holds strictly. Because $\bar{\sigma}^2(\alpha)$ is bounded away from zero and has a quadratic form of α with its coefficients bounded by Lemma 4(1)-(3), 6 and 7(2), this together with (A.4), we get that $n^{-1}Q(\alpha)$ is uniformly equicontinuous in α on Δ.

By the compactness of $N^c(\alpha_0,\delta)$, we suppose there would exist an $\delta > 0$ and a sequence $\{\alpha_n\}$ in $N^c(\alpha_0,\delta)$ converging to a point $\alpha^* \ne \alpha_0$ such that $\lim_{n\to\infty} n^{-1}[Q(\alpha_n) - Q(\alpha_0)] = 0$. Next, as $\alpha_n \to \alpha^*$, we have $\lim_{n\to\infty} n^{-1}[Q(\alpha_n) - Q(\alpha^*)] = 0$. Hence, it follows that

$$\lim_{n\to\infty} n^{-1}[Q(\alpha^*) - Q(\alpha_0)] = 0. \qquad (A.5)$$

Since we have known that $Q_a(\alpha^*) - Q_a(\alpha_0) \le 0$ and $\lim_{n\to\infty}[\sigma_a^2(\alpha^*) - \bar{\sigma}^2(\alpha^*)] \le 0$, (A.5) is possible only if (i) $\lim_{n\to\infty}[\sigma_a^2(\alpha^*) - \bar{\sigma}^2(\alpha^*)] = 0$ and (ii) $\lim_{n\to\infty} n^{-1}[Q_a(\alpha^*) -$

$Q_a(\alpha_0)] = 0$ both hold. However, (i) is a contradiction when Condition (7) holds as

$$\lim_{n\to\infty}[\sigma_a^2(\alpha^*) - \bar{\sigma}^2(\alpha^*)] = -(\alpha_0 - \alpha^*)^2 \lim_{n\to\infty} n^{-1}E[(G\mathbf{m} - Z)^{\mathrm{T}}(G\mathbf{m} - Z)] = 0,$$

by Lemma 4(1)-(3), 6 and 7(2). If Condition ($\tilde{7}$) holds, the contradiction follows from (ii) by Condition (8).

For the consistency of $\hat{\sigma}^2$, as it follows by some calculation, $A(\hat{\alpha})A^{-1} = I_n + (\alpha_0 - \hat{\alpha})G$, Lemma 6, 7, 8(1)-(3), Chebyshev inequality and $\hat{\alpha} \xrightarrow{P} \alpha_0$ that

$$\begin{aligned}
\hat{\sigma}^2 &= \frac{1}{n}(A(\hat{\alpha})Y - SA(\hat{\alpha})Y)^{\mathrm{T}}(A(\hat{\alpha})Y - SA(\hat{\alpha})Y) \\
&= \frac{1}{n}(A(\hat{\alpha})A^{-1}\mathbf{m})^{\mathrm{T}}PA(\hat{\alpha})A^{-1}\mathbf{m} + \frac{2}{n}(A(\hat{\alpha})A^{-1}\mathbf{m})^{\mathrm{T}}PA(\hat{\alpha})A^{-1}\epsilon \\
&\quad + \frac{1}{n}(A(\hat{\alpha})A^{-1}\epsilon)^{\mathrm{T}}PA(\hat{\alpha})A^{-1}\epsilon \\
&= \frac{1}{n}\epsilon^{\mathrm{T}}P\epsilon + o_P(1) = \sigma_0^2 + o_P(1).
\end{aligned}$$

Proof of Theorem 2: Denote $\boldsymbol{\theta} = (\alpha, \sigma^2)^{\mathrm{T}}$ and $\boldsymbol{\theta}_0 = (\alpha_0, \sigma_0^2)^{\mathrm{T}}$, we get by Taylor expansion that

$$0 = \frac{\partial l(\hat{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta}} = \frac{\partial l(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} + \frac{\partial^2 l(\tilde{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}^{\mathrm{T}}}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0),$$

where $\tilde{\boldsymbol{\theta}} = (\tilde{\alpha}, \tilde{\sigma}^2)^{\mathrm{T}}$ lies between $\hat{\boldsymbol{\theta}}$ and $\boldsymbol{\theta}_0$ and thus converges to $\boldsymbol{\theta}_0$ in probability by Theorem 1. Then the asymptotic distribution of $\hat{\boldsymbol{\theta}}$ can be obtained by showing that

$$-\frac{1}{n}\frac{\partial^2 l(\tilde{\boldsymbol{\theta}})}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}^{\mathrm{T}}} \xrightarrow{P} \Omega \quad \text{and} \quad \frac{1}{\sqrt{n}}\frac{\partial l(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} \xrightarrow{D} N(\mathbf{0}, \Sigma + \Omega)$$

where Ω is a nonsingular matrix by Theorem 1.

By straightforward calculation, it can be easily obtained that

$$\begin{aligned}
\tfrac{1}{n}\tfrac{\partial^2 l(\boldsymbol{\theta})}{\partial \alpha^2} &= -\tfrac{1}{n}\mathrm{tr}([WA^{-1}(\alpha)]^2) - \tfrac{1}{\sigma^2 n}(WY)^{\mathrm{T}}PWY, \\
\tfrac{1}{n}\tfrac{\partial^2 l(\boldsymbol{\theta})}{\partial \sigma^2\partial \sigma^2} &= \tfrac{1}{2\sigma^4} - \tfrac{1}{\sigma^6 n}(A(\alpha)Y)^{\mathrm{T}}PA(\alpha)Y, \\
\tfrac{1}{n}\tfrac{\partial^2 l(\boldsymbol{\theta})}{\partial \alpha\partial \sigma^2} &= -\tfrac{1}{\sigma^4 n}(WY)^{\mathrm{T}}PA(\alpha)Y.
\end{aligned} \tag{A.6}$$

As $A(\tilde{\alpha})A^{-1} = I_n + (\alpha_0 - \tilde{\alpha})G$ by $G = WA^{-1}$, we have

$$\frac{1}{n}\frac{\partial^2 l(\tilde{\boldsymbol{\theta}})}{\partial \sigma^2\partial \sigma^2} - \frac{1}{n}\frac{\partial^2 l(\boldsymbol{\theta}_0)}{\partial \sigma^2\partial \sigma^2} = o_P(1) \quad \text{and} \quad \frac{1}{n}\frac{\partial^2 l(\tilde{\boldsymbol{\theta}})}{\partial \alpha\partial \sigma^2} - \frac{1}{n}\frac{\partial^2 l(\boldsymbol{\theta}_0)}{\partial \alpha\partial \sigma^2} = o_P(1).$$

using Lemma 6, 7, 8(1)-(3), Chebyshev inequality and $\tilde{\boldsymbol{\theta}} \xrightarrow{P} \boldsymbol{\theta}_0$. Let $G(\alpha) = WA^{-1}(\alpha)$, then it follows by the mean value theorem that

$$\begin{aligned} & \frac{1}{n}\frac{\partial^2 l(\tilde{\boldsymbol{\theta}})}{\partial \alpha^2} - \frac{1}{n}\frac{\partial^2 l(\boldsymbol{\theta}_0)}{\partial \alpha^2} \\ = \; & -\frac{2}{n}\mathrm{tr}(G^3(\bar{\alpha}))(\tilde{\alpha} - \alpha_0) + \left(\frac{1}{\sigma_0^2} - \frac{1}{\tilde{\sigma}^2}\right)\frac{1}{n}(G\mathbf{m} + G\epsilon)^{\mathrm{T}} P(G\mathbf{m} + G\epsilon) \end{aligned}$$

for some $\bar{\alpha}$ between $\tilde{\alpha}$ and α_0. Note that $G(\alpha)$ is bounded in row and column sums uniformly in a neighborhood of α_0 by Condition (5)-(6). Therefore, $\frac{1}{n}\mathrm{tr}(G^3(\bar{\alpha})) = O(1/\rho_n)$. Since we have $\frac{1}{n}(G\mathbf{m} + G\epsilon)^{\mathrm{T}} P(G\mathbf{m} + G\epsilon) = O_P(1)$ by Lemma 6(2), 7, 8(3) and Markov inequality, it follows that $\frac{1}{n}\frac{\partial^2 l(\tilde{\boldsymbol{\theta}})}{\partial \alpha^2} - \frac{1}{n}\frac{\partial^2 l(\boldsymbol{\theta}_0)}{\partial \alpha^2} = o_P(1)$ by $\tilde{\alpha} \xrightarrow{P} \alpha_0$ and $\tilde{\sigma}^2 \xrightarrow{P} \sigma_0^2$.

Next input $\boldsymbol{\theta}_0$ into (A.6) and we can get by Lemma 6 that

$$\begin{aligned} &-\tfrac{1}{n}\tfrac{\partial^2 l(\boldsymbol{\theta}_0)}{\partial \alpha^2} = \tfrac{1}{n}\mathrm{tr}(G^2) + \tfrac{1}{\sigma_0^2 n}(G\mathbf{m})^{\mathrm{T}} P(G\mathbf{m}) + \tfrac{1}{\sigma_0^2 n}\epsilon^{\mathrm{T}} G^{\mathrm{T}} PG\epsilon + o_P(1), \\ &-\tfrac{1}{n}\tfrac{\partial^2 l(\boldsymbol{\theta}_0)}{\partial \sigma^2 \partial \sigma^2} = -\tfrac{1}{2\sigma_0^4} + \tfrac{1}{\sigma_0^6 n}\epsilon^{\mathrm{T}} P\epsilon + o_P(1), \\ &-\tfrac{1}{n}\tfrac{\partial^2 l(\boldsymbol{\theta}_0)}{\partial \alpha \partial \sigma^2} = \tfrac{1}{\sigma_0^4 n}\epsilon^{\mathrm{T}} G^{\mathrm{T}} P\epsilon + o_P(1). \end{aligned}$$

Thus, the result of $-\frac{1}{n}\frac{\partial^2 l(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^{\mathrm{T}}} \xrightarrow{P} \Omega$ can be obtained using Lemma 7, Lemma 8(1)-(3) and Chebyshev inequality.

In the following we will establish the asymptotic distribution of $\frac{1}{\sqrt{n}}\frac{\partial l(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}}$. It follows by Lemma 5(2) that $\frac{1}{\sqrt{n}}(G\mathbf{m})^{\mathrm{T}} P\mathbf{m} = O_P(n^{1/2}h^3 + \{h^2 \log n\}^{1/2}) = o_P(1)$ when $nh^6 \to 0$ and $h^2 \log n \to 0$. Then we have by straightforward calculation, Lemma 6(1) and Lemma 8 that

$$\begin{aligned} \frac{1}{\sqrt{n}}\frac{\partial l(\boldsymbol{\theta}_0)}{\partial \alpha} &= -\frac{1}{\sqrt{n}}\mathrm{tr}(G) + \frac{1}{\sigma_0^2\sqrt{n}}(WY)^{\mathrm{T}} PAY \\ &= \frac{1}{\sigma_0^2\sqrt{n}}\Big[(G\mathbf{m} - SG\mathbf{m})^{\mathrm{T}}\epsilon + \{\epsilon^{\mathrm{T}} G\epsilon - \sigma_0^2\mathrm{tr}(G)\}\Big] + o_P(1), \end{aligned}$$

and

$$\begin{aligned} \frac{1}{\sqrt{n}}\frac{\partial l(\boldsymbol{\theta}_0)}{\partial \sigma^2} &= -\frac{\sqrt{n}}{2\sigma_0^2} + \frac{1}{2\sigma_0^4\sqrt{n}}(AY)^{\mathrm{T}} PAY \\ &= \frac{1}{2\sigma_0^4\sqrt{n}}\{\epsilon^{\mathrm{T}}\epsilon - n\sigma_0^2\} + o_P(1). \end{aligned}$$

Next we have by straightforward calculation that

$$\mathrm{var}((G\mathbf{m} - SG\mathbf{m})^{\mathrm{T}}\epsilon + \{\epsilon^{\mathrm{T}} G\epsilon - \sigma_0^2\mathrm{tr}(G)\})$$

$$
\begin{aligned}
&= \sigma_0^2 E[(G\mathbf{m} - SG\mathbf{m})^{\mathrm{T}}(G\mathbf{m} - SG\mathbf{m})] + (\mu_4 - 3\sigma_0^4)\sum_{i=1}^{n} g_{ii}^2 + \sigma_0^4[\mathrm{tr}(GG^{\mathrm{T}}) + \mathrm{tr}(G^2)] \\
&\quad +2\mu_3 E[(G\mathbf{m} - SG\mathbf{m})^{\mathrm{T}} G_c],
\end{aligned}
$$

$\mathrm{var}(\boldsymbol{\epsilon}^{\mathrm{T}}\boldsymbol{\epsilon} - n\sigma_0^2) = n(\mu_4 - \sigma_0^4)$ and

$$
\begin{aligned}
&\mathrm{cov}\{(G\mathbf{m} - SG\mathbf{m})^{\mathrm{T}}\epsilon + \{\epsilon^{\mathrm{T}} G\epsilon - \sigma_0^2 \mathrm{tr}(G)\}, \epsilon^{\mathrm{T}}\epsilon - n\sigma_0^2\} \\
= \;& \mu_3 E[(G\mathbf{m} - SG\mathbf{m})^{\mathrm{T}} \mathbf{1}_n] + (\mu_4 - \sigma_0^4)\mathrm{tr}(G).
\end{aligned}
$$

Hence, it follows by Lemma 7(2) and some calculation that

$$
E\left(\frac{1}{n}\frac{\partial l(\boldsymbol{\theta}_0)}{\partial \theta}\frac{\partial l(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}^{\mathrm{T}}}\right) = \Sigma + \Omega + o(1).
$$

As the components of $\frac{1}{\sqrt{n}}\frac{\partial l(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} = \left(\frac{1}{\sqrt{n}}\frac{\partial l(\boldsymbol{\theta}_0)}{\partial \alpha}, \frac{1}{\sqrt{n}}\frac{\partial l(\boldsymbol{\theta}_0)}{\partial \sigma^2}\right)^{\mathrm{T}}$ are linear-quadratic forms of double arrays, using Lemma 9 we gain $\frac{1}{\sqrt{n}}\frac{\partial l(\boldsymbol{\theta}_0)}{\partial \boldsymbol{\theta}} \xrightarrow{D} N(\mathbf{0}, \Sigma + \Omega)$.

