Submitted to the Annals of Statistics

SUPPLEMENT TO "A SEMIPARAMETRIC SPATIAL DYNAMIC MODEL"

By Yan Sun[‡], Hongjia Yan[§], Wenyang Zhang[§,†] and Zudi Lu[¶]

Shanghai University of Finance and Economics [‡], The University of York[§] and The University of Southampton[¶]

Appendix A. Proofs of the lemmas . In this section, we set $m_i = X_i^{\mathrm{T}}\boldsymbol{\beta}(s_i)$, 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 element of the matrix D, and c is a positive finite constant which may take different values at each appearance. Moreover, the operator $\mathrm{Vec}(\cdot)$ creates a column vector from the matrix by simply stacking its column vectors below one another.

Frequently we will use the facts (see Lee [10]) 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 [7]. First, we deal with the truncation error in replacing Y_i with the truncated process $Y_i 1(|Y_i| \leq \tau_n)$. 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 $R(s) = \frac{1}{n}\sum\limits_{i=1}^{n} K_h(\|s_i - s\|)Y_i 1(|Y_i| > \tau_n)$. Since $P(|Y_n| > \tau_n) \leq \tau_n^{-q} E|Y_n|^q$ and $\sum\limits_{n=1}^{\infty} \tau_n^{-q} = \sum\limits_{n=1}^{\infty} n^{-1}(\log n)^{-q/2} < \infty$ for $q > 2$. It follows that with probability one $|Y_n| \leq \tau_n$ for all sufficient large n. Since τ_n is increasing, we have for all sufficient large n, $|Y_i| \leq \tau_n$ for all $i \leq n$. This implies that $\sup_s |R(s)|$ is eventually zero with probability one.

*Supported by National Science Foundation of China (Grant 11271242), program for New Century Excellent Talents in China's University (NECT-10-0562), Key Laboratory of Mathematical Economics (SUFE), Ministry of Education of China, and the Singapore National Research Foundation under its Cooperative Basic Research Grant and administered by the Singapore Ministry of Health's National Medical Research Council (Grant No. NMRC/CBRG/0014/2012).

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

Next by a standard argument and Condition (4)

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

it follows that with probability one $\sup_s E|R(s)| = O(r_n)$.

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 $\mathcal{S} \subseteq \{s : \|s\| \leq c_1\}$. Next we create a grid using regions of the form $N_l = \{s : \|s - s_l\| \leq r_n h\}$. By selecting s_l to lay on a grid, the region $\{s : \|s\| \leq c_1\}$ can be covered with $N \leq c_1^2 h^{-2} r_n^{-2}$ 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 $s, s' \in R^2$, $|K(\|s\|) - K(\|s'\|)| \leq c_2\Big|\|s\| - \|s'\|\Big| \leq c_2\|s - s'\|$. Define $W^*(\|s\|) = c_2 I(\|s\| \leq 2L)$, thus for $s \in N_l$, we have $\|\frac{s-s_l}{h}\| \leq r_n$ and

$$|K(\|\frac{s_i - s}{h}\|) - K(\|\frac{s_i - s_l}{h}\|)| \leq r_n W^*(\|\frac{s_i - s_l}{h}\|). \tag{A.1}$$

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

$$\begin{aligned}
&\sup_{s\in N_l} |R_1(s) - ER_1(s)| \\
\leq\ & |R_1(s_l) - ER_1(s_l)| + r_n[|\tilde{R}_1(s_l)| + E|\tilde{R}_1(s_l)|] \\
\leq\ & |R_1(s_l) - ER_1(s_l)| + r_n|\tilde{R}_1(s_l) - E\tilde{R}_1(s_l)| + 2r_n E|\tilde{R}_1(s_l)| \\
\leq\ & |R_1(s_l) - ER_1(s_l)| + |\tilde{R}_1(s_l) - E\tilde{R}_1(s_l)| + 2c_3 r_n
\end{aligned}$$

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

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

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,

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

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

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

$$
n^{-1}H^{-1}\mathcal{X}^{\mathrm{T}}\mathcal{W}\mathcal{X}H^{-1}=
\begin{pmatrix}
\frac{1}{n}\sum_{i=1}^n X_iX_i^{\mathrm{T}}K_h(\|s_i-s\|) & \frac{1}{n}\sum_{i=1}^n X_iX_i^{\mathrm{T}}\otimes(\frac{s_i-s}{h})^{\mathrm{T}}K_h(\|s_i-s\|) \\
\frac{1}{n}\sum_{i=1}^n X_iX_i^{\mathrm{T}}\otimes\frac{s_i-s}{h}K_h(\|s_i-s\|) & \frac{1}{n}\sum_{i=1}^n X_iX_i^{\mathrm{T}}\otimes\frac{s_i-s}{h}(\frac{s_i-s}{h})^{\mathrm{T}}K_h(\|s_i-s\|)
\end{pmatrix}.
$$

