Appendix B

B.1 Proof of Theorem 4.1

Put $\Delta = \boldsymbol{\eta} - \boldsymbol{\eta}^0$ and $A_i = (a_{isr})$, and observe that, by (4.5), for $\boldsymbol{\eta}$ within any given but fixed radius of $\boldsymbol{\eta}^0$,

$$\begin{aligned} \lambda_s(X_i, \boldsymbol{\eta}) &= \lambda_s(X_i, \boldsymbol{\eta}^0) + \lambda_s'(X_i, \boldsymbol{\eta}^0)^{\mathrm{T}} \Delta + \Theta_{is}(\boldsymbol{\eta}) \, \|\boldsymbol{\eta} - \boldsymbol{\eta}^0\|^2 \,, \\ a_{isr}(X_i, \boldsymbol{\eta}) &= a_{isr}(X_i, \boldsymbol{\eta}^0) + a_{isr}'(X_i, \boldsymbol{\eta}^0)^{\mathrm{T}} \Delta + \Theta_{isr}(\boldsymbol{\eta}) \, \|\boldsymbol{\eta} - \boldsymbol{\eta}^0\|^2 \,, \end{aligned} \tag{B.1}$$

where $a_{isr}'(X_i, \boldsymbol{\eta}) = \partial a_{isr}(X_i, \boldsymbol{\eta}) / \partial \boldsymbol{\eta}$ and, here and below, $\Theta_{i\ldots}(\boldsymbol{\eta})$ denotes a generic random variable satisfying, with probability 1, $|\Theta_{i\ldots}(\boldsymbol{\eta})| \le C_1$ for $\|\boldsymbol{\eta} - \boldsymbol{\eta}^0\| \le C_2$ and $\|X_i\| \le C_2$, for any $C_2 > 0$, where C_1 depends on C_2 but not on i. Hence, for distinct integers $t_1, \ldots, t_\ell$ depending on $s_1, \ldots, s_\ell$,

$$\begin{aligned} Q_{ir}(\boldsymbol{\eta}) &= \sum_{\ell=1}^{r} \sum_{s_1,\ldots,s_\ell} \nu_{s_1,\ldots,s_\ell}(r) \left(\prod_{t=1}^{\ell} W_{is_t} \right) \\ &\qquad \times \Big\{ \lambda_{s_1}(X_i, \boldsymbol{\eta}^0) + \lambda_{s_1}'(X_i, \boldsymbol{\eta}^0)^{\mathrm{T}} \Delta + \Theta_{it_1}(\boldsymbol{\eta}) \, \|\boldsymbol{\eta} - \boldsymbol{\eta}^0\|^2 \Big\} \\ &\qquad \times \ldots \times \Big\{ \lambda_{s_\ell}(X_i, \boldsymbol{\eta}^0) + \lambda_{s_\ell}'(X_i, \boldsymbol{\eta}^0)^{\mathrm{T}} \Delta + \Theta_{it_\ell}(\boldsymbol{\eta}) \, \|\boldsymbol{\eta} - \boldsymbol{\eta}^0\|^2 \Big\} \\ &= \sum_{\ell=1}^{r} \sum_{s_1,\ldots,s_\ell} \nu_{s_1,\ldots,s_\ell}(r) \left(\prod_{t=1}^{\ell} W_{is_t} \right) \Bigg[\prod_{t=1}^{\ell} \lambda_{s_t}(X_i, \boldsymbol{\eta}^0) \\ &\qquad + \sum_{t=1}^{\ell} \Big\{ \prod_{u \,:\, u \ne t} \lambda_{s_u}(X_i, \boldsymbol{\eta}^0) \Big\} \lambda_{s_t}'(X_i, \boldsymbol{\eta}^0)^{\mathrm{T}} \Delta + \Theta_i(\boldsymbol{\eta}) \, \|\boldsymbol{\eta} - \boldsymbol{\eta}^0\|^2 \Bigg] \\ &= Q_{ir}(X_i, \boldsymbol{\eta}^0) + u_{ir}^{\mathrm{T}} \Delta + \Theta_i(\boldsymbol{\eta}) \, \|\boldsymbol{\eta} - \boldsymbol{\eta}^0\|^2 \,. \end{aligned} \tag{B.2}$$