Proof of Theorem 3: It can be easily shown that

$$
\begin{aligned}
\sqrt{nh_1^2 f(s)}(\hat{\boldsymbol{\beta}}(s) - \boldsymbol{\beta}(s)) &= \sqrt{nh_1^2 f(s)}(I_p, \mathbf{0}_{p\times 2p})(\mathcal{X}_1^{\mathrm{T}}\mathcal{W}_1\mathcal{X}_1)^{-1}\mathcal{X}_1^{\mathrm{T}}\mathcal{W}_1\epsilon \\
&\quad + \sqrt{nh_1^2 f(s)}(\alpha_0 - \hat{\alpha})(I_p, \mathbf{0}_{p\times 2p})(\mathcal{X}_1^{\mathrm{T}}\mathcal{W}_1\mathcal{X}_1)^{-1}\mathcal{X}_1^{\mathrm{T}}\mathcal{W}_1 WY \\
&\quad + \sqrt{nh_1^2 f(s)}(I_p, \mathbf{0}_{p\times 2p})\{(\mathcal{X}_1^{\mathrm{T}}\mathcal{W}_1\mathcal{X}_1)^{-1}\mathcal{X}_1^{\mathrm{T}}\mathcal{W}_1\mathbf{m} - \boldsymbol{\beta}(s)\} \\
&\equiv J_{n1} + J_{n2} + J_{n3}
\end{aligned}
$$

where $\mathcal{X}_1$ and $\mathcal{W}_1$ are $\mathcal{X}$ and $\mathcal{W}$ respectively with h replaced by h_1.

Let H_1 be H with h replaced by h_1. It follows by straightforward calculation that

$$
\sqrt{n^{-1}h_1^2 f(s)} E\{H_1^{-1}\mathcal{X}_1^{\mathrm{T}}\mathcal{W}_1\epsilon\} = \mathbf{0}_{3p\times 1},
$$

and

$$
\begin{aligned}
&n^{-1}h_1^2 f(s)\mathrm{cov}\{H_1^{-1}\mathcal{X}_1^{\mathrm{T}}\mathcal{W}_1\epsilon\} = \sigma_0^2 n^{-1} h_1^2 f(s) E\{H_1^{-1}\mathcal{X}_1^{\mathrm{T}}\mathcal{W}_1^2\mathcal{X}_1 H_1^{-1}\} \\
= \;& \sigma_0^2 f^2(s)\begin{pmatrix} \nu_0\Psi + o_P(\mathbf{1}_p\mathbf{1}_p^{\mathrm{T}}) & o_P(\mathbf{1}_p\mathbf{1}_{2p}^{\mathrm{T}}) \\ o_P(\mathbf{1}_{2p}\mathbf{1}_p^{\mathrm{T}}) & \nu_2\Psi\otimes I_2 + o_P(\mathbf{1}_{2p}\mathbf{1}_{2p}^{\mathrm{T}}) \end{pmatrix}
\end{aligned}
$$

then it follows by central limit theorem, Lemma 2(1) and Slutsky's Theorem that

$$
J_{n1} \xrightarrow{D} N\left(\mathbf{0}, \nu_0\kappa_0^{-2}\sigma_0^2\Psi^{-1}\right).
$$

Moreover, it follows immediately from Lemma 3 that

$$n^{-1}\{H_1^{-1}\mathcal{X}_1^{\mathrm{T}}\mathcal{W}_1G(\mathbf{m}+\epsilon) - E[H_1^{-1}\mathcal{X}_1^{\mathrm{T}}\mathcal{W}_1G(\mathbf{m}+\epsilon)]\} = o_P(1)$$

This together with Lemma 2(1) and Condition (4) leads to

$$(I_p, \mathbf{0}_{p\times 2p})(\mathcal{X}_1^{\mathrm{T}}\mathcal{W}_1\mathcal{X}_1)^{-1}\mathcal{X}_1^{\mathrm{T}}\mathcal{W}_1G(\mathbf{m}+\epsilon) = O_P(1).$$

Next when $nh_1^6 = O(1)$ and $h/h_1 \to 0$, we have $\sqrt{\frac{h_1^2}{n}}(G\mathbf{m})^{\mathrm{T}}P\mathbf{m} = o_P(n^{1/2}h_1h^2) = o_P(1)$ using Lemma 5(1). Hence it can be seen from the proof of Theorem 2 that $\sqrt{\frac{h_1^2}{n}}\frac{\partial l(\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}} = o_P(1)$ and $\sqrt{nh_1^2}(\hat{\alpha}-\alpha_0) = o_P(1)$ under the assumptions of Theorem 3. Therefore,

$$J_{n2} = \sqrt{f(s)}\sqrt{nh_1^2}(\alpha_0-\hat{\alpha})(I_p, \mathbf{0}_{p\times 2p})(\mathcal{X}_1^{\mathrm{T}}\mathcal{W}_1\mathcal{X}_1)^{-1}\mathcal{X}_1^{\mathrm{T}}\mathcal{W}_1(G\mathbf{m}+G\epsilon) = o_P(1).$$

For J_{n3}, it can be obtained by Lemma 2(2) that

$$J_{n3} = \frac{\kappa_2 h_1^2}{2\kappa_0}\{\boldsymbol{\beta}_{uu}(s) + \boldsymbol{\beta}_{vv}(s)\} + o_P(h_1^2\mathbf{1}_p).$$

Finally combing the results of J_{n1}, J_{n2} and J_{n3}, when $nh_1^6 = O(1)$ and $h/h_1 \to 0$ we get the theorem.

Proof of Theorem 4: It is obvious from the proof of nonsingularity of Ω in Theorem 1 that under Condition (9), Ω is singular.

Next like Lee (2004), to prove the consistency of $\hat{\alpha}$, it suffices to show that

$$\frac{\rho_n}{n}\{l_c(\alpha) - l_c(\alpha_0) - [Q(\alpha) - Q(\alpha_0)]\} = o_P(1) \text{ uniformly on } \Delta,$$

where $Q(\alpha) = -n/2\cdot\log\bar{\sigma}^2(\alpha) + \log|A(\alpha)|$ defined as in the proof of Theorem 1 and α_0 is a unique maximizer.

It follows by the mean value theorem that

$$\begin{aligned}
& \frac{\rho_n}{n}\{l_c(\alpha) - l_c(\alpha_0) - [Q(\alpha) - Q(\alpha_0)]\} \\
= & -\frac{\rho_n}{2}\frac{\partial[\log\tilde{\sigma}^2(\tilde{\alpha}) - \log\bar{\sigma}^2(\tilde{\alpha})]}{\partial\alpha}(\alpha-\alpha_0) \\
= & \frac{1}{\tilde{\sigma}^2(\tilde{\alpha})}\frac{\rho_n}{n}\left\{[(WY)^{\mathrm{T}}PA(\tilde{\alpha})Y - L_n(\tilde{\alpha})] - \frac{\tilde{\sigma}^2(\tilde{\alpha}) - \bar{\sigma}^2(\tilde{\alpha})}{\bar{\sigma}^2(\tilde{\alpha})}L_n(\tilde{\alpha})\right\}(\alpha-\alpha_0)
\end{aligned}$$

where $\tilde{\alpha}$ lies between α and α_0, and $L_n(\tilde{\alpha}) = E[(WY)^{\mathrm{T}} PA(\tilde{\alpha})Y]$.

Note that $A(\tilde{\alpha})A^{-1} = I_n + (\alpha_0 - \tilde{\alpha})G$, by applying Lemma 4(5)(6), 11 and Chebyshev inequality we can get

$$\frac{\rho_n}{n}\{(WY)^{\mathrm{T}} PA(\tilde{\alpha})Y - L_n(\tilde{\alpha})\} = o_P(1) \quad \text{and} \quad \frac{\rho_n}{n} L_n(\tilde{\alpha}) = O(1).$$

Moreover, using the same lines as in the proof of Theorem 1, we can establish that $\tilde{\sigma}^2(\tilde{\alpha}) - \bar{\sigma}^2(\tilde{\alpha}) = o_P(1)$ for any $\tilde{\alpha}$ on Δ with $\bar{\sigma}^2(\alpha)$ being uniformly bounded away from zero on Δ. Thus $\tilde{\sigma}^2(\alpha)$ is uniformly bounded away from zero in probability. Consequently,

$$\frac{\rho_n}{n}\{l_c(\alpha) - l_c(\alpha_0) - [Q(\alpha) - Q(\alpha_0)]\} = o_P(1) \text{ uniformly on } \Delta.$$

The uniqueness condition of α_0 can be obtained by the uniform equicontinuity of $\frac{\rho_n}{n}[Q(\alpha) - Q(\alpha_0)]$ on Δ and $\lim\limits_{n\to\infty} \frac{\rho_n}{n}[Q(\alpha) - Q(\alpha_0)] < 0$ when $\alpha \neq \alpha_0$ using a counter argument as in the proof of Theorem 1.

Write

$$\begin{aligned}
\frac{\rho_n}{n}[Q(\alpha) - Q(\alpha_0)] &= -\frac{\rho_n}{2}[\log \bar{\sigma}^2(\alpha) - \log \bar{\sigma}^2(\alpha_0)] + \frac{\rho_n}{n}[\log|A(\alpha)| - \log|A(\alpha_0)|] \\
&\equiv -\frac{1}{2}J_{n1} + J_{n2}.
\end{aligned}$$

It follows by the mean value theorem

$$J_{n1} = \frac{\rho_n}{\bar{\sigma}^{*2}(\alpha)}(\bar{\sigma}^2(\alpha) - \bar{\sigma}^2(\alpha_0))$$

where $\bar{\sigma}^{*2}(\alpha)$ lies between $\bar{\sigma}^2(\alpha)$ and $\bar{\sigma}^2(\alpha_0)$. As $\bar{\sigma}^2(\alpha)$ is uniformly bounded away from zero on Δ, $\bar{\sigma}^{*2}(\alpha)$ is also uniformly bounded away from zero on Δ. Further, we can see by Lemma 4(5)(6) and Lemma 11 that $\rho_n(\bar{\sigma}^2(\alpha) - \bar{\sigma}^2(\alpha_0))$ is a quadratic form of α with its coefficients bounded. Therefore, J_{n1} is uniformly equicontinuous on Δ by the above results.

For J_{n2}, it can be seen by the mean value theorem that

$$J_{n2} = -\frac{\rho_n}{n}\mathrm{tr}(WA^{-1}(\tilde{\alpha}))(\alpha - \alpha_0)$$

where $\tilde{\alpha}$ lies between α and α_0, and $\mathrm{tr}(WA^{-1}(\tilde{\alpha})) = O\big(n/\rho_n\big)$ by Condition (5)-(6). Therefore, J_{n2} is uniformly equicontinuous on Δ.

In conclusion, $\frac{\rho_n}{n}[Q(\alpha) - Q(\alpha_0)]$ is uniformly equicontinuous on Δ.

Next we will show that when $\alpha \neq \alpha_0$, $\lim_{n\to\infty} \frac{\rho_n}{n}[Q(\alpha) - Q(\alpha_0)] < 0$. Using similar lines as in the proof of Theorem 1, let $Q_a(\alpha) = -\frac{n}{2}\log \sigma_a^2(\alpha) + \log|A(\alpha)|$, and write

$$\begin{aligned}\frac{\rho_n}{n}[Q(\alpha) - Q(\alpha_0)] &= \frac{\rho_n}{n}[Q_a(\alpha) - Q_a(\alpha_0)] - \frac{\rho_n}{2}[\log \bar{\sigma}^2(\alpha) - \log \sigma_a^2(\alpha)] \\ &\quad + \frac{\rho_n}{2}[\log \bar{\sigma}^2(\alpha_0) - \log \sigma_0^2]. \qquad (A.7)\end{aligned}$$

As it follows by the mean value theorem, Lemma 4(4)-(6) and Lemma 11(1)(2) that

$$\begin{aligned}-\frac{\rho_n}{2}[\log \bar{\sigma}^2(\alpha) - \log \sigma_a^2(\alpha)] &= -\frac{\rho_n}{2\sigma^{*2}(\alpha)}[\bar{\sigma}^2(\alpha) - \sigma_a^2(\alpha)] \\ &= -\frac{1}{2\sigma^{*2}(\alpha)}(\alpha_0 - \alpha)^2 \frac{\rho_n}{n} E[(G\mathrm{m})^{\mathrm{T}} PG\mathrm{m}] + o(1)\end{aligned}$$

where $\sigma^{*2}(\alpha)$ lies between $\bar{\sigma}^2(\alpha)$ and $\sigma_a^2(\alpha)$ and it therefore uniformly bounded away from zero on Δ. Then for any $\alpha \neq \alpha_0$, when condition (9) holds, $-\frac{\rho_n}{2}[\log \bar{\sigma}^2(\alpha) - \log \sigma^2(\alpha)] < 0$ for sufficient large n.

For the third term on the right side in (A.7), it can be obtained by the mean value theorem, Lemma 4(4) and Lemma 11(1) that

$$\frac{\rho_n}{2}[\log \bar{\sigma}^2(\alpha_0) - \log \sigma_0^2] = \frac{\rho_n}{2\sigma^{*2}}\{\bar{\sigma}^2(\alpha_0) - \sigma_0^2\} = o(1)$$

where σ^{*2} lies between $\bar{\sigma}^2(\alpha_0)$ and σ_0^2, and is bounded away from zero.

In consequence, $\lim_{n\to\infty} \frac{\rho_n}{n}\{Q(\alpha) - Q(\alpha_0)\} < 0$ when $\alpha \neq \alpha_0$, as we have shown $Q_a(\alpha) - Q_a(\alpha_0) \leq 0$ in the proof of Theorem 1.

Proof of Theorem 5: By Taylor expansion, we have that

$$0 = \frac{\partial l_c(\hat{\alpha})}{\partial \alpha} = \frac{\partial l_c(\alpha_0)}{\partial \alpha} + \frac{\partial^2 l_c(\tilde{\alpha})}{\partial \alpha^2}(\hat{\alpha} - \alpha_0)$$

where $\tilde{\alpha}$ lies between $\hat{\alpha}$ and α_0, and thus converges to α_0 in probability by Theorem 4. Then the asymptotic distribution of $\hat{\alpha}$ can be obtained by proving that when $\rho_n \to \infty$,

$$-\frac{\rho_n}{n}\frac{\partial^2 l_c(\tilde{\alpha})}{\partial \alpha^2} \xrightarrow{P} \sigma_1^2 \ \text{ and } \ \sqrt{\frac{\rho_n}{n}}\frac{\partial l_c(\alpha_0)}{\partial \alpha} \xrightarrow{D} N(0, \sigma_2^2/\sigma_0^4),$$

where $\sigma_1^2 = \frac{1}{\sigma_0^2}\lim_{n\to\infty}\frac{\rho_n}{n}E[(G\mathbf{m} - SG\mathbf{m})^{\mathrm{T}}(G\mathbf{m} - SG\mathbf{m})]$ and $\sigma_2^2 = \sigma_0^4\sigma_1^2$.