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

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

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

(2) Note that

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

As

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

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

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

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

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

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

$$\begin{aligned} &\frac{1}{n}\sum_{i=1}^{n} X_i x_{il}(\frac{s_i - s}{h})^{\mathrm{T}} \ddot{\boldsymbol{\beta}}_l(s_i^*)(\frac{s_i - s}{h}) K_h(\|s_i - s\|) \\ = \quad & \kappa_2 f(s) E(X_1 x_{1l})(1,0,0,1)\mathrm{Vec}(\ddot{\boldsymbol{\beta}}_l(s)) + o_P(\mathbf{1}_p), \end{aligned}$$

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,

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

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

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

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

Proof of Lemma 3: Note that

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

In the following we will show that

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

where $V_{ij} = g_{ij} m_j X_i$ and $V_{ij} = g_{ij} m_j X_i \otimes \frac{s_i - s}{h}$, respectively.

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

$$\begin{aligned}
& \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n} g_{ij} m_j X_i K_h(\|s_i - s\|) \\
= \quad & \frac{1}{n}\sum_{i=1}^{n} g_{ii} m_i X_i K_h(\|s_i - s\|) + \frac{1}{n}\sum_{i=1}^{n}\sum_{j\neq i}^{n} g_{ij} E m_j X_i K_h(\|s_i - s\|) \\
& + \frac{1}{n}\sum_{i=1}^{n}\sum_{j\neq i}^{n} g_{ij}(m_j - E m_j) X_i K_h(\|s_i - s\|).
\end{aligned}$$

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

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

and

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

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

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

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

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

$$S = \kappa_0^{-1} n^{-1}(1 + o_P(1)) \cdot \Big(V_{ij}\Big)_{1\le i,j\le n} \tag{A.2}$$

where $V_{ij} = f^{-1}(s_i) X_i^{\mathrm{T}} \Psi^{-1} X_j K_h(\|s_j - s_i\|)$.

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

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

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

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

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

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

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 (A.2), Lemma 2(1), Lemma 1, Condition (4), continuity of $f(\cdot)$ and the second partial derivative 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

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

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 derivative of $\boldsymbol{\beta}(s)$ are Lipshitz continuous, then we can obtain by Lemma 1, Condition (4) and similar arguments as in Lemma 2(2) that

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

This together with Lemma 2(1) leads to

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

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

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

If $f(\cdot)$ and the second partial derivative of $\boldsymbol{\beta}(\cdot)$ are Lipschitz continuous, then

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

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

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

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 $L = \mathbf{m}$, $\boldsymbol{\epsilon}$ and $G\boldsymbol{\epsilon}$.

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

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

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 (A.3), (A.4) and Chebyshev inequality that

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

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

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

(2) Here, we will show that $n^{-1}L^{\mathrm{T}} PG\mathbf{m} = o_P(1)$ for $L = \mathbf{m}, \boldsymbol{\epsilon}$ 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 $E(\frac{1}{n}\boldsymbol{\epsilon}^{\mathrm{T}} G\mathbf{m}) = 0$, and

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

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

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

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

Therefore,

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

by law of large numbers.

Similarly,

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

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

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

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

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

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

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

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

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

and

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

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

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

$$\begin{aligned}
& \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k\neq j}^{n} g_{ij}g_{ik}m_j m_k - E\{\frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{n}\sum_{k\neq j}^{n} g_{ij}g_{ik}m_j m_k\} \\
= \; & \frac{1}{n}\sum_{j=1}^{n}\sum_{k\neq j}^{n}\sum_{i=1}^{n} g_{ij}g_{ik}\bar{m}_j\bar{m}_k + \frac{1}{n}\sum_{j=1}^{n}\sum_{k\neq j}^{n}\sum_{i=1}^{n} g_{ij}g_{ik}\bar{m}_j Em_k \\
& + \frac{1}{n}\sum_{k=1}^{n}\sum_{j\neq k}^{n}\sum_{i=1}^{n} g_{ij}g_{ik}\bar{m}_k Em_j.
\end{aligned}$$

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

and $J_{n3} = \frac{1}{n}\sum_{k=1}^{n}\sum_{j\neq k}^{n}\sum_{i=1}^{n} g_{ij}g_{ik}\bar{m}_k Em_j$, with

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

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

and

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

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

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

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

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

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

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

and

$$\mathrm{tr}(S) = \frac{pK(0)}{\kappa_0 nh^2}\sum_{i=1}^{n}f^{-1}(s_i) + o_p(1).$$

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

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

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 $n^{-1/2}\boldsymbol{\epsilon}^{\mathrm{T}}S\boldsymbol{\epsilon} = o_P(1)$.