Combining (B.2) and (B.1) we deduce that

$$\{A_i(X_i,\boldsymbol{\eta})\,R_i(\boldsymbol{\eta})\}_s = \sum_{r=1}^{q} a_{isr}(X_i,\boldsymbol{\eta})\,\{\bar{Y}_i^r - Q_{ir}(\boldsymbol{\eta})\}$$
$$= \{A_i(X_i,\boldsymbol{\eta}^0)\,R_i(\boldsymbol{\eta}^0)\}_s + \sum_{r=1}^{q}\Big[\{\bar{Y}_i^r - Q_{ir}(\boldsymbol{\eta}^0)\}\,a'_{isr}(X_i,\boldsymbol{\eta}^0) - a_{isr}(X_i,\boldsymbol{\eta}^0)\,u_{ir}\Big]^{\mathrm{T}}\Delta$$
$$+\,\Theta_{is}(\boldsymbol{\eta})\left(\left|\bar{Y}_i^q\right|+1\right)\|\boldsymbol{\eta}-\boldsymbol{\eta}^0\|^2\,.$$

(If $a_{isr}(x,\boldsymbol{\eta})$ does not actually depend on $\boldsymbol{\eta}$ then derivatives of a_{isr} with respect $\boldsymbol{\eta}$ vanish, and so the term in $|\bar{Y}_i^r|$ can be dropped. This is the reason it is not necessary, when $A_i(x,\boldsymbol{\eta})$ does not depend on $\boldsymbol{\eta}$, to have $E(|Y|^c\,|\,X=x)$ bounded in x for all values of c, although boundedness of $E(|Y|^p\,|\,X=x)$ is needed in order to work with qth cumulants.) Summing over i, and equating to zero as entailed by (2.12), we deduce that $\boldsymbol{\eta}=\hat{\boldsymbol{\eta}}$ satisfies

$$\frac{1}{k}\sum_{i=1}^{k}\sum_{r=1}^{q}\Big[a_{isr}(X_i,\boldsymbol{\eta}^0)\,u_{ir} - \{\bar{Y}_i^r - Q_{ir}(\boldsymbol{\eta}^0)\}\,a'_{isr}(X_i,\boldsymbol{\eta}^0)\Big]^{\mathrm{T}}(\boldsymbol{\eta}-\boldsymbol{\eta}^0)$$
$$+\Theta_s(\boldsymbol{\eta})\left(\left|\bar{Y}^r\right|+1\right)\|\boldsymbol{\eta}-\boldsymbol{\eta}^0\|^2 = \frac{1}{k}\sum_{i=1}^{k}\{A_i(X_i,\boldsymbol{\eta}^0)\,R_i(\boldsymbol{\eta}^0)\}_s\,, \tag{B.3}$$

or equivalently,

$$(M+L)\,(\boldsymbol{\eta}-\boldsymbol{\eta}^0) + \Theta(\boldsymbol{\eta})\,\|\boldsymbol{\eta}-\boldsymbol{\eta}^0\|^2 = S\,. \tag{B.4}$$

In (B.3), Θ_s is a random variable satisfying, with probability 1, $|\Theta_s(\boldsymbol{\eta})| \le p^{-1/2}C_1$ for $\|\boldsymbol{\eta}-\boldsymbol{\eta}^0\| \le C_2$ and C_2 sufficiently small, where C_1 depends on C_2, and in (B.4), $\Theta(\boldsymbol{\eta})$ is the p-vector with sth component $\Theta_s(\boldsymbol{\eta})$.