As we have by $A(\alpha)A^{-1} = I_n + (\alpha_0 - \alpha)G$, Lemma 11 and Chebyshev inequality that $\frac{\rho_n}{n}(WY)^{\mathrm{T}}PWY = \frac{\rho_n}{n}(G\mathbf{m} + G\epsilon)^{\mathrm{T}}P(G\mathbf{m} + G\epsilon) = O_P(1)$ and $\frac{\rho_n}{n}(WY)^{\mathrm{T}}PA(\alpha)Y = O_P(1)$, then when $\rho_n \to \infty$,

$$\begin{aligned}
&\frac{\rho_n}{n}\frac{\partial^2 l_c(\alpha)}{\partial\alpha^2}\\
= \;& \frac{\rho_n}{n}\Big\{\frac{2}{\tilde{\sigma}^4(\alpha)n}[(WY)^{\mathrm{T}}PA(\alpha)Y]^2 - \frac{1}{\tilde{\sigma}^2(\alpha)}(WY)^{\mathrm{T}}PWY - \mathrm{tr}([WA^{-1}(\alpha)]^2)\Big\}\\
= \;& -\frac{1}{\tilde{\sigma}^2(\alpha)}\cdot\frac{\rho_n}{n}(WY)^{\mathrm{T}}PWY - \frac{\rho_n}{n}\mathrm{tr}([WA^{-1}(\alpha)]^2) + o_P(1).
\end{aligned}$$

Further using Lemma 6(1), 8(1) and the above results, we can get when $\rho_n \to \infty$ that

$$\tilde{\sigma}^2(\alpha) = \frac{1}{n}\epsilon^{\mathrm{T}}P\epsilon + o_P(1) = \sigma_0^2 + o_P(1)$$

for any $\alpha \in \Delta$. Therefore it follows by the mean value theorem that

$$\begin{aligned}
&\frac{\rho_n}{n}\left\{\frac{\partial^2 l_c(\tilde{\alpha})}{\partial\alpha^2} - \frac{\partial^2 l_c(\alpha_0)}{\partial\alpha^2}\right\}\\
= \;& \Big\{\frac{1}{\tilde{\sigma}^2(\alpha_0)} - \frac{1}{\tilde{\sigma}^2(\tilde{\alpha})}\Big\}\frac{\rho_n}{n}(WY)^{\mathrm{T}}PWY - \frac{\rho_n}{n}\{\mathrm{tr}(G^2(\tilde{\alpha})) - \mathrm{tr}(G^2(\alpha_0))\} + o_P(1)\\
= \;& -\frac{\rho_n}{n}\mathrm{tr}(G^3(\bar{\alpha}))(\tilde{\alpha} - \alpha_0) + o_P(1)
\end{aligned}$$

where $G(\alpha) = WA^{-1}(\alpha)$. As $\mathrm{tr}(G^3(\bar{\alpha})) = O(n/\rho_n)$ uniformly on Δ by Condition (5)-(6), we obtain that $\frac{\rho_n}{n}\left\{\frac{\partial^2 l_c(\tilde{\alpha})}{\partial\alpha^2} - \frac{\partial^2 l_c(\alpha_0)}{\partial\alpha^2}\right\} = o_P(1)$ using $\tilde{\alpha} \xrightarrow{P} \alpha_0$.

Next it follows from $\tilde{\sigma}^2(\alpha_0) \xrightarrow{P} \sigma_0^2$, Lemma 11 and Chebyshev inequality that

$$-\frac{\rho_n}{n}\frac{\partial^2 l_c(\alpha_0)}{\partial\alpha^2} = \frac{1}{\sigma_0^2}\frac{\rho_n}{n}E[(G\mathbf{m})^{\mathrm{T}}PG\mathbf{m}] + \frac{\rho_n}{n}[\mathrm{tr}(G^2) + \mathrm{tr}(GG^{\mathrm{T}})] + o_P(1).$$

Therefore, $-\frac{\rho_n}{n}\frac{\partial^2 l_c(\tilde{\alpha})}{\partial\alpha^2} \xrightarrow{P} \sigma_1^2$ by the row sums of G being uniform order $O(1/\sqrt{\rho_n})$.

In the following we will establish the asymptotic distribution of $\sqrt{\frac{\rho_n}{n}}\frac{\partial l_c(\alpha_0)}{\partial\alpha}$. As it follows that $\sqrt{\frac{\rho_n}{n}}(G\mathbf{m})^{\mathrm{T}}P\mathbf{m} = o_P(n^{1/2}h^3 + \{h^2\log n\}^{1/2}) = o_P(1)$ when $nh^6 \to 0$, $h^2\log n \to 0$ by Lemma 10(2) and $\sqrt{\frac{\rho_n}{n}}(G\epsilon)^{\mathrm{T}}P\mathbf{m} = o_P(1)$ by Lemma

11(3). Then we have by straightforward calculation and Lemma 6(1), 8(1), 11(5)(7) that the first order derivative of $\sqrt{\frac{\rho_n}{n}} l_c(\alpha)$ at α_0 is

$$\sqrt{\frac{\rho_n}{n}}\frac{\partial l_c(\alpha_0)}{\partial \alpha} = \frac{1}{\tilde{\sigma}^2(\alpha_0)}\sqrt{\frac{\rho_n}{n}}\Big\{(WY)^{\mathrm{T}}PAY - \tilde{\sigma}^2(\alpha_0)\mathrm{tr}(G)\Big\},$$

with

$$\begin{aligned}&\sqrt{\frac{\rho_n}{n}}\{(WY)^{\mathrm{T}}PAY - \tilde{\sigma}^2(\alpha_0)\mathrm{tr}(G)\}\\ = &\sqrt{\frac{\rho_n}{n}}\Big\{(G\mathbf{m} - SG\mathbf{m})^{\mathrm{T}}\epsilon + \epsilon^{\mathrm{T}}[G - \frac{1}{n}\mathrm{tr}(G)I_n]\epsilon\Big\} + o_P(1),\end{aligned}$$

and

$$\begin{aligned}\sigma_{qn}^2 &\equiv \mathrm{var}\Big\{(G\mathbf{m} - SG\mathbf{m})^{\mathrm{T}}\epsilon + \epsilon^{\mathrm{T}}[G - \frac{1}{n}\mathrm{tr}(G)I_n]\epsilon\Big\}\\ &= \sigma_0^2 E[(G\mathbf{m} - SG\mathbf{m})^{\mathrm{T}}(G\mathbf{m} - SG\mathbf{m})] + (\mu_4 - 3\sigma_0^4)\sum_{i=1}^n\{g_{ii} - \frac{\mathrm{tr}(G)}{n}\}^2\\ &\quad +\sigma_0^4[\mathrm{tr}((G+G^{\mathrm{T}})G) - \frac{2}{n}\mathrm{tr}^2(G)] + 2\mu_3 E[(G\mathbf{m} - SG\mathbf{m})^{\mathrm{T}}(G_c - \frac{1}{n}\mathrm{tr}(G)1_n)].\end{aligned}$$

As we have by Lemma 12 that

$$\sigma_{qn}^{-1}\Big\{(G\mathbf{m} - SG\mathbf{m})^{\mathrm{T}}\epsilon + \epsilon^{\mathrm{T}}[G^{\mathrm{T}} - \frac{1}{n}\mathrm{tr}(G)I_n]\epsilon\Big\} \xrightarrow{D} N(0,1),$$

it follows that

$$\sqrt{\frac{n}{\rho_n}}(\hat{\alpha} - \alpha_0) = \Big(-\frac{\rho_n}{n}\frac{\partial^2 l_c(\tilde{\alpha})}{\partial \alpha^2}\Big)^{-1}\cdot\sqrt{\frac{\rho_n}{n}}\frac{\partial l_c(\alpha_0)}{\partial\alpha} \xrightarrow{D} N\Big(0, \sigma_0^2\lambda_4^{-1}\Big).$$

by $\frac{\rho_n}{n}\sigma_{qn}^2 \to \sigma_2^2$ and $\tilde{\sigma}^2(\alpha_0) \xrightarrow{P} \sigma_0^2$.

Proof of Theorem 6: By straightforward calculation, Lemma 6(1), Lemma 8(1), 11, Chebyshev inequality and Theorem 5, we get when $\rho_n \to \infty$ that

$$\begin{aligned}\sqrt{n}(\hat{\sigma}^2 - \sigma_0^2) &= \frac{1}{\sqrt{n}}(A(\hat{\alpha})Y - SA(\hat{\alpha})Y)^{\mathrm{T}}(A(\hat{\alpha})Y - SA(\hat{\alpha})Y) - \sqrt{n}\sigma_0^2\\ &= \frac{1}{\sqrt{n}}(\mathbf{m}+\epsilon)^{\mathrm{T}}P(\mathbf{m}+\epsilon) - \sqrt{n}\sigma_0^2\\ &\quad+\frac{2}{\sqrt{\rho_n}}\sqrt{\frac{n}{\rho_n}}(\alpha_0 - \hat{\alpha})\frac{\rho_n}{n}(G\mathbf{m} + G\epsilon)^{\mathrm{T}}P(\mathbf{m}+\epsilon)\\ &\quad+\frac{1}{\sqrt{n}}\{\sqrt{\frac{n}{\rho_n}}(\alpha_0 - \hat{\alpha})\}^2\frac{\rho_n}{n}(G\mathbf{m} + G\epsilon)^{\mathrm{T}}P(G\mathbf{m} + G\epsilon)\\ &= \frac{1}{\sqrt{n}}\sum_{i=1}^n(\epsilon_i^2 - \sigma_0^2) + o_P(1)\end{aligned}$$

This together with central limit theorem for iid random variables leads to

$$\sqrt{n}(\hat{\sigma}^2 - \sigma_0^2) \xrightarrow{D} N(0, \mu_4 - \sigma_0^4).$$

Proof of Theorem 7: The result can be obtained using the same lines as the proof of Theorem 3, except that here $J_{n2} = \sqrt{f(s)}\sqrt{\frac{nh_1^2}{\rho_n}}(\alpha_0 - \hat{\alpha})(I_p, 0_{p\times 2p})(n^{-1}H_1^{-1}\mathcal{X}_1^{\mathrm{T}}\mathcal{W}_1\mathcal{X}_1H_1^{-1})^{-1}\frac{\sqrt{\rho_n}}{n}H_1^{-1}\mathcal{X}_1^{\mathrm{T}}\mathcal{W}_1G(\mathbf{m}+\epsilon)$. It follows by Lemma 2(1), Markov inequality, the row sums of the matrix G having uniform order $O(1/\sqrt{\rho_n})$ and Condition (4) that

$$(I_p, 0_{p\times 2p})(n^{-1}H_1^{-1}\mathcal{X}_1^{\mathrm{T}}\mathcal{W}_1\mathcal{X}_1H_1^{-1})^{-1}\frac{\sqrt{\rho_n}}{n}H_1^{-1}\mathcal{X}_1^{\mathrm{T}}\mathcal{W}_1G(\mathbf{m}+\epsilon) = O_P(1).$$

Next, it can be seen from the proof of Theorem 5 and Lemma 10(1) that $\sqrt{\frac{\rho_n h_1^2}{n}}(G\mathbf{m})^{\mathrm{T}}P\mathbf{m} = o_P(n^{1/2}h_1h^2) = o_P(1)$ when $nh_1^6 = O(1)$ and $h/h_1 \to 0$. Hence, $\sqrt{\frac{nh_1^2}{\rho_n}}(\hat{\alpha}-\alpha) \xrightarrow{P} 0$ according to the arguments establishing Theorem 5. Consequently we have that $J_{n2} = o_P(1)$.

Appendix B. Proofs of the lemmas

In this section, we set $m_i = X_i^{\mathrm{T}}\beta(s_i)$, x_{il} be the lth ($l = 1, \cdots, p$) element of X_i, $i = 1, \cdots, n$, $r_n = (\frac{\log n}{nh^2})^{1/2}$, $[D]_{ij}$ be the (i,j)th elements of the matrix D, and c is a positive finite constant which may take different values at each appearance. Moreover, the operator Vec$(\cdot)$ creates a column vector from the matrix by simply stacking its column vectors below one another.

Frequently we will use the facts (see Lee, 2004) without clearly pointed out that the matrix G is uniformly bounded in both row and column sums, and the elements g_{ij} of G are $O(1/\rho_n)$ uniformly in all i, j.

Proof of Lemma 1: Let $\tau_n = n^{1/q}(\log n)^{1/2}$ and the following proof is organized as Hansen (2008). First, we deal with the truncation error in replacing Y_i with the truncated process $Y_i1(|Y_i| \leq \tau_n)$. Second, we replace the supremum with a maximization over a finite N-point grid. Third, we use Bernstein inequality to bound the remainder.

The first step is to truncate Y_i. Define $R(s) = \frac{1}{n}\sum_{i=1}^{n} K_h(\|s_i - s\|)Y_i 1(|Y_i| > \tau_n)$. Since $P(|Y_n| > \tau_n) \le \tau_n^{-q} E|Y_n|^q$ and $\sum_{n=1}^{\infty} \tau_n^{-q} = \sum_{n=1}^{\infty} n^{-1}(\log n)^{-q/2} < \infty$ for $q > 2$. It follows that with probability one $|Y_n| \le \tau_n$ for all sufficient large n. Since τ_n is increasing, we have for all sufficient large n, $|Y_i| \le \tau_n$ for all $i \le n$. This implies that $\sup_s |R(s)|$ is eventually zero with probability one.

Next by a standard argument and Condition (4)

$$E[R(s)] \le \frac{1}{n}\sum_{i=1}^{n} K_h(\|s_i - s\|)E|Y_i|^q/\tau_n^{q-1} \le c\tau_n^{1-q},$$

it follows that with probability one $\sup_s E|R(s)| = O(\tau_n^{1-q}) = O(r_n)$.

Combing the above results, we have that with probability one

$$\sup_{s\in\mathcal{S}} |R(s) - ER(s)| = O(r_n).$$

For the second step we create a grid to cover the region $\mathcal{S}$. As $\mathcal{S}$ is a compact region, we can find a finite positive constant c_1 such that $\mathcal{S} \subseteq \{s : \|s\| \le c_1\}$. Next we create a grid using regions of the form $N_l = \{s : \|s - s_l\| \le r_n h\}$. By selecting s_l to lay on a grid, the region $\{s : \|s\| \le c_1\}$ can be covered with $N \le c_1^2 h^{-2} r_n^{-2}$ such regions N_l. Therefore the supremum can be replaced by a maximization over these N-point grid.