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

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

(4) Note that

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

Moreover, $E[n^{-1/2}(G\mathbf{m}-SG\mathbf{m})^{\mathrm{T}}S\boldsymbol{\epsilon}] = 0$ and $\mathrm{var}[n^{-1/2}(G\mathbf{m}-SG\mathbf{m})^{\mathrm{T}}S\boldsymbol{\epsilon}] = \sigma_0^2 n^{-1}E[(G\mathbf{m}-SG\mathbf{m})^{\mathrm{T}}SS^{\mathrm{T}}(G\mathbf{m}-SG\mathbf{m})]$. It follows by (A.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 $\tilde{X} = (X_1^{\mathrm{T}}\mathbf{1}_p, \cdots, X_n^{\mathrm{T}}\mathbf{1}_p)^{\mathrm{T}}$. It can be seen by (A.5) 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 $n^{-1/2}(G\mathbf{m} - SG\mathbf{m})^{\mathrm{T}}S\boldsymbol{\epsilon} = o_P(1)$. Therefore, $n^{-1/2}\{(G\mathbf{m})^{\mathrm{T}}P\boldsymbol{\epsilon} - (G\mathbf{m} - SG\mathbf{m})^{\mathrm{T}}\boldsymbol{\epsilon}\} = o_P(1)$.

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 [8] and Lee [10].

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

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

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

Applying Holder inequality we obtain that

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

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

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

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 (A.6) that

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

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

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

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

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

Therefore,

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

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

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 $\sum\limits_{i=1}^{n}|b_{ik}|$ that

$$\begin{aligned}E(J_{n1}^2) &= \frac{1}{n^2}\sum_{k=1}^{n-1}(\sum_{i=k+1}^{n} b_{ik}^2)^2 D(\epsilon_1^2)\\ &\leq \frac{1}{n^2}D(\epsilon_1^2)\max_{i,k}|b_{ik}|^2\sum_{k=1}^{n-1}(\sum_{i=k+1}^{n}|b_{ik}|)^2=O(\frac{1}{n}).\end{aligned}$$

Since $J_{n2}=\frac{1}{n}\sum\limits_{k=1}^{n-1}\sum\limits_{l\neq k}^{n-1}(\sum\limits_{i=\max\{k,l\}+1}^{n} b_{ik}b_{il})\epsilon_k\epsilon_l$, 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 $J_{n3}=\frac{1}{n}\sum\limits_{k=1}^{n-1}[\sum\limits_{i=k+1}^{n}(\sigma_0^2 E[\sum\limits_{j=1}^{n} g_{ij}m_j-g_i^s]+\mu_3 b_{ii})b_{ik}]\epsilon_k$, it follows that

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

by $|E[\sum_{j=1}^{n} g_{ij}m_j - g_i^s]| \le \sum_{j=1}^{n}|g_{ij}Em_j| + E|g_i^s| = O(1)$ for any i, where $E|g_i^s| = O(1)$ is obtained using (A.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 $\frac{\sqrt{\rho_n}}{n}(G\mathbf{m})^{\mathrm{T}}(X_1^{\mathrm{T}}\mathbf{1}_p, \cdots X_n^{\mathrm{T}}\mathbf{1}_p)^{\mathrm{T}} = \frac{\sqrt{\rho_n}}{n}\sum_{i=1}^{n}\sum_{j=1}^{n} g_{ij}m_j X_i^{\mathrm{T}}\mathbf{1}_p$, and

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

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

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

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

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

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

Together with (A.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 (A.4) that

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

by law of large numbers.

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

It follows immediately from Lemma 10(1) and $\rho_n h^4 \to 0$ 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

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

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 $\frac{\rho_n}{n}\boldsymbol{\epsilon}^{\mathrm{T}}PG\mathbf{m} = o_P(1)$ and $\frac{\rho_n}{n}(G\boldsymbol{\epsilon})^{\mathrm{T}}PG\mathbf{m} = o_P(1)$.

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

As we have by Lemma 2(1) and (A.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 $\tilde{Z}(s) = \lim\limits_{n\to\infty}\frac{\sqrt{\rho_n}}{n}\sum\limits_{i=1}^{n}\sum\limits_{j\neq i}^{n} g_{ij}\boldsymbol{\beta}(s_j)K_h(\|s_i-s\|)$, 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.