In view of (4.5), the summands in the definition of b_{st} at (4.3) are uniformly bounded, and all moments of the summands in (4.4) are finite. Moreover, ℓ_{rj} has zero mean and variance of order k^{-1}. Therefore Rosenthal's inequality implies that for all $C_3, C_4 > 1$,

$$P(\|L\| > C_3) \leq C_5(C_4) \left(C_3 \, k^{1/2}\right)^{-C_4}, \tag{B.5}$$

where $C_5(C_4) > 0$ depends on C_4 but not on C_3 or k. (In (B.5), and below, we write the norm $\|Q\|$ of a $p \times p$ matrix to denote the supremum of $\|Q\, v\|$ over all p-vectors v for which $\|v\| = 1$.)

Let $c \in (0, 1)$ denote a lower bound to the least eigenvalue of $M^{\mathrm{T}} M$, and put $C_3 = \frac{1}{3} c$. If $\|\boldsymbol{\eta} - \boldsymbol{\eta}^0\| \leq \min(C_1^{-1} C_3, C_2)$ and $\|L\| \leq C_3$ then

$$\left\|(M + L)\,(\boldsymbol{\eta} - \boldsymbol{\eta}^0)\right\| - \|\Theta(\boldsymbol{\eta})\| \, \|\boldsymbol{\eta} - \boldsymbol{\eta}^0\|^2 \geq (c - \|L\| - C_3)\, \|\boldsymbol{\eta} - \boldsymbol{\eta}^0\| \geq C_3 \, \|\boldsymbol{\eta} - \boldsymbol{\eta}^0\|,$$

and therefore equations (2.12) have a solution whenever

$$\|S\| \leq C_3 \, \min\left(C_1^{-1} C_3, C_2\right).$$

Hence, the probability of a solution is not less than $\pi_k \equiv 1 - C_6 \, k^{-C_4/2}$, for some $C_6 > 0$. Moreover, we can deduce from (B.4) that under the same conditions, any solution $\hat{\boldsymbol{\eta}}$ of (B.1) satisfies

$$\left\|(\hat{\boldsymbol{\eta}} - \boldsymbol{\eta}^0) - (M + L)^{-1} \, S\right\| \leq (3/2c)\, C_1 \, \|\hat{\boldsymbol{\eta}} - \boldsymbol{\eta}^0\|^2, \tag{B.6}$$

and therefore, if $\|\hat{\boldsymbol{\eta}} - \boldsymbol{\eta}^0\| \leq C_3$, we can see from (B.4) that

$$\tfrac{1}{2} \, \|\hat{\boldsymbol{\eta}} - \boldsymbol{\eta}^0\| \leq \|\hat{\boldsymbol{\eta}} - \boldsymbol{\eta}^0\| \left\{1 - (3/2c)\, \|\hat{\boldsymbol{\eta}} - \boldsymbol{\eta}^0\|\right\} \leq \left\|(M + L)^{-1} \, S\right\| \leq (3/2c)\, \|S\|. \tag{B.7}$$

Together, (B.6) and (B.7) imply that if $\|\hat{\boldsymbol{\eta}} - \boldsymbol{\eta}^0\| \leq \min(C_1^{-1} C_3, C_2, C_3)$ then with probability not less than π_k,

$$\Big\|(\hat{\boldsymbol{\eta}} - \boldsymbol{\eta}^0) - (M+L)^{-1}\, S\Big\| \le (3/2c)\, C_1\, C_3^{-2}\, \|S\|^2 \,,$$

from which follows the second-last inequality in the Theorem. The last inequality in the theorem follows via a Taylor expansion, enabled by (B.5).

B.2 Proof of Corollary 4.1

Define the event $\mathcal{E} = \{\|S\| \le \epsilon, \|L\| \le \epsilon\}$, and let $\widetilde{\mathcal{E}}$ denote the complement of $\mathcal{E}$. Let $\epsilon > 0$, let $C > 0$ be as in the condition $\max_{1\le i\le k} n_i = O(k^C)$ in the statement of the corollary, and note that

$$\begin{aligned}(\sup|f|)^{-1}\, E\{|f(\hat{\boldsymbol{\eta}})|\, I(\widetilde{\mathcal{E}})\} \le P(\widetilde{\mathcal{E}}) &\le P(\|S\| > \epsilon) + P(\|L\| > \epsilon) \\ &\le \epsilon^{-2}\, E\big(\|S\|^2\big) + \epsilon^{-2(C+1)}\, E\big(\|L\|^{2(C+1)}\big) \\ &= O\big(K^{-1} + k^{-(C+1)}\big) = O\big(K^{-1}\big)\,. \qquad \text{(B.8)}\end{aligned}$$