From the assumption of the kernel function, we know that there exists a finite positive constant L, when $\|s\| > L$, $K(\|s\|) = 0$, and there exists a finite positive constant c_2 such that for all $s, s' \in R^2$, $|K(\|s\|) - K(\|s'\|)| \le c_2\Big|\|s\| - \|s'\|\Big| \le c_2\|s - s'\|$. Define $W^*(\|s\|) = c_2 I(\|s\| \le 2L)$, thus for $s \in N_l$, we have $\|\frac{s-s_l}{h}\| \le \tau_n$ and

$$|K(\|\frac{s_i - s}{h}\|) - K(\|\frac{s_i - s_l}{h}\|)| \le \tau_n W^*(\|\frac{s_i - s_l}{h}\|). \tag{B.1}$$

Now define $R_1(s) = \frac{1}{n}\sum_{i=1}^{n} K_h(\|s_i - s\|)Y_i 1(|Y_i| \le \tau_n)$ and $\tilde{R}_1(s) = \frac{1}{n}\sum_{i=1}^{n} W_h^*(\|s_i - s\|)|Y_i| 1(|Y_i| \le \tau_n)$ where $W_h^*(\|s_i - s\|) = W^*(\|\frac{s_i - s}{h}\|)/h^2$. Note that $E|\tilde{R}_1(s)| \le \frac{1}{n}\sum_{i=1}^{n} W_h^*(\|s_i - s\|)E|Y_i| < c_3$ for some positive constant c_3 by Condition (4). Then we have by (B.1) that

$$\sup_{s\in N_l} |R_1(s) - ER_1(s)| \quad \le \quad |R_1(s_l) - ER_1(s_l)| + r_n[|\tilde{R}_1(s_l)| + E|\tilde{R}_1(s_l)|]$$

$$
\begin{aligned}
&\leq \quad |R_1(s_l) - ER_1(s_l)| + r_n|\tilde{R}_1(s_l) - E\tilde{R}_1(s_l)| + 2r_nE|\tilde{R}_1(s_l)| \\
&\leq \quad |R_1(s_l) - ER_1(s_l)| + |\tilde{R}_1(s_l) - E\tilde{R}_1(s_l)| + 2c_3r_n
\end{aligned}
$$

with the final inequality because $r_n \leq 1$ for sufficient large n. Therefore, for sufficient large n

$$
\begin{aligned}
P(\sup_{s\in\mathcal{S}}|R_1(s) - ER_1(s)| > 4c_3r_n) \quad &\leq \quad N \max_{1\leq l\leq N} P(\sup_{s\in N_l}|R_1(s) - ER_1(s)| > 4c_3r_n) \\
&\leq \quad N \max_{1\leq l\leq N} P(|R_1(s_l) - ER_1(s_l)| > c_3r_n) \\
&\quad\quad +N \max_{1\leq l\leq N} P(|\tilde{R}_1(s_l) - E\tilde{R}_1(s_l)| > c_3r_n).
\end{aligned}
$$

Third, we will use Bernstein inequality to bound the above probabilites. Let $V_i(s) = Y_{i1}K(\|\frac{s_i-s}{h}\|) - E[Y_{i1}K(\|\frac{s_i-s}{h}\|)]$ where $Y_{i1} = Y_i1(|Y_i| \leq \tau_n)$. As $|Y_{i1}| \leq \tau_n$ and $K(\|\frac{s_i-s}{h}\|) \leq c_4$ for some positive constant c_4, it follows that $|V_i(s)| \leq 2c_4\tau_n$ and for any s, $\sum_{i=1}^n \mathrm{var}(V_i(s)) = \sum_{i=1}^n K^2(\|\frac{s_i-s}{h}\|)D(Y_{i1}) \leq c_5nh^2$ by Condition (4) for some positive constant c_5. Then by Bernstein inequality for independent variables it follows that for any s and sufficient large n,

$$
\begin{aligned}
P(|R_1(s) - ER_1(s)| > c_3r_n) \quad &= \quad P(|\sum_{i=1}^n V_i(s)| > c_3r_nnh^2) \\
&\leq \quad 2\exp\left\{\frac{-c_3^2r_n^2n^2h^4}{2\sum_{i=1}^n \mathrm{var}(V_i(s)) + \frac{4}{3}c_3c_4\tau_nr_nnh^2}\right\} \\
&\leq \quad 2\exp\left\{\frac{-c_3^2\log n}{2c_5+4c_4}\right\} \leq 2n^{-c_3}
\end{aligned}
$$

since $(c_3/3\tau_nr_n)^2 = c_3^2/9\log^2 n/(n^{1-2/q}h^2) \to 0$ and taking $c_3 > \max\{2c_5 + 4c_4, 1\}$.

Using the same arguments, we can get that for any s and sufficient large n $P(|\tilde{R}_1(s) - E\tilde{R}_1(s)| > c_3r_n) \leq 2n^{-c_3}$. Therefore,

$$
P(\sup_{s\in\mathcal{S}}|R_1(s) - ER_1(s)| > 4c_3r_n) \leq ch^{-2}r_n^{-2}n^{-c_3} = o(1).
$$

Proof of Lemma 2:

(1) Note that

$n^{-1}H^{-1}\mathcal{X}^{\mathrm{T}}\mathcal{W}\mathcal{X}H^{-1} =$

$$\left(\begin{array}{cc} \frac{1}{n}\sum\limits_{i=1}^{n} X_i X_i^{\mathrm{T}} K_h(\|s_i - s\|) & \frac{1}{n}\sum\limits_{i=1}^{n} X_i X_i^{\mathrm{T}} \otimes (\frac{s_i - s}{h})^{\mathrm{T}} K_h(\|s_i - s\|) \\ \frac{1}{n}\sum\limits_{i=1}^{n} X_i X_i^{\mathrm{T}} \otimes \frac{s_i - s}{h} K_h(\|s_i - s\|) & \frac{1}{n}\sum\limits_{i=1}^{n} X_i X_i^{\mathrm{T}} \otimes \frac{s_i - s}{h} (\frac{s_i - s}{h})^{\mathrm{T}} K_h(\|s_i - s\|) \end{array} \right).$$

Then by Lemma 1, Lipschitz continuity of $f(\cdot)$, Condition (4) and symmetry of the kernel funtion we have that

$$n^{-1}H^{-1}\mathcal{X}^{\mathrm{T}}\mathcal{W}\mathcal{X}H^{-1} =$$
$$\left(\begin{array}{cc} \kappa_0 f(s)\Psi + O_p(\{h + r_n\}\mathbf{1}_p\mathbf{1}_p^{\mathrm{T}}) & O_p(\{h + r_n\}\mathbf{1}_p\mathbf{1}_{2p}^{\mathrm{T}}) \\ O_p(\{h + r_n\}\mathbf{1}_{2p}\mathbf{1}_p^{\mathrm{T}}) & \kappa_2 f(s)\Psi \otimes I_2 + O_p(\{h + r_n\}\mathbf{1}_{2p}\mathbf{1}_{2p}^{\mathrm{T}}) \end{array} \right)$$

holds uniformly in $s \in \mathcal{S}$.

(2) Note that

$$\boldsymbol{\beta}(s) - (I_p, \mathbf{0}_{p\times 2p})(\mathcal{X}^{\mathrm{T}}\mathcal{W}\mathcal{X})^{-1}\mathcal{X}^{\mathrm{T}}\mathcal{W}\mathbf{m} =$$
$$(I_p, \mathbf{0}_{p\times 2p})H^{-1}(n^{-1}H^{-1}\mathcal{X}^{\mathrm{T}}\mathcal{W}\mathcal{X}H^{-1})^{-1}n^{-1}H^{-1}\mathcal{X}^{\mathrm{T}}\mathcal{W}\Big\{\mathcal{X}\left(\begin{array}{c} \boldsymbol{\beta}(s) \\ \mathrm{Vec}(\dot{\boldsymbol{\beta}}^{\mathrm{T}}(s)) \end{array} \right) - \mathbf{m}\Big\}.$$

As

$$n^{-1}H^{-1}\mathcal{X}^{\mathrm{T}}\mathcal{W}\Big\{\mathcal{X}\left(\begin{array}{c} \boldsymbol{\beta}(s) \\ \mathrm{Vec}(\dot{\boldsymbol{\beta}}^{\mathrm{T}}(s)) \end{array} \right) - \mathbf{m}\Big\} =$$
$$\left(\begin{array}{c} \frac{1}{n}\sum\limits_{i=1}^{n} X_i X_i^{\mathrm{T}}\{\boldsymbol{\beta}(s) + \dot{\boldsymbol{\beta}}(s)(s_i - s) - \boldsymbol{\beta}(s_i)\}K_h(\|s_i - s\|) \\ \frac{1}{n}\sum\limits_{i=1}^{n} X_i \otimes \frac{s_i - s}{h} X_i^{\mathrm{T}}\{\boldsymbol{\beta}(s) + \dot{\boldsymbol{\beta}}(s)(s_i - s) - \boldsymbol{\beta}(s_i)\}K_h(\|s_i - s\|) \end{array} \right),$$

it follows by the second order Taylor expansion that for s_i in a small neighborhood of s,

$$\boldsymbol{\beta}(s_i) = \boldsymbol{\beta}(s) + \dot{\boldsymbol{\beta}}(s)(s_i - s) + \frac{1}{2}\left(\begin{array}{c} (s_i - s)^{\mathrm{T}}\ddot{\boldsymbol{\beta}}_1(s_i^*)(s_i - s) \\ \vdots \\ (s_i - s)^{\mathrm{T}}\ddot{\boldsymbol{\beta}}_p(s_i^*)(s_i - s) \end{array} \right),$$

where $\ddot{\boldsymbol{\beta}}_l(s) = \frac{\partial^2 \beta_l(s)}{\partial s \partial s^{\mathrm{T}}}$, $l = 1, \cdots, p$, and $s_i^* = s + \theta(s_i - s)$ with $\theta \in (0, 1)$. Then we can obtain that

$$n^{-1}H^{-1}\mathcal{X}^{\mathrm{T}}\mathcal{W}\Big\{\mathcal{X}\left(\begin{array}{c} \boldsymbol{\beta}(s) \\ \mathrm{Vec}(\dot{\boldsymbol{\beta}}^{\mathrm{T}}(s)) \end{array} \right) - \mathbf{m}\Big\} =$$

$$-\frac{h^2}{2}\sum_{l=1}^{p}\begin{pmatrix} \frac{1}{n}\sum_{i=1}^{n} X_i x_{il}(\frac{s_i-s}{h})^{\mathrm{T}}\ddot{\boldsymbol{\beta}}_l(s_i^*)(\frac{s_i-s}{h})K_h(\|s_i-s\|) \\ \frac{1}{n}\sum_{i=1}^{n} X_i\otimes(\frac{s_i-s}{h})x_{il}(\frac{s_i-s}{h})^{\mathrm{T}}\ddot{\boldsymbol{\beta}}_l(s_i^*)(\frac{s_i-s}{h})K_h(\|s_i-s\|) \end{pmatrix}.$$

Now using Lemma 1, Condition (4), symmetry of the kernel function and continuity of the second order partial derivatives of $\boldsymbol{\beta}(s)$, it is easy to show that

$$\begin{aligned} &\frac{1}{n}\sum_{i=1}^{n} X_i x_{il}(\frac{s_i-s}{h})^{\mathrm{T}}\ddot{\boldsymbol{\beta}}_l(s_i^*)(\frac{s_i-s}{h})K_h(\|s_i-s\|) \\ =\ & \kappa_2 f(s)E(X_1x_{1l})(1,0,0,1)\mathrm{Vec}(\ddot{\boldsymbol{\beta}}_l(s))+o_P(\mathbf{1}_p), \end{aligned}$$

and

$$\frac{1}{n}\sum_{i=1}^{n} X_i\otimes(\frac{s_i-s}{h})x_{il}(\frac{s_i-s}{h})^{\mathrm{T}}\ddot{\boldsymbol{\beta}}_l(s_i^*)(\frac{s_i-s}{h})K_h(\|s_i-s\|)=o_P(\mathbf{1}_{2p})$$

hold uniformly in $s\in\mathcal{S}$. Therefore,

$$\begin{aligned} & n^{-1}H^{-1}\mathcal{X}^{\mathrm{T}}\mathcal{W}\Big\{\mathcal{X}\begin{pmatrix}\boldsymbol{\beta}(s)\\ \mathrm{Vec}(\dot{\boldsymbol{\beta}}^{\mathrm{T}}(s))\end{pmatrix}-\mathrm{m}\Big\} \\ =\ & \begin{pmatrix} -\frac{1}{2}h^2\kappa_2 f(s)\Psi\Big\{\boldsymbol{\beta}_{uu}(s)+\boldsymbol{\beta}_{vv}(s)\Big\} \\ \mathbf{0}_{2p\times 1}\end{pmatrix}+o_P(h^2\mathbf{1}_{3p}) \end{aligned}$$

holds uniformly in $s\in\mathcal{S}$.

Next, it follows from Lemma 2(1) that

$$\begin{aligned} &(n^{-1}H^{-1}\mathcal{X}^{\mathrm{T}}\mathcal{W}\mathcal{X}H^{-1})^{-1}= \\ &\begin{pmatrix}\kappa_0^{-1}f^{-1}(s)\Psi^{-1} & \mathbf{0}_{p\times 2p}\\ \mathbf{0}_{2p\times p} & \kappa_2^{-1}f^{-1}(s)\Psi^{-1}\otimes I_2\end{pmatrix}+O_p(\{h+r_n\}\mathbf{1}_{3p}\mathbf{1}_{3p}^{\mathrm{T}}) \end{aligned}$$

holds uniformly in s. Hence, by the above results, we have

$$\boldsymbol{\beta}(s)-(I_p,\mathbf{0}_{p\times 2p})(\mathcal{X}^{\mathrm{T}}\mathcal{W}\mathcal{X})^{-1}\mathcal{X}^{\mathrm{T}}\mathcal{W}\mathbf{m}=-\frac{\kappa_2h^2}{2\kappa_0}\Big\{\boldsymbol{\beta}_{uu}(s)+\boldsymbol{\beta}_{vv}(s)\Big\}+o_p(h^2\mathbf{1}_p)$$

holds uniformly in $s\in\mathcal{S}$.