Appendix B. Detail proofs of the theorems . In this section, we will give the more detailed proofs of the theorems.

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, $\mathrm{tr}\{(G+G^{\mathrm{T}})G\} - \frac{2}{n}\mathrm{tr}^2(G) = \frac{1}{2}\mathrm{tr}\{(\tilde{G}^{\mathrm{T}} + \tilde{G})(\tilde{G}^{\mathrm{T}} + \tilde{G})^{\mathrm{T}}\} \geq 0$ 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 [10] 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 ([18], 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

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

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

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

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

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

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

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

uniformly on Δ. 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

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

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: $Y=\alpha_0WY+\boldsymbol{\epsilon}$ with $\boldsymbol{\epsilon}\sim N(\mathbf{0},\sigma_0^2I_n)$. 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 $\max_{\sigma^2}E_a[l_a(\alpha,\sigma^2)]$ 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 $\alpha \in \Delta$, $\max_{\sigma^2} E_a[l_a(\alpha,\sigma^2)] \le E_a[l_a(\alpha_0,\sigma_0^2)]$, thus $Q_a(\alpha) \le Q_a(\alpha_0)$. 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 uniformly on Δ,

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

If we can show that

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

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 (B.3), it follows by the mean value theorem and Condition (5)-(6) that

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

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

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

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

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 (B.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 (B.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 $\alpha^* \neq \alpha_0$ 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

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

Since we have known that $Q_a(\alpha^*) - Q_a(\alpha_0) \leq 0$ and $\lim_{n\to\infty}[\sigma_a^2(\alpha^*) - \bar{\sigma}^2(\alpha^*)] \leq 0$, (B.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

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

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 4(1)-(3), 6, 7, 8(1)-(3), Chebyshev inequality and $\hat{\alpha} \xrightarrow{P} \alpha_0$ that

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

Proof of Theorem 2: Denoting $\boldsymbol{\theta} = (\alpha, \sigma^2)^{\mathrm{T}}$ 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

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

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

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

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 $\frac{1}{n}(G\mathbf{m} + G\boldsymbol{\epsilon})^{\mathrm{T}}P(G\mathbf{m} + G\boldsymbol{\epsilon}) = O_P(1)$ 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 (B.6) and we can get by Lemma 6 that

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

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

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

and

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

Next we have by straightforward calculation that

$$\begin{aligned}&\mathrm{var}((G\mathbf{m}-SG\mathbf{m})^{\mathrm{T}}\boldsymbol{\epsilon}+\{\boldsymbol{\epsilon}^{\mathrm{T}}G\boldsymbol{\epsilon}-\sigma_0^2\mathrm{tr}(G)\})\\ =\ &\sigma_0^2E[(G\mathbf{m}-SG\mathbf{m})^{\mathrm{T}}(G\mathbf{m}-SG\mathbf{m})]+(\mu_4-3\sigma_0^4)\sum_{i=1}^n g_{ii}^2\\ &+\sigma_0^4[\mathrm{tr}(GG^{\mathrm{T}})+\mathrm{tr}(G^2)]+2\mu_3E[(G\mathbf{m}-SG\mathbf{m})^{\mathrm{T}}G_c],\end{aligned}$$

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

$$\begin{aligned}&\mathrm{cov}\{(G\mathbf{m}-SG\mathbf{m})^{\mathrm{T}}\boldsymbol{\epsilon}+\{\boldsymbol{\epsilon}^{\mathrm{T}}G\boldsymbol{\epsilon}-\sigma_0^2\mathrm{tr}(G)\},\boldsymbol{\epsilon}^{\mathrm{T}}\boldsymbol{\epsilon}-n\sigma_0^2\}\\ =\ &\mu_3E[(G\mathbf{m}-SG\mathbf{m})^{\mathrm{T}}\mathbf{1}_n]+(\mu_4-\sigma_0^4)\mathrm{tr}(G).\end{aligned}$$