Assume that f has two bounded derivatives within radius C_7 of $\boldsymbol{\eta}^0$, where $C_7 > 0$, and let D_4 be as in Theorem 4.1. Let $c \in (0,1)$ be a lower bound to the least eigenvalue of $M^{\mathrm{T}} M$, and write f' for the p-vector of first derivatives of f. By choosing $\epsilon = \epsilon(C_7)$ sufficiently small we can show, as in the proof of Theorem 4.1, that when $\mathcal{E}$ obtains we have $\|\hat{\boldsymbol{\eta}} - \boldsymbol{\eta}^0\| \le \frac{1}{2}\, C_7$, $\|\{(M+L)^{-1}\, S\}^{\mathrm{T}}\, f'(\boldsymbol{\eta}^0)\| \le \frac{1}{2}\, C_7$ and $D_4\, \|S\|^2 \le \frac{1}{2}\, C_7$, and that the least eigenvalues of $(M+L)^{\mathrm{T}}(M+L)$ and $(I+LM^{-1})^{\mathrm{T}}(I+LM^{-1})$ exceed $\frac{1}{2}\, c$. Hence, by Taylor expansion of $f(\hat{\boldsymbol{\eta}})$ about $\boldsymbol{\eta}^0$, we deduce from Theorem 4.1 that there exists a constant $C_8 > 0$ such that, provided $\mathcal{E}$ obtains,

$$\Big|f(\hat{\boldsymbol{\eta}}) - f(\boldsymbol{\eta}^0) - \big\{(M+L)^{-1}\, S\big\}^{\mathrm{T}}\, f'(\boldsymbol{\eta}^0)\Big| \le C_8\, \|S\|^2 \,.$$

Therefore,

$$\begin{aligned}&\Big|E\{f(\hat{\boldsymbol{\eta}})\,I(\mathcal{E})\} - f(\boldsymbol{\eta}^0)\,P(\mathcal{E}) - E\Big[\big\{(M+L)^{-1}\,S\big\}^{\mathrm{T}}\,I(\mathcal{E})\Big]\,f'(\boldsymbol{\eta}^0)\Big| \\ &= O\big(E\|S\|^2\big) = O\big(K^{-1}\big)\,. \qquad \text{(B.9)}\end{aligned}$$

Let $j_0 \geq 1$ denote an integer. Noting the properties discussed in the previous paragraph, we deduce that

$$\begin{aligned}&E\{(M+L)^{-1}\,S\,I(\mathcal{E})\} = M^{-1}\,E\Big\{\big(I + L\,M^{-1}\big)^{-1}\,S\,I(\mathcal{E})\Big\} \\ &= M^{-1}\sum_{j=0}^{j_0} E\Big\{\big(-\,LM^{-1}\big)^j\,S\,I(\mathcal{E})\Big\} + O\big[E\{\|L\|^{j_0+1}\,\|S\|\,I(\mathcal{E})\}\big] \\ &= M^{-1}\sum_{j=0}^{j_0}\Big[E\{(-\,LM^{-1})^j\,S\} - E\{(-\,LM^{-1})^j\,S\,I(\|S\| > \epsilon)\}\Big] \\ &\quad + O\bigg\{\sum_{j=0}^{j_0}\Big[E\{\|L\|^{2j}\,I(\|L\| > \epsilon)\}\Big]^{1/2}\big(E\|S\|^2\big)^{1/2} + E\{\|L\|^{j_0+1}\,\|S\|\,I(\mathcal{E})\}\bigg\}, \qquad \text{(B.10)}\end{aligned}$$