Proof of Lemma 3: Note that

$$n^{-1}H^{-1}\mathcal{X}^{\mathrm{T}}\mathcal{W}G\mathbf{m}=\begin{pmatrix}\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n} g_{ij}m_jX_iK_h(\|s_i-s\|)\\ \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n} g_{ij}m_jX_i\otimes\frac{s_i-s}{h}K_h(\|s_i-s\|)\end{pmatrix}.$$

In the following we will show that

$$\sup_{s\in\mathcal{S}}\Big|\frac{1}{n}\sum_{i=1}^{n}\Big\{\sum_{j=1}^{n}g_{ij}m_jX_iK_h(\|s_i-s\|)-E[\sum_{j=1}^{n}g_{ij}m_jX_iK_h(\|s_i-s\|)]\Big\}\Big|=o_P(1),$$

and

$$\sup_{s\in\mathcal{S}}\Big|\frac{1}{n}\sum_{i=1}^{n}\Big\{\sum_{j=1}^{n}g_{ij}m_jX_i\otimes\frac{s_i-s}{h}K_h(\|s_i-s\|)-E[\sum_{j=1}^{n}g_{ij}m_jX_i\otimes\frac{s_i-s}{h}K_h(\|s_i-s\|)]\Big\}\Big|=o_P(1).$$

It is obvious that these two results can be established by the same arguments, here we only show the first one. Note that

$$\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}g_{ij}m_jX_iK_h(\|s_i-s\|)\\
=\;&\frac{1}{n}\sum_{i=1}^{n}g_{ii}m_iX_iK_h(\|s_i-s\|)+\frac{1}{n}\sum_{i=1}^{n}\sum_{j\neq i}^{n}g_{ij}Em_jX_iK_h(\|s_i-s\|)\\
&+\frac{1}{n}\sum_{i=1}^{n}\sum_{j\neq i}^{n}g_{ij}(m_j-Em_j)X_iK_h(\|s_i-s\|).
\end{aligned}$$

As g_{ii} and $\sum\limits_{j\neq i}^{n}g_{ij}Em_j$ are both bounded for any i, it follows by Lemma 1 that

$$\begin{aligned}
\frac{1}{n}\sum_{i=1}^{n}g_{ii}m_iX_iK_h(\|s_i-s\|) &= \frac{1}{n}\sum_{i=1}^{n}E[g_{ii}m_iX_iK_h(\|s_i-s\|)]+O_P(r_n)\\
&= \Psi\frac{1}{n}\sum_{i=1}^{n}g_{ii}\beta(s_i)K_h(\|s_i-s\|)+o_P(1),
\end{aligned}$$

and

$$\begin{aligned}
\frac{1}{n}\sum_{i=1}^{n}\sum_{j\neq i}^{n}g_{ij}Em_jX_iK_h(\|s_i-s\|) &= \frac{1}{n}\sum_{i=1}^{n}\sum_{j\neq i}^{n}E[X_ig_{ij}Em_jK_h(\|s_i-s\|)]+O_P(r_n)\\
&= \Gamma\Gamma^{\mathrm{T}}\frac{1}{n}\sum_{i=1}^{n}\sum_{j\neq i}^{n}g_{ij}\beta(s_j)K_h(\|s_i-s\|)+o_P(1)
\end{aligned}$$

hold uniformly in $s\in\mathcal{S}$.

In the following we only need to show that for any d ($d=1,\cdots,p$)

$$\sup_{s\in\mathcal{S}}\Big|\frac{1}{n}\sum_{i=1}^{n}\sum_{j\neq i}^{n}g_{ij}(m_j-Em_j)x_{id}K_h(\|s_i-s\|)\Big|=o_P(1).$$

This result can be established using the second step in Lemma 1 where we take $r_n=(\log n)^{-1/2}$ and then Chebyshev inequality instead of Bernstein inequality.

Proof of Lemma 4: (1) It follows by Lemma 2(1) and some calculation that

$$S = \kappa_0^{-1} n^{-1}(1 + o_P(1)) \cdot \begin{pmatrix} f^{-1}(s_1)X_1^{\mathrm{T}}\Psi^{-1}X_1K_h(\|s_1 - s_1\|) & \cdots & f^{-1}(s_1)X_1^{\mathrm{T}}\Psi^{-1}X_nK_h(\|s_n - s_1\|) \\ \vdots & \vdots & \vdots \\ f^{-1}(s_n)X_n^{\mathrm{T}}\Psi^{-1}X_1K_h(\|s_1 - s_n\|) & \cdots & f^{-1}(s_n)X_n^{\mathrm{T}}\Psi^{-1}X_nK_h(\|s_n - s_n\|) \end{pmatrix} \tag{B.2}$$

As $P = (I_n - S)^{\mathrm{T}}(I_n - S) = I_n - S^{\mathrm{T}} - S + S^{\mathrm{T}}S$, we note that

$$\begin{aligned} E[\mathrm{tr}(S)] &= \frac{1}{\kappa_0 n}\sum_{i=1}^{n} E[f^{-1}(s_i)X_i^{\mathrm{T}}\Psi^{-1}X_iK_h(\|s_i - s_i\|)](1 + o(1)) \\ &= \frac{K(0)}{\kappa_0 nh^2}\sum_{i=1}^{n} E[f^{-1}(s_i)X_i^{\mathrm{T}}\Psi^{-1}X_i](1 + o(1)) \\ &= \frac{pK(0)}{\kappa_0 nh^2}\sum_{i=1}^{n} f^{-1}(s_i) = O(h^{-2}), \end{aligned}$$

hence it follows by $nh^2 \to \infty$ that $n^{-1}E[\mathrm{tr}(S)] = n^{-1}E[\mathrm{tr}(S^{\mathrm{T}})] = o(1)$.

Since by straightforward calculation we have that the (k,l)th $(k,l = 1,\cdots,n)$ element of matrix $S^{\mathrm{T}}S$ takes the form

$$\begin{aligned} [S^{\mathrm{T}}S]_{kl} &= \kappa_0^{-2}n^{-2}(1 + o_P(1)) \\ &\quad \cdot X_k^{\mathrm{T}}\{\sum_{i=1}^{n} f^{-2}(s_i)\Psi^{-1}X_iX_i^{\mathrm{T}}\Psi^{-1}K_h(\|s_k - s_i\|)K_h(\|s_l - s_i\|)\}X_l, \end{aligned}$$

and it follows by Lemma 1, continuity of $f(\cdot)$ and Condition (4) that

$$\begin{aligned} & \frac{1}{nh^2}\sum_{i=1}^{n} f^{-2}(s_i)\Psi^{-1}X_iX_i^{\mathrm{T}}\Psi^{-1}K^2(\|\frac{s_i - s}{h}\|) \\ =\ & \frac{1}{nh^2}\sum_{i=1}^{n} f^{-2}(s_i)\Psi^{-1}K^2(\|\frac{s_i - s}{h}\|) + O_P(r_n) \\ =\ & \nu_0 f^{-1}(s)\Psi^{-1}(1 + o_P(1)) \end{aligned}$$

holds uniformly in $s \in \mathcal{S}$. Thus

$$\begin{aligned} n^{-1}E[\mathrm{tr}(S^{\mathrm{T}}S)] &= \frac{\nu_0}{\kappa_0^2 n^2 h^2} E[\sum_{k=1}^{n} f^{-1}(s_k)X_k^{\mathrm{T}}\Psi^{-1}X_k](1 + o(1)) \\ &= \frac{\nu_0 p}{\kappa_0^2 n^2 h^2}\sum_{k=1}^{n} f^{-1}(s_k)(1 + o(1)) = o(1). \end{aligned}$$

Consequently, $n^{-1}\mathrm{tr}(P) = n^{-1}\mathrm{tr}(I_n) - 2n^{-1}\mathrm{tr}(S) + n^{-1}\mathrm{tr}(S^{\mathrm{T}}S) = 1 + o(1)$.

Results (2) and (3) can be established by the same arguments as in (1) and straightforward calculation.

Next it can be seen clearly from the above proof that when $nh^2/\rho_n \to \infty$, we can obtain results (4)-(6) by the fact that the elements of G having the uniform order $O(1/\rho_n)$.

Proof of Lemma 5: (1) It follows from Lemma 2(2) and some calculation that

$$
\begin{aligned}
(G\mathbf{m})^{\mathrm{T}}P\mathbf{m} &= -\frac{\kappa_2 h^2}{2\kappa_0}(G\mathbf{m})^{\mathrm{T}}(I_n - S^{\mathrm{T}})\begin{pmatrix} X_1^{\mathrm{T}}[\boldsymbol{\beta}_{uu}(s_1) + \boldsymbol{\beta}_{vv}(s_1)] \\ \vdots \\ X_n^{\mathrm{T}}[\boldsymbol{\beta}_{uu}(s_n) + \boldsymbol{\beta}_{vv}(s_n)] \end{pmatrix} \\
&\quad + (G\mathbf{m})^{\mathrm{T}}(I_n - S^{\mathrm{T}})(X_1, \cdots, X_n)^{\mathrm{T}}\mathbf{1}_p o_P(h^2).
\end{aligned}
$$

Next we use (B.2), Lemma 2(1), Lemma 1, Condition (4), continuity of $f(\cdot)$ and the second partial derivatives of $\boldsymbol{\beta}(\cdot)$ to get that

$$
\begin{aligned}
& S^{\mathrm{T}}\begin{pmatrix} X_1^{\mathrm{T}}[\boldsymbol{\beta}_{uu}(s_1) + \boldsymbol{\beta}_{vv}(s_1)] \\ \vdots \\ X_n^{\mathrm{T}}[\boldsymbol{\beta}_{uu}(s_n) + \boldsymbol{\beta}_{vv}(s_n)] \end{pmatrix} \\
= & \begin{pmatrix} X_1^{\mathrm{T}}[\boldsymbol{\beta}_{uu}(s_1) + \boldsymbol{\beta}_{vv}(s_1)] \\ \vdots \\ X_n^{\mathrm{T}}[\boldsymbol{\beta}_{uu}(s_n) + \boldsymbol{\beta}_{vv}(s_n)] \end{pmatrix} + (X_1, \cdots, X_n)^{\mathrm{T}}\mathbf{1}_p o_P(1),
\end{aligned}
$$

and

$$
S^{\mathrm{T}}(X_1, \cdots, X_n)^{\mathrm{T}}\mathbf{1}_p = (X_1^{\mathrm{T}}\mathbf{1}_p, \cdots, X_n^{\mathrm{T}}\mathbf{1}_p)^{\mathrm{T}}O_P(1). \tag{B.3}
$$

Consequently we have by Markov inequality that

$$
(G\mathbf{m})^{\mathrm{T}}P\mathbf{m} = n^{-1}(G\mathbf{m})^{\mathrm{T}}(X_1^{\mathrm{T}}\mathbf{1}_p, \cdots, X_n^{\mathrm{T}}\mathbf{1}_p)^{\mathrm{T}}o_P(nh^2) = o_P(nh^2)
$$

(2) If $f(\cdot)$ and the second partial derivatives of $\boldsymbol{\beta}(s)$ are all Lipshitz continuous, then we can obtain by Lemma 1, Condition (4) and similar arguments as

in Lemma 2(2) that

$$
\begin{aligned}
& n^{-1}H^{-1}\mathcal{X}^{\mathrm{T}}\mathcal{W}\Big\{\mathcal{X}\left(\begin{array}{c}\beta(s)\\ \mathrm{Vec}(\dot{\beta}^{\mathrm{T}}(s))\end{array}\right)-\mathbf{m}\Big\}\\
= & \left(\begin{array}{c}-\frac{1}{2}h^2\kappa_2 f(s)\Psi\Big\{\beta_{uu}(s)+\beta_{vv}(s)\Big\}\\ \mathbf{0}_{2p\times 1}\end{array}\right)+O_P(\{h^3+h^2r_n\}\mathbf{1}_{3p}).
\end{aligned}
$$

This together with Lemma 2(1) leads to

$$
\beta(s)-(I_p,\mathbf{0}_{p\times 2p})(\mathcal{X}^{\mathrm{T}}\mathcal{W}\mathcal{X})^{-1}\mathcal{X}^{\mathrm{T}}\mathcal{W}\mathbf{m}=-\frac{\kappa_2 h^2}{2\kappa_0}\Big\{\beta_{uu}(s)+\beta_{vv}(s)\Big\}+O_p(\{h^3+h^2r_n\}\mathbf{1}_p)
$$

holding uniformly in $s\in\mathcal{S}$. Hence

$$
\begin{aligned}
(G\mathbf{m})^{\mathrm{T}}P\mathbf{m} &= -\frac{\kappa_2 h^2}{2\kappa_0}(G\mathbf{m})^{\mathrm{T}}(I_n-\mathcal{S}^{\mathrm{T}})\left(\begin{array}{l}X_1^{\mathrm{T}}[\beta_{uu}(s_1)+\beta_{vv}(s_1)]\\ \vdots\\ X_n^{\mathrm{T}}[\beta_{uu}(s_n)+\beta_{vv}(s_n)]\end{array}\right)\\
&\quad +(G\mathbf{m})^{\mathrm{T}}(I_n-\mathcal{S}^{\mathrm{T}})(X_1,\cdots,X_n)^{\mathrm{T}}\mathbf{1}_p O_P(\{h^3+h^2r_n\})
\end{aligned}
$$

If $f(\cdot)$ and the second partial derivatives of $\beta(\cdot)$ are all Lipschitz continuous, then

$$
\begin{aligned}
& \mathcal{S}^{\mathrm{T}}\left(\begin{array}{c}X_1^{\mathrm{T}}[\beta_{uu}(s_1)+\beta_{vv}(s_1)]\\ \vdots\\ X_n^{\mathrm{T}}[\beta_{uu}(s_n)+\beta_{vv}(s_n)]\end{array}\right)\\
= & \left(\begin{array}{c}X_1^{\mathrm{T}}[\beta_{uu}(s_1)+\beta_{vv}(s_1)]\\ \vdots\\ X_n^{\mathrm{T}}[\beta_{uu}(s_n)+\beta_{vv}(s_n)]\end{array}\right)+(X_1,\cdots,X_n)^{\mathrm{T}}\mathbf{1}_p O_P(h+r_n).
\end{aligned}
$$

Therefore, we have by (B.3) and Markov inequality that

$$
\begin{aligned}
(G\mathbf{m})^{\mathrm{T}}P\mathbf{m} &= n^{-1}(G\mathbf{m})^{\mathrm{T}}(X_1^{\mathrm{T}}\mathbf{1}_p,\cdots,X_n^{\mathrm{T}}\mathbf{1}_p)^{\mathrm{T}}O_P(n\{h^3+h^2r_n\})\\
&= O_P(nh^3+\{nh^2\log n\}^{1/2})
\end{aligned}
$$

Proof of Lemma 6: (1) In the following, we will show that $n^{-1/2}L^{\mathrm{T}}P\mathbf{m}=o_P(1)$ for $L=\mathbf{m},\ \epsilon$ and $G\epsilon$.

Note that $n^{-1/2}\mathbf{m}^{\mathrm{T}}P\mathbf{m} = n^{-1/2}(\mathbf{m} - S\mathbf{m})^{\mathrm{T}}(\mathbf{m} - S\mathbf{m})$, and it follows by Lemma 2(2) that

$$\mathbf{m} - S\mathbf{m} = (X_1, \cdots, X_n)^{\mathrm{T}}\mathbf{1}_p O_P(h^2). \tag{B.4}$$

Therefore,

$$n^{-1/2}\mathbf{m}^{\mathrm{T}}P\mathbf{m} = n^{-1}\sum_{i=1}^{n}(X_i^{\mathrm{T}}\mathbf{1}_p)^2 O_P(n^{1/2}h^4) = o_P(1)$$

using law of large numbers and $nh^8 \to 0$.