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

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

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

$$\begin{aligned}&\sqrt{nh_1^2f(s)}(\hat{\boldsymbol{\beta}}(s)-\boldsymbol{\beta}(s))\\ =\ &\sqrt{nh_1^2f(s)}(I_p,\mathbf{0}_{p\times 2p})(\mathcal{X}_1^{\mathrm{T}}\mathcal{W}_1\mathcal{X}_1)^{-1}\mathcal{X}_1^{\mathrm{T}}\mathcal{W}_1\boldsymbol{\epsilon}\\ &+\sqrt{nh_1^2f(s)}(\alpha_0-\hat{\alpha})(I_p,\mathbf{0}_{p\times 2p})(\mathcal{X}_1^{\mathrm{T}}\mathcal{W}_1\mathcal{X}_1)^{-1}\mathcal{X}_1^{\mathrm{T}}\mathcal{W}_1WY\\ &+\sqrt{nh_1^2f(s)}\{(I_p,\mathbf{0}_{p\times 2p})(\mathcal{X}_1^{\mathrm{T}}\mathcal{W}_1\mathcal{X}_1)^{-1}\mathcal{X}_1^{\mathrm{T}}\mathcal{W}_1\mathbf{m}-\boldsymbol{\beta}(s)\}\\ \equiv\ &J_{n1}+J_{n2}+J_{n3}\end{aligned}$$

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

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

and

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

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

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

Moreover, it follows immediately from Lemma 3 that

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

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

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

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,

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

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

$$J_{n3} = \sqrt{nh_1^2 f(s)}\Big[\frac{\kappa_2 h_1^2}{2\kappa_0}\{\boldsymbol{\beta}_{uu}(s)+\boldsymbol{\beta}_{vv}(s)\} + o_P(h_1^2\mathbf{1}_p)\Big].$$

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 [10], 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

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

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

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

uniformly on Δ.

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 $\lim_{n\to\infty}\frac{\rho_n}{n}[Q(\alpha) - Q(\alpha_0)] < 0$ when $\alpha \neq \alpha_0$ using a counter argument as in the proof of Theorem 1.

Write

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

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 equicontinuious 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

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

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

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

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 (B.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 $\frac{\rho_n}{n}(WY)^{\mathrm{T}}PWY = \frac{\rho_n}{n}(G\mathbf{m}+G\boldsymbol{\epsilon})^{\mathrm{T}}P(G\mathbf{m}+G\boldsymbol{\epsilon}) = O_P(1)$ 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

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

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

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

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 $\sqrt{\frac{\rho_n}{n}}(G\boldsymbol{\epsilon})^{\mathrm{T}}P\mathbf{m} = o_P(1)$ 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

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

and

$$\begin{aligned}\sigma_{qn}^2 \ \equiv\ & \mathrm{var}\Big\{(G\mathbf{m} - SG\mathbf{m})^{\mathrm{T}}\boldsymbol{\epsilon} + \boldsymbol{\epsilon}^{\mathrm{T}}[G - \frac{1}{n}\mathrm{tr}(G)I_n]\boldsymbol{\epsilon}\Big\}\\ =\ & \sigma_0^2E[(G\mathbf{m} - SG\mathbf{m})^{\mathrm{T}}(G\mathbf{m} - SG\mathbf{m})] + (\mu_4 - 3\sigma_0^4)\sum_{i=1}^n\{g_{ii} - \frac{\mathrm{tr}(G)}{n}\}^2\\ & +\sigma_0^4[\mathrm{tr}((G + G^{\mathrm{T}})G) - \frac{2}{n}\mathrm{tr}^2(G)]\\ & +2\mu_3E[(G\mathbf{m} - SG\mathbf{m})^{\mathrm{T}}(G_c - \frac{1}{n}\mathrm{tr}(G)\mathbf{1}_n)].\end{aligned}$$

As we have by Lemma 12 that

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

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

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

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 $J_{n2}=\sqrt{f(s)}\sqrt{\frac{nh_1^2}{\rho_n}}(\alpha_0-\hat{\alpha})(I_p,\mathbf{0}_{p\times 2p})(n^{-1}H_1^{-1}\mathcal{X}_1^{\mathrm{T}}\mathcal{W}_1\mathcal{X}_1H_1^{-1})^{-1}\frac{\sqrt{\rho_n}}{n}H_1^{-1}\mathcal{X}_1^{\mathrm{T}}\mathcal{W}_1G(\mathbf{m}+\boldsymbol{\epsilon})$. 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

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

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 $J_{n2}=o_P(1)$.

REFERENCES

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

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

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

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

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

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

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

[8] 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.

[9] 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.

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

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

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

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

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

[15] 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.

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

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

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

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

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

Yan Sun
School of Economics
Shanghai University of Finance
and Economics, P. R. China
E-mail: sunyan@mail.shufe.edu.cn

Hongjia Yan and Wenyang Zhang
Department of Mathematics
The University of York, UK
E-mail: wenyang.zhang@york.ac.uk

Zudi Lu
School of Mathematical Sciences
The University of Southampton, UK
E-mail: Z.Lu@soton.ac.uk