where, here and below, order-of-magnitude expressions for vectors are interpreted component by component. We can further expand terms in L, writing $L = L_1 + L_2$ where L_1 is the $p \times p$ matrix with (s,t)th component equal to $b_{st} - E(b_{st})$, and $L_2 = L - L_1$ (see (4.4) for a definition of L). Condition (4.5) implies that $E\|L_1\|^j = O(k^{-j/2})$ and $E\|L_2\|^j + E\|S\|^j = O(K^{-j/2})$ for all integers j, and so we can deduce from (B.10) that the same expansion holds if we replace L by L_1, replace ϵ by $\frac{1}{2}\,\epsilon$ in one place, and add a remainder of $O(K^{-1} + k^{-(j_0+1)}\,K^{-1/2})$ to the right-hand side:

$$\begin{aligned}
&E\big\{(M+L_1)^{-1}\,S\,I(\mathcal{E})\big\} \\
=\;& M^{-1}\sum_{j=0}^{j_0}\Big[E\big\{(-\,L_1M^{-1})^j\,S\big\}-E\big\{(-\,L_1M^{-1})^j\,S\,I(\|S\|>\epsilon)\big\}\Big] \\
&\quad +O\bigg\{\sum_{j=0}^{j_0}\Big[E\big\{\|L_1\|^{2j}\,I\big(\|L_1\|>\tfrac{1}{2}\,\epsilon\big)\big\}\Big]^{1/2}\big(E\|S\|^2\big)^{1/2} \\
&\qquad +E\big\{\|L_1\|^{j_0+1}\,\|S\|\,I(\mathcal{E})\big\}\bigg\}+O\big(K^{-1}+k^{-(j_0+1)}\,K^{-1/2}\big)\,. \qquad \text{(B.11)}
\end{aligned}$$

Since $E(\|S\|^2)=O(K^{-1})$ then

$$\begin{aligned}
E\big\{\|L_1\|^{j_0+1}\,\|S\|\,I(\mathcal{E})\big\} &\le \Big[E\big(\|L_1\|^{2(j_0+1)}\big)\,E\big(\|S\|^2\big)\Big]^{1/2} \\
&= O\Big\{\big(E\|L_1\|^{2(j_0+1)}\big)^{1/2}\,K^{-1/2}\Big\}\,.
\end{aligned}$$

Using the argument leading to (B.8) we can show that if j_0 is sufficiently large,

$$E\big(\|L_1\|^{2(j_0+1)}\big)=O\big(k^{-(j_0+1)}\big)=O\big(K^{-1}\big)\,, \qquad \text{(B.12)}$$

and therefore

$$E\big\{\|L_1\|^{j_0+1}\,\|S\|\,I(\mathcal{E})\big\}=O\big(K^{-1}\big)\,. \qquad \text{(B.13)}$$

Moreover, $E(S\,|\,\mathcal{F}_X)=0$, where $\mathcal{F}_X$ denotes the sigma-field generated by $X_1,X_2,\ldots$, and so

$$E\big\{(-\,L_1M^{-1})^j\,S\big\}=E\big\{(-\,L_1M^{-1})^j\,E(S\,|\,\mathcal{F}_X)\big\}=0\,. \qquad \text{(B.14)}$$

Also, for $j \geq 1$,

$$E\big\{(-\,L_1M^{-1})^j\,S\,I(\|S\|>\epsilon)\big\} \;=\; E\Big[(-\,L_1M^{-1})^j\,E\big\{S\,I(\|S\|>\epsilon)\;\big|\;\mathcal{F}_X\big\}\Big]$$
$$=\; O\Big[E\big\{\|L_1\|^j\,E\big(\|S\|^2\;\big|\;\mathcal{F}_X\big)\big\}\Big]=O\big(K^{-1}\big)\,. \quad \text{(B.15)}$$

Furthermore, using (B.5) we deduce that

$$E\big\{\|L_1\|^{2j}\,I\big(\|L_1\|>\tfrac{1}{2}\,\epsilon\big)\big\} \;\leq\; \big(E\|L_1\|^{4j}\big)^{1/2}\,P\big(\|L_1\|>\tfrac{1}{2}\,\epsilon\big)^{1/2}$$
$$=\; O\big(k^{-2j}\,k^{-C_4/2}\big)\,, \quad \