Since we have by (B.3), (B.4) and Chebyshev inequality that

$$\begin{aligned} n^{-1/2}\boldsymbol{\epsilon}^{\mathrm{T}}P\mathbf{m} &= n^{-1/2}(\boldsymbol{\epsilon} - S\boldsymbol{\epsilon})^{\mathrm{T}}(\mathbf{m} - S\mathbf{m}) \\ &= \{n^{-1/2}\boldsymbol{\epsilon}^{\mathrm{T}}(X_1, \cdots, X_n)^{\mathrm{T}}\mathbf{1}_p - n^{-1/2}\boldsymbol{\epsilon}^{\mathrm{T}}S^{\mathrm{T}}(X_1, \cdots, X_n)^{\mathrm{T}}\mathbf{1}_p\}O_P(h^2) \\ &= n^{-1/2}\sum_{i=1}^{n}X_i^{\mathrm{T}}\mathbf{1}_p\epsilon_i O_P(h^2) = O_P(h^2), \end{aligned}$$

Hence $n^{-1/2}\boldsymbol{\epsilon}^{\mathrm{T}}P\mathbf{m} = o_P(1)$.

Similarly, we can show that $n^{-1/2}(G\epsilon)^{\mathrm{T}}P\mathbf{m} = O_P(h^2) = o_P(1)$.

(2) Here, we will show that $n^{-1}L^{\mathrm{T}}PG\mathbf{m} = o_P(1)$ for $L = \mathbf{m}, \epsilon$ and $G\boldsymbol{\epsilon}$.

Clearly, it follows by Lemma 5(1) that $n^{-1}\mathbf{m}^{\mathrm{T}}PG\mathbf{m} = o_P(h^2) = o_P(1)$.

For simplification, in the following we set $\tilde{X} = (X_1^{\mathrm{T}}\mathbf{1}_p, \cdots, X_n^{\mathrm{T}}\mathbf{1}_p)^{\mathrm{T}}$ and $V = (f^{-1}(s_1)X_1^{\mathrm{T}}\Psi^{-1}\mathbf{1}_p, \cdots, f^{-1}(s_n)X_n^{\mathrm{T}}\Psi^{-1}\mathbf{1}_p)^{\mathrm{T}}$.

Note that

$$\frac{1}{n}\boldsymbol{\epsilon}^{\mathrm{T}}PG\mathbf{m} = \frac{1}{n}\boldsymbol{\epsilon}^{\mathrm{T}}G\mathbf{m} - \frac{1}{n}\boldsymbol{\epsilon}^{\mathrm{T}}S^{\mathrm{T}}G\mathbf{m} - \frac{1}{n}\boldsymbol{\epsilon}^{\mathrm{T}}SG\mathbf{m} + \frac{1}{n}\boldsymbol{\epsilon}^{\mathrm{T}}S^{\mathrm{T}}SG\mathbf{m}.$$

As $E(\frac{1}{n}\epsilon^{\mathrm{T}}G\mathrm{m}) = 0$, and

$$\mathrm{var}(\frac{1}{n}\epsilon^{\mathrm{T}}G\mathrm{m}) = \frac{\sigma_0^2}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}g_{ij}^2Em_j^2 + \frac{\sigma_0^2}{n^2}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k\neq j}^{n}g_{ij}g_{ik}Em_jEm_k = O(\frac{1}{n}),$$

we obtain by Chebyshev inequality that $n^{-1}\boldsymbol{\epsilon}^{\mathrm{T}}G\mathbf{m} = o_P(1)$.

It follows by (B.2), Lipschitz continuity of $f(\cdot)$, Lemma 3 and Condition (4) that

$$S^{\mathrm{T}}G\mathbf{m} = \{Z + V \cdot o_P(1) + \tilde{X} \cdot o_P(1)\}(1 + o_P(1)). \tag{B.5}$$

Therefore,

$$\frac{1}{n}\epsilon^{\mathrm{T}}S^{\mathrm{T}}G\mathbf{m} = \{\frac{1}{n}\epsilon^{\mathrm{T}}Z + \frac{1}{n}\epsilon^{\mathrm{T}}V \cdot o_P(1) + \frac{1}{n}\epsilon^{\mathrm{T}}\tilde{X} \cdot o_P(1)\}(1 + o_P(1)) = o_P(1)$$

by law of large numbers.

Similarly,

$$SG\mathbf{m} = \{Z + V \cdot o_P(1)\}(1 + o_P(1)). \tag{B.6}$$

by Lemma 3 and Condition (4). Therefore, we have by law of large numbers that

$$\frac{1}{n}\epsilon^{\mathrm{T}}SG\mathbf{m} = \{\frac{1}{n}\epsilon^{\mathrm{T}}Z + \frac{1}{n}\epsilon^{\mathrm{T}}V \cdot o_P(1)\}(1 + o_P(1)) = o_P(1).$$

Next it follows by (B.2) and Lemma 1 that $S\epsilon = V \cdot o_P(1)$. This together with (B.6), we obtain that

$$\frac{1}{n}(S\epsilon)^{\mathrm{T}}SG\mathbf{m} = \{\frac{1}{n}V^{\mathrm{T}}Z + \frac{1}{n}V^{\mathrm{T}}V\}o_P(1).$$

Therefore, $n^{-1}\epsilon^{\mathrm{T}}S^{\mathrm{T}}SG\mathbf{m} = o_P(1)$ by law of large numbers.

Similarly, we can show that $n^{-1}(G\epsilon)^{\mathrm{T}}PG\mathbf{m} = o_P(1)$.

Proof of Lemma 7: (1) It can be seen that

$$\frac{1}{n}(G\mathbf{m})^{\mathrm{T}}PG\mathbf{m} = \frac{1}{n}(G\mathbf{m})^{\mathrm{T}}G\mathbf{m} - \frac{2}{n}(G\mathbf{m})^{\mathrm{T}}SG\mathbf{m} + \frac{1}{n}(SG\mathbf{m})^{\mathrm{T}}SG\mathbf{m}.$$

For the first term we have that

$$\frac{1}{n}(G\mathbf{m})^{\mathrm{T}}G\mathbf{m} = \frac{1}{n}\sum_{j=1}^{n}(\sum_{i=1}^{n}g_{ij}^2)m_j^2 + \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k\neq j}^{n}g_{ij}g_{ik}m_jm_k,$$

and

$$\mathrm{var}\{\frac{1}{n}\sum_{j=1}^{n}(\sum_{i=1}^{n}g_{ij}^2)m_j^2\} = \frac{1}{n^2}\sum_{j=1}^{n}(\sum_{i=1}^{n}g_{ij}^2)^2D(m_j^2) = O(\frac{1}{n}),$$

it follows by Chebyshev inequality that $\frac{1}{n}\sum_{j=1}^{n}\sum_{i=1}^{n}g_{ij}^2m_j^2 - E\{\frac{1}{n}\sum_{j=1}^{n}\sum_{i=1}^{n}g_{ij}^2m_j^2\} = o_P(1)$.

Let $\bar{m}_i = m_i - Em_i,\ i = 1, \cdots, n$, then

$$\begin{aligned}
&\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k\neq j}^{n} g_{ij}g_{ik}m_jm_k - E\{\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k\neq j}^{n} g_{ij}g_{ik}m_jm_k\} \\
= &\frac{1}{n}\sum_{j=1}^{n}\sum_{k\neq j}^{n}\sum_{i=1}^{n} g_{ij}g_{ik}\bar{m}_j\bar{m}_k + \frac{1}{n}\sum_{j=1}^{n}\sum_{k\neq j}^{n}\sum_{i=1}^{n} g_{ij}g_{ik}\bar{m}_jEm_j \\
&+\frac{1}{n}\sum_{k=1}^{n}\sum_{j\neq k}^{n}\sum_{i=1}^{n} g_{ij}g_{ik}\bar{m}_kEm_k.
\end{aligned}$$

Define $J_{n1} = \frac{1}{n}\sum_{j=1}^{n}\sum_{k\neq j}^{n}\sum_{i=1}^{n} g_{ij}g_{ik}\bar{m}_j\bar{m}_k$, $J_{n2} = \frac{1}{n}\sum_{j=1}^{n}\sum_{k\neq j}^{n}\sum_{i=1}^{n} g_{ij}g_{ik}\bar{m}_jEm_j$ and $J_{n3} = \frac{1}{n}\sum_{k=1}^{n}\sum_{j\neq k}^{n}\sum_{i=1}^{n} g_{ij}g_{ik}\bar{m}_kEm_k$, with

$$\begin{aligned}
\mathrm{var}(J_{n1}) &= E(J_{n1}^2) \\
&= \frac{2}{n^2}\sum_{j=1}^{n}\sum_{k\neq j}^{n}(\sum_{i=1}^{n} g_{ij}g_{ik})^2[\beta^{\mathrm{T}}(s_j)D(X_1)\beta(s_j)][\beta^{\mathrm{T}}(s_k)D(X_1)\beta(s_k)] \\
&\leq \max_{j,k}(\sum_{i=1}^{n}|g_{ij}g_{ik}|)\frac{2}{n^2}\sum_{j=1}^{n}\{\sum_{i=1}^{n}|g_{ij}|\beta^{\mathrm{T}}(s_j)D(X_1)\beta(s_j)\}^2 = O(\frac{1}{n}),
\end{aligned}$$

$$\mathrm{var}(J_{n2}) = \frac{1}{n^2}\sum_{j=1}^{n}(\sum_{k\neq j}^{n}\sum_{i=1}^{n} g_{ij}g_{ik})^2D(m_j)(Em_j)^2 = O(\frac{1}{n}),$$

and

$$\mathrm{var}(J_{n3}) = \frac{1}{n^2}\sum_{k=1}^{n}(\sum_{j\neq k}^{n}\sum_{i=1}^{n} g_{ij}g_{ik})^2D(m_k)(Em_k)^2 = O(\frac{1}{n}).$$

Therefore, by Chebyshev inequality that $J_{n1} = o_P(1)$, $J_{n2} = o_P(1)$ and $J_{n3} = o_P(1)$. In conclusion, we obtain that $\frac{1}{n}(G\mathbf{m})^{\mathrm{T}}G\mathbf{m} - E\{\frac{1}{n}(G\mathbf{m})^{\mathrm{T}}G\mathbf{m}\} = o_P(1)$.

It follows from (B.6) that

$$\frac{1}{n}(G\mathbf{m})^{\mathrm{T}}SG\mathbf{m} = \{\frac{1}{n}(G\mathbf{m})^{\mathrm{T}}Z + \frac{1}{n}(G\mathbf{m})^{\mathrm{T}}V\cdot o_P(1)\}(1+o_P(1)).$$

Using similar arguments as establishing $\frac{1}{n}(G\mathbf{m})^{\mathrm{T}}G\mathbf{m} - E\{\frac{1}{n}(G\mathbf{m})^{\mathrm{T}}G\mathbf{m}\} = o_P(1)$, we have that $\frac{1}{n}(G\mathbf{m})^TL - E\{\frac{1}{n}(G\mathbf{m})^TL\} = o_P(1)$ for $L = Z$ and V. Moreover, $E\{\frac{1}{n}(G\mathbf{m})^{\mathrm{T}}SG\mathbf{m}\} = E\{\frac{1}{n}(G\mathbf{m})^{\mathrm{T}}Z\} + o(1)$. Therefore,

$$\frac{1}{n}(G\mathbf{m})^{\mathrm{T}}SG\mathbf{m} - E\{\frac{1}{n}(G\mathbf{m})^{\mathrm{T}}SG\mathbf{m}\} = o_P(1).$$

For the term $\frac{1}{n}(SG\mathbf{m})^{\mathrm{T}}SG\mathbf{m}$, again by (B.6) we have that

$$\begin{aligned}\frac{1}{n}(SG\mathbf{m})^{\mathrm{T}}SG\mathbf{m} &= \Big\{\frac{1}{n}Z^{\mathrm{T}}Z+\frac{1}{n}V^{\mathrm{T}}V\cdot o_P(1)+\frac{2}{n}Z^{\mathrm{T}}V\cdot o_P(1)\Big\}(1+o_P(1)) \\ &= \frac{1}{n}E(Z^{\mathrm{T}}Z)+o_P(1)\end{aligned}$$

by law of large numbers, and $E\{\frac{1}{n}(SG\mathbf{m})^{\mathrm{T}}SG\mathbf{m}\}=\frac{1}{n}E(Z^{\mathrm{T}}Z)+o(1)$. Thus

$$\frac{1}{n}(SG\mathbf{m})^{\mathrm{T}}SG\mathbf{m}-E\{\frac{1}{n}(SG\mathbf{m})^{\mathrm{T}}SG\mathbf{m}\}=o_P(1).$$

In conclusion, we obtain that $\frac{1}{n}(G\mathbf{m})^{\mathrm{T}}PG\mathbf{m}-E\{\frac{1}{n}(G\mathbf{m})^{\mathrm{T}}PG\mathbf{m}\}=o_P(1)$.

(2) We have seen from (B.6) that

$$\begin{aligned}\frac{1}{n}E\{(G\mathbf{m})^{\mathrm{T}}PG\mathbf{m}\} &= \frac{1}{n}E\{(G\mathbf{m}-Z+V\cdot o_P(1))^{\mathrm{T}}(G\mathbf{m}-Z+V\cdot o_P(1))\}(1+o(1)) \\ &= \frac{1}{n}E[(G\mathbf{m}-Z)^{\mathrm{T}}(G\mathbf{m}-Z)]+o(1).\end{aligned}$$

Proof of Lemma 8: (1) Since

$$n^{-1/2}\{\epsilon^{\mathrm{T}}P\epsilon-\epsilon^{\mathrm{T}}\epsilon\}=-2n^{-1/2}\epsilon^{\mathrm{T}}S\epsilon+n^{-1/2}\epsilon^{\mathrm{T}}S^{\mathrm{T}}S\epsilon,$$

$E[n^{-1/2}\epsilon^{\mathrm{T}}S\epsilon]=\sigma_0^2n^{-1/2}E[\mathrm{tr}(S)]=O(\{nh^4\}^{-1/2})=o(1)$, and

$$\begin{aligned}\mathrm{var}(n^{-1/2}\epsilon^{\mathrm{T}}S\epsilon) &\le \frac{1}{n}E(\epsilon^{\mathrm{T}}S\epsilon)^2 \\ &= \frac{1}{n}\Big[(\mu_4-3\sigma_0^4)\sum_{i=1}^{n}E[S]_{ii}^2+\sigma_0^4E\{[\mathrm{tr}(S)]^2+\mathrm{tr}(SS^{\mathrm{T}})+\mathrm{tr}(S^2)\}\Big]\end{aligned}$$

It can be seen from the proof of Lemma 4(1) that $n^{-1}E[\mathrm{tr}(SS^{\mathrm{T}})]=o(1)$,

$$\frac{1}{n}\sum_{i=1}^{n}E[S]_{ii}^2=\frac{1}{\kappa_0^2n^3h^4}\sum_{i=1}^{n}E[f^{-2}(s_i)(X_i^{\mathrm{T}}\Psi^{-1}X_i)^2]K^2(0)=O(\frac{1}{n^2h^4})=o(1),$$

and

$$\mathrm{tr}(S)=\frac{pK(0)}{\kappa_0nh^2}\sum_{i=1}^{n}f^{-1}(s_i)+o_p(1).$$

It can be seen that $n^{-1/2}\mathrm{tr}(S)=O_p(\{nh^4\}^{-1/2})=o_P(1)$. Hence, $n^{-1}[\mathrm{tr}(S)]^2=o_P(1)$ and $n^{-1}E\{[\mathrm{tr}(S)]^2\}=o(1)$.

It follows by straightforward calculation, Lemma 1 and Condition (4) that

$$\begin{aligned}[S^2]_{ii} &= \frac{1+o_P(1)}{\kappa_0^2n^2}f^{-1}(s_i)X_i^{\mathrm{T}}\Psi^{-1}\sum_{j=1}^{n}f^{-1}(s_j)X_jX_j^{\mathrm{T}}\Psi^{-1}K_h^2(\|s_j-s_i\|)X_i \\ &= \frac{\nu_0}{\kappa_0^2nh^2}f^{-1}(s_i)X_i^{\mathrm{T}}\Psi^{-1}X_i(1+o_P(1)).\end{aligned}$$

Thus

$$\frac{1}{n}E[\mathrm{tr}(S^2)] = \frac{\nu_0(1+o(1))}{\kappa_0^2 n^2 h^2}\sum_{i=1}^n E[f^{-1}(s_i)X_i^{\mathrm{T}}\Psi^{-1}X_i] = O(\frac{1}{nh^2}) = o(1).$$

Consequently, we have by Chebyshev inequality that $n^{-1/2}\epsilon^{\mathrm{T}}S\epsilon = o_P(1)$.

Similarly, it can be shown that $n^{-1/2}\epsilon^{\mathrm{T}}S^{\mathrm{T}}S\epsilon = o_P(1)$. Hence we have shown that $n^{-1/2}(\epsilon^{\mathrm{T}}P\epsilon - \epsilon^{\mathrm{T}}\epsilon) = o_P(1)$.

Results (2) and (3) can be obtained by the same arguments as in (1) and straightforward calculation.

(4) Note that

$$n^{-1/2}\{(G\mathbf{m})^{\mathrm{T}}P\epsilon - (G\mathbf{m} - SG\mathbf{m})^{\mathrm{T}}\epsilon\} = -n^{-1/2}(G\mathbf{m} - SG\mathbf{m})^{\mathrm{T}}S\epsilon.$$

Moreover, $E[n^{-1/2}(G\mathbf{m} - SG\mathbf{m})^{\mathrm{T}}S\epsilon] = 0$ and $\mathrm{var}[n^{-1/2}(G\mathbf{m} - SG\mathbf{m})^{\mathrm{T}}S\epsilon] = \sigma_0^2 n^{-1}E[(G\mathbf{m} - SG\mathbf{m})^{\mathrm{T}}SS^{\mathrm{T}}(G\mathbf{m} - SG\mathbf{m})]$. As it follows by (B.6), Lemma 1 and Condition (4) that

$$\begin{aligned} S^{\mathrm{T}}SG\mathbf{m} &= \{S^{\mathrm{T}}Z + S^{\mathrm{T}}V\cdot o_P(1)\}(1+o_P(1)) \\ &= \{Z + \tilde{X}\cdot o_P(1) + V\cdot o_P(1)\}(1+o_P(1)) \end{aligned}$$

where $V = (f^{-1}(s_1)X_1^{\mathrm{T}}\Psi^{-1}\mathbf{1}_p, \cdots, f^{-1}(s_n)X_n^{\mathrm{T}}\Psi^{-1}\mathbf{1}_p)^{\mathrm{T}}$ and $\tilde{X} = (X_1^{\mathrm{T}}\mathbf{1}_p, \cdots, X_n^{\mathrm{T}}\mathbf{1}_p)^{\mathrm{T}}$. This together with (B.5) it can be seen that

$$S^{\mathrm{T}}G\mathbf{m} - S^{\mathrm{T}}SG\mathbf{m} = \{Z + \tilde{X} + V\}o_P(1).$$

Hence $n^{-1}E[(G\mathbf{m} - SG\mathbf{m})^{\mathrm{T}}SS^{\mathrm{T}}(G\mathbf{m} - SG\mathbf{m})] = o(1)$. Consequently, it can be obtained by Chebyshev inequality that $n^{-1/2}(G\mathbf{m} - SG\mathbf{m})^{\mathrm{T}}S\epsilon = o_P(1)$. Therefore, $n^{-1/2}\{(G\mathbf{m})^{\mathrm{T}}P\epsilon - (G\mathbf{m} - SG\mathbf{m})^{\mathrm{T}}\epsilon\} = o_P(1)$.

Proof of Lemma 9: The asymptotic distribution of the linear-quadratic random form Q_n can be established via the martingale central limit theorem. Our proof of this lemma follows closely the arguments in Kelejian and Prucha (2001) and Lee (2004).

Note that

$$Q_n = \sum_{i=1}^n(\sum_{j=1}^n g_{ij}m_j - g_i^s)\epsilon_i + \sum_{i=1}^n b_{ii}\epsilon_i^2 + 2\sum_{i=1}^n\sum_{k=1}^{i-1}b_{ik}\epsilon_i\epsilon_k - \sigma_0^2\mathrm{tr}(B) = \sum_{i=1}^n V_{ni}$$

where g_i^s is the ith element of $SG\mathbf{m}$ and $V_{ni} = (\sum_{j=1}^{n} g_{ij}m_j - g_i^s)\epsilon_i + b_{ii}(\epsilon_i^2 - \sigma_0^2) + 2\epsilon_i \sum_{k=1}^{i-1} b_{ik}\epsilon_k$.

Define $\sigma-$ fields $\mathcal{T}_i =< \epsilon_1, \cdots, \epsilon_i >$ generated by $\epsilon_1, \cdots, \epsilon_i$. Because $\{\epsilon_i\}_{i=1}^n$ are iid with zero mean, finite variance and independent with $\{X_j\}_{j=1}^n$,

$$E(V_{ni}|\mathcal{T}_{i-1}) = E(\sum_{j=1}^{n} g_{ij}m_j - g_i^s)E\epsilon_i + b_{ii}(E\epsilon_i^2 - \sigma_0^2) + 2E\epsilon_i \sum_{k=1}^{i-1} b_{ik}\epsilon_k = 0.$$

Hence, the $\{(V_{ni}, \mathcal{T}_i)|1 \leq i \leq n\}$ forms a martingale difference double array and $\sigma_{Q_n}^2 = \sum_{i=1}^{n} E(V_{ni}^2)$ with $\sigma_{Q_n}^2$ being bounded away from zero at n rate. Define the normalized variables $V_{ni}^* = V_{ni}/\sigma_{Q_n}$. Then $\{(V_{ni}^*, \mathcal{T}_i)|1 \leq i \leq n\}$ is a martingale difference double array and $\frac{Q_n}{\sigma_{Q_n}} = \sum_{i=1}^{n} V_{ni}^*$. In order for the martingale central limit theorem to be applicable we would show that there exists a $\delta > 0$ such that $\sum_{i=1}^{n} E|V_{ni}^*|^{2+\delta} = o(1)$ and $\sum_{i=1}^{n} E(V_{ni}^{*2}|\mathcal{T}_{i-1}) \xrightarrow{P} 1$.

For any positive constant p and q satisfying $\frac{1}{p} + \frac{1}{q} = 1$, we have

$$\begin{aligned}
|V_{ni}| &\leq |b_{ii}| \cdot |\epsilon_i^2 - \sigma_0^2| + |\epsilon_i|(|\sum_{j=1}^{n} g_{ij}m_j - g_i^s| + 2\sum_{k=1}^{i-1} |b_{ik}| \cdot |\epsilon_k|) \\
&= |b_{ii}|^{\frac{1}{p}}(|b_{ii}|^{\frac{1}{q}} \cdot |\epsilon_i^2 - \sigma_0^2|) + |\sum_{j=1}^{n} g_{ij}m_j - g_i^s|^{\frac{1}{p}}(|\sum_{j=1}^{n} g_{ij}m_j - g_i^s|^{\frac{1}{q}}|\epsilon_i|) \\
&\quad + \sum_{k=1}^{i-1} |b_{ik}|^{\frac{1}{p}}(|b_{ik}|^{\frac{1}{q}} 2|\epsilon_k| \cdot |\epsilon_i|).
\end{aligned}$$

Applying Holder inequality we obtain that

$$\begin{aligned}
|V_{ni}|^q &\leq \Big[|\sum_{j=1}^{n} g_{ij}m_j - g_i^s| + \sum_{k=1}^{i} |b_{ik}|\Big]^{\frac{q}{p}} \\
&\quad \cdot \Big[|\sum_{j=1}^{n} g_{ij}m_j - g_i^s| \cdot |\epsilon_i|^q + |b_{ii}| \cdot |\epsilon_i^2 - \sigma_0^2|^q + \sum_{k=1}^{i-1} |b_{ik}| 2^q |\epsilon_i|^q |\epsilon_k|^q\Big]
\end{aligned}$$

Let $c_1 > 1$ be a finite constant such that $E(|\epsilon_1^2 - \sigma_0^2|) \leq c_1$, $E|\epsilon_1|^q \leq c_1$, and $(E|\epsilon_1|^q)^2 \leq c_1$. Set $\mathcal{D} = \{X_i\}_{i=1}^n$, we have

$$E[|V_{ni}|^q|\mathcal{D}] \leq 2^q c_1 \Big[|\sum_{j=1}^{n} g_{ij}m_j - g_i^s| + \sum_{k=1}^{i} |b_{ik}|\Big]^q$$

As the the matrix B are uniformly bounded in row sums, there exists a constant c_2 such that $\sum_{j=1}^n |b_{ij}| \le c_2$ for all i. Take $q = 2+\delta$, it follows by Cr inequality and (B.6) that

$$\begin{aligned}
&\sum_{i=1}^n E[|V_{ni}|^{2+\delta}] = \sum_{i=1}^n E\{E[|V_{ni}|^{2+\delta}|\mathcal{D}]\} \\
\le\;& c_1 2^{3+2\delta} \sum_{i=1}^n \{E|\sum_{j=1}^n g_{ij}m_j - g_i^s|^{2+\delta} + (\sum_{k=1}^i |b_{ik}|)^{2+\delta}\} \\
\le\;& c_1 2^{3+2\delta}\{2^{2+2\delta} \sum_{i=1}^n \Big[E|\sum_{j=1}^n g_{ij}(m_j - Em_j)|^{2+\delta} + |\sum_{j=1}^n g_{ij}Em_j|^{2+\delta}\Big] + cn\}.
\end{aligned}$$

Because $\{m_j\}$ are independent variables, we have that

$$\begin{aligned}
&E|\sum_{j=1}^n g_{ij}(m_j - Em_j)|^{2+\delta} \\
\le\;& c\{\sum_{j=1}^n E|g_{ij}(m_j - Em_j)|^{2+\delta} + (\sum_{j=1}^n E[g_{ij}(m_j - Em_j)]^2)^{\frac{2+\delta}{2}}\} \le c
\end{aligned}$$

by $\sum_{j=1}^n |g_{ij}|$ being uniformly bounded for all i. Therefore, $\sum_{i=1}^n E[|V_{ni}|^{2+\delta}] = O(n)$. Hence $\sum_{i=1}^n E|V_{ni}^*|^{2+\delta} = \frac{1}{(\sigma_{Q_n}^2)^{\frac{2+\delta}{2}}} \sum_{i=1}^n E|V_{ni}|^{2+\delta} = O(\frac{n}{n^{1+\delta/2}}) = o(1)$.

It remains to show that $\sum_{i=1}^n E(V_{ni}^{*2}|\mathcal{T}_{i-1}) \xrightarrow{P} 1$. As $E(V_{ni}^2|\mathcal{D}, \mathcal{T}_{i-1}) = (\mu_4 - \sigma_0^4)b_{ii}^2 + [(\sum_{j=1}^n g_{ij}m_j - g_i^s) + 2\sum_{k=1}^{i-1} b_{ik}\epsilon_k]^2\sigma_0^2 + 2\mu_3 b_{ii}[(\sum_{j=1}^n g_{ij}m_j - g_i^s) + 2\sum_{k=1}^{i-1} b_{ik}\epsilon_k]$, it follows that

$$\begin{aligned}
&E(V_{ni}^2|\mathcal{T}_{i-1}) - E(V_{ni}^2) \\
=\;& 4\sigma_0^2\{\sum_{k=1}^{i-1} b_{ik}^2(\epsilon_k^2 - \sigma_0^2) + \sum_{k=1}^{i-1}\sum_{l\ne k}^{i-1} b_{ik}b_{il}\epsilon_k\epsilon_l\} + 4[\sigma_0^2 E(\sum_{j=1}^n g_{ij}m_j - g_i^s) + \mu_3 b_{ii}]\sum_{k=1}^{i-1} b_{ik}\epsilon_k
\end{aligned}$$

Therefore,

$$\begin{aligned}
\sum_{i=1}^n E(V_{ni}^{*2}|\mathcal{T}_{i-1}) - 1 &= \frac{1}{\sigma_{Q_n}^2}\sum_{i=1}^n [E(V_{ni}^2|\mathcal{T}_{i-1}) - E(V_{ni}^2)] \\
&= \frac{4\sigma_0^2}{\frac{1}{n}\sigma_{Q_n}^2} \cdot \frac{1}{n}\sum_{i=1}^n\{\sum_{k=1}^{i-1} b_{ik}^2(\epsilon_k^2 - \sigma_0^2) + \sum_{k=1}^{i-1}\sum_{l\ne k}^{i-1} b_{ik}b_{il}\epsilon_k\epsilon_l\} \\
&\quad + \frac{4}{\frac{1}{n}\sigma_{Q_n}^2} \cdot \frac{1}{n}\sum_{i=1}^n [\sigma_0^2 E(\sum_{j=1}^n g_{ij}m_j - g_i^s) + \mu_3 b_{ii}]\sum_{k=1}^{i-1} b_{ik}\epsilon_k \\
&= \frac{4\sigma_0^2}{\frac{1}{n}\sigma_{Q_n}^2}(J_{n1} + J_{n2}) + \frac{4}{\frac{1}{n}\sigma_{Q_n}^2} J_{n3}
\end{aligned}$$

with $J_{n1} = \frac{1}{n}\sum_{i=1}^{n}\sum_{k=1}^{i-1} b_{ik}^2(\epsilon_k^2 - \sigma_0^2)$, $J_{n2} = \frac{1}{n}\sum_{i=1}^{n}\sum_{k=1}^{i-1}\sum_{l\neq k}^{i-1} b_{ik}b_{il}\epsilon_k\epsilon_l$, and $J_{n3} = \frac{1}{n}\sum_{i=1}^{n}[\sigma_0^2 E(\sum_{j=1}^{n} g_{ij}m_j - g_i^s) + \mu_3 b_{ii}]\sum_{k=1}^{i-1} b_{ik}\epsilon_k$.

Clearly, $EJ_{nl} = 0, l = 1, 2, 3$. By Chebyshev inequality, to show $J_{nl} = o_P(1)$, it is only need to prove $EJ_{nl}^2 = o(1)$. It is obvious by uniform boundness of b_{ik} and uniform boundness of $\sum_{i=1}^{n}|b_{ik}|$ that $E(J_{n1}^2) = \frac{1}{n^2}\sum_{k=1}^{n-1}(\sum_{i=k+1}^{n} b_{ik}^2)^2 D(\epsilon_1^2) \leq \frac{1}{n^2}D(\epsilon_1^2)\max_{i,k}|b_{ik}|^2\sum_{k=1}^{n-1}(\sum_{i=k+1}^{n}|b_{ik}|)^2 = O(\frac{1}{n})$.

Since $J_{n2} = \frac{1}{n}\sum_{k=1}^{n-1}\sum_{l\neq k}^{n-1}(\sum_{i=\max\{k,l\}+1}^{n} b_{ik}b_{il})\epsilon_k\epsilon_l$, we have

$$\begin{aligned} E(J_{n2}^2) &= \frac{2\sigma_0^4}{n^2}\sum_{k=1}^{n-1}\sum_{l\neq k}^{n-1}(\sum_{i=\max\{k,l\}+1}^{n} b_{ik}b_{il})^2 \leq \frac{2\sigma_0^4}{n^2}\sum_{k=1}^{n}\sum_{l=1}^{n}(\sum_{i=1}^{n}|b_{ik}b_{il}|)^2 \\ &\leq \frac{2\sigma_0^4}{n^2}\max_{i,l}|b_{il}|\max_k\sum_{i=1}^{n}|b_{ik}|\sum_{i=1}^{n}\sum_{k=1}^{n}\sum_{l=1}^{n}|b_{ik}b_{il}| = O(\frac{1}{n}) \end{aligned}$$

As J_{n3} can be written as $J_{n3} = \frac{1}{n}\sum_{k=1}^{n-1}[\sum_{i=k+1}^{n}(\sigma_0^2 E[\sum_{j=1}^{n} g_{ij}m_j - g_i^s] + \mu_3 b_{ii})b_{ik}]\epsilon_k$, it follows that

$$\begin{aligned} E(J_{n3}^2) &= \frac{\sigma_0^2}{n^2}\sum_{k=1}^{n-1}[\sum_{i=k+1}^{n}(\sigma_0^2 E[\sum_{j=1}^{n} g_{ij}m_j - g_i^s] + \mu_3 b_{ii})b_{ik}]^2 \\ &\leq \frac{\sigma_0^2}{n^2}\max_i\{\sigma_0^2|E[\sum_{j=1}^{n} g_{ij}m_j - g_i^s]| + \mu_3|b_{ii}|\}^2\sum_{k=1}^{n}(\sum_{i=1}^{n}|b_{ik}|)^2 = O(\frac{1}{n}) \end{aligned}$$

by $|E[\sum_{j=1}^{n} g_{ij}m_j - g_i^s]| \leq \sum_{j=1}^{n}|g_{ij}Em_j| + E|g_i^s| = O(1)$ for any i, where $E|g_i^s| = O(1)$ is obtained using (B.6).

Because $J_{nl} = o_P(1)$ for $l = 1, 2, 3$ and $\lim_{n\to\infty}\frac{\sigma_{Q_n}^2}{n} > 0$, $\sum_{i=1}^{n} E(V_{ni}^{*2}|\mathcal{T}_{i-1})$ converges in probability to 1. The central limit theorem for martingale difference double array is thus applicable to establish the result.

Proof of Lemma 10: (1) Here we will show that $(G\mathbf{m})^{\mathrm{T}}P\mathbf{m} = o_P(\rho_n^{-1/2}nh^2)$. It can be seen from the proof of Lemma 5(1) that

$$(G\mathbf{m})^{\mathrm{T}}P\mathbf{m} = (G\mathbf{m})^{\mathrm{T}}(X_1^{\mathrm{T}}\mathbf{1}_p, \cdots X_n^{\mathrm{T}}\mathbf{1}_p)^{\mathrm{T}}o_P(h^2).$$

As $\frac{\sqrt{\rho_n}}{n}(G\mathrm{m})^{\mathrm{T}}(X_1^{\mathrm{T}}\mathbf{1}_p, \cdots X_n^{\mathrm{T}}\mathbf{1}_p)^{\mathrm{T}} = \frac{\sqrt{\rho_n}}{n}\sum_{i=1}^{n}\sum_{j=1}^{n} g_{ij}m_jX_i^{\mathrm{T}}\mathbf{1}_p$, and

$$\begin{aligned} E|\frac{\sqrt{\rho_n}}{n}\sum_{i=1}^{n}\sum_{j=1}^{n} g_{ij}m_jX_i^{\mathrm{T}}\mathbf{1}_p| &\le \frac{\sqrt{\rho_n}}{n}\sum_{i=1}^{n}\sum_{j=1}^{n} E|g_{ij}m_jX_i^{\mathrm{T}}\mathbf{1}_p| \\ &\le c\frac{\sqrt{\rho_n}}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}|g_{ij}| = O(1), \end{aligned}$$

using that $\max_i \sum_{j=1}^{n}|g_{ij}| = O(1/\sqrt{\rho_n})$, then by Markov inequality we have

$$\frac{\sqrt{\rho_n}}{n}(G\mathbf{m})^{\mathrm{T}}(X_1^{\mathrm{T}}\mathbf{1}_p, \cdots X_n^{\mathrm{T}}\mathbf{1}_p)^{\mathrm{T}} = O_P(1). \tag{B.7}$$

Therefore, $(G\mathrm{m})^{\mathrm{T}}P\mathbf{m} = o_P(\rho_n^{-1/2}nh^2)$.

(2) If $f(\cdot)$ and the second partial derivatives of $\beta(s)$ are all Lipschitz continuous, then it follows from the proof of Lemma 5(2) that

$$(G\mathrm{m})^{\mathrm{T}}P\mathbf{m} = (G\mathrm{m})^{\mathrm{T}}(X_1^{\mathrm{T}}\mathbf{1}_p, \cdots X_n^{\mathrm{T}}\mathbf{1}_p)^{\mathrm{T}}O_P(h^3 + h^2r_n).$$

Together with (B.7) we have

$$(G\mathbf{m})^{\mathrm{T}}P\mathbf{m} = O_P(\rho_n^{-1/2}nh^3 + \{nh^2\log n/\rho_n\}^{1/2}).$$

Proof of Lemma 11: In the following proofs we will always use the facts that the elements of G having the uniform order $O(1/\rho_n)$ and the row sums of the matrix G having the uniform order $O(1/\sqrt{\rho_n})$.

First we will show that $\frac{\rho_n}{n}\mathbf{m}^{\mathrm{T}}P\mathbf{m} = o_P(1)$. It can be seen from (B.4) that

$$\begin{aligned} \frac{\rho_n}{n}\mathbf{m}^{\mathrm{T}}P\mathbf{m} &= \frac{\rho_n}{n}(\mathbf{m} - S\mathbf{m})^{\mathrm{T}}(\mathbf{m} - S\mathbf{m}) \\ &= \frac{1}{n}(X_1^{\mathrm{T}}\mathbf{1}_p, \cdots, X_n^{\mathrm{T}}\mathbf{1}_p)(X_1^{\mathrm{T}}\mathbf{1}_p, \cdots, X_n^{\mathrm{T}}\mathbf{1}_p)^{\mathrm{T}}O_P(\rho_nh^4) = o_P(1) \end{aligned}$$

by law of large numbers.

Now we will show that $\frac{\rho_n}{n}L^{\mathrm{T}}PG\mathrm{m} = o_P(1)$ for $L = \mathbf{m}, \epsilon$ and $G\epsilon$.

It follows immediately from Lemma 10(1) and $\rho_nh^4 \to 0$ that $\frac{\rho_n}{n}\mathbf{m}^{\mathrm{T}}PG\mathbf{m} = o_P(\rho_n^{1/2}h^2) = o_P(1)$.

Next by the same lines as establishing Lemma 3 and Condition (4) that

$$\frac{\sqrt{\rho_n}}{n} H^{-1} \mathcal{X}^{\mathrm{T}} \mathcal{W} G \mathbf{m} = \begin{pmatrix} \Gamma\Gamma^{\mathrm{T}} \frac{\sqrt{\rho_n}}{n} \sum_{i=1}^{n} \sum_{j \neq i}^{n} g_{ij} \beta(s_j) K_h(\|s_i - s\|) \\ 0_{2p\times 1} \end{pmatrix} + o_P(\mathbf{1}_{3p}) \tag{B.8}$$

holds uniformly in $s \in \mathcal{S}$. Then using the same lines as establishing Lemma 6(2), the facts that the elements of G having the uniform order $O(1/\rho_n)$, the row sums of the matrix G having the uniform order $O(1/\sqrt{\rho_n})$ and $\rho_n/n \to 0$, we obtain that $\frac{\rho_n}{n}\epsilon^{\mathrm{T}} PG\mathbf{m} = o_P(1)$ and $\frac{\rho_n}{n}(G\epsilon)^{\mathrm{T}} PG\mathbf{m} = o_P(1)$.

Next it follows the same lines as establishing $n^{-1/2}(G\epsilon)^{\mathrm{T}} P\mathbf{m} = o_P(1)$ in Lemma 6(1) that $\sqrt{\rho_n/n}(G\epsilon)^{\mathrm{T}} P\mathbf{m} = o_P(1)$ when $\rho_n h^4 \to 0$.

As we have by Lemma 2(1) and (B.8) that

$$\sqrt{\rho_n} SG\mathbf{m} = \begin{pmatrix} \kappa_0^{-1} f^{-1}(s_1) X_1^{\mathrm{T}} \Psi^{-1} \Gamma\Gamma^{\mathrm{T}} \tilde{Z}(s_1) \\ \vdots \\ \kappa_0^{-1} f^{-1}(s_n) X_n^{\mathrm{T}} \Psi^{-1} \Gamma\Gamma^{\mathrm{T}} \tilde{Z}(s_n) \end{pmatrix} + o_P(1),$$

where $\tilde{Z}(s) = \lim_{n\to\infty} \frac{\sqrt{\rho_n}}{n} \sum_{i=1}^{n} \sum_{j\neq i}^{n} g_{ij}\beta(s_j) K_h(\|s_i - s\|)$, and

$$\frac{\rho_n}{n}(G\mathbf{m})^{\mathrm{T}} PG\mathbf{m} = \frac{1}{n}(\sqrt{\rho_n}G\mathbf{m} - \sqrt{\rho_n}SG\mathbf{m})^{\mathrm{T}}(\sqrt{\rho_n}G\mathbf{m} - \sqrt{\rho_n}SG\mathbf{m}),$$

the results (4) can be obtained similarly using the same lines as showing Lemma 7 with $G\mathbf{m}$ and $SG\mathbf{m}$ replaced by $\sqrt{\rho_n}G\mathbf{m}$ and $\sqrt{\rho_n}SG\mathbf{m}$ respectively.

The results (5) and (6) can be obtained from the proof of Lemma 8(2) and 8(3) under the assumptions of Lemma 11.

Finally, the result (7) can be obtained as Lemma 8(4).

Proof of Lemma 12: The proof can be established using the same lines as Lemma 9 under the assumptions of Lemma 12.

References

Anselin, L. (1988): *Spatial Econometrics: Methods and Models.* The Netherlands: Kluwer Academic Publishers.

Cheng, M., Zhang, W. and Chen, L. (2009). Statistical estimation in generalized multiparameter likelihood models. *Journal of the American Statistical Association*, **104**, 1179-1191.

Fan, J. and Gijbels, I. (1996). *Local Polynomial Modelling and Its Applications.* Chapman and Hall, London.

Fan, J. and Zhang, W. (1999). Statistical estimation in varying coefficient models. *Ann. Statist.*, **27**, 1491-1518.

Fan, J. and Zhang, W. (2000). Simultaneous confidence bands and hypothesis testing in varying-coefficient models. *Scandinavian Journal of Statistics*, **27**, 715-731.

Gao, J., Lu, Z. and Tjostheim, D. (2006). Estimation in semiparametric spatial regression. *The Annals of Statistics*, **34**, 1395-1435

Hansen Bruce E.(2008). Uniform convergence rates for kernel estimation with dependent data. *Econometric Theory*, **24**, 726-748.

Kelejian, H. H. and Prucha, I. R. (2001). On the aymptotic distribution of the Moran I test statistic with applications. *Journal of Econometrics* , **104**, 219-257.

Kelejian, H. H. and Prucha, I. R., 2010. Specification and estimation of spatial autoregressive models with autoregressive and heteroskedastic disturbances. *Journal of Econometrics*, **157**, 53-67.

Lee, L.-F. (2004). Asymptotic distribution of quasi-maximum likelihood estimators for spatial autoregressive models. *Econometrica*, **72** , 1899-1925.

Li, J. and Zhang, W. (2011). A semiparametric threshold model for censored longitudinal data analysis. *Journal of the American Statistical Association*, **106**, 685-696.

Linton O. (1995). Second order approximation in the partially linear regression model. *Econometrica*, **63**, 1079-1112.

Ord, J. K., 1975. Estimation methods for models of spatial interaction. *Journal of the American Statistical Association*, **70**, 120-126.

Su L. and Jin S. (2010). Profile quasi-maximum likelihood estimation of partially linear spatial autoregressive models. *Journal of Econometrics* , **157**, 18-33.

Sun, Y., Zhang, W. and Tong, H. (2007). Estimation of the covariance matrix of random effects in longitudinal studies. *The Annals of Statistics*. **35**, 2795-2814.

Tao, H. and Xia, Y. (2011) Adaptive semi-varying coefficient model selection, *Statistica Sinica*, **22**, 575-599.

Wang, H. and Xia, Y. (2009) Shrinkage estimation of the varying coefficient model. *Journal of the American Statistical Association*. **104**, 747-757.

White, H. (1994). *Estimation, inference and specification analysis*. Cambridge University Press.

Zhang, W., Fan, J. and Sun, Y. (2009). A semiparametric model for cluster data. *The Annals of Statistics*, **37**, 2377-2408.

Zhang, W., Lee, S. Y. and Song, X. (2002). Local polynomial fitting in semi-varying coefficient models, *Journal of Multivariate Analysis*, **82**, 166-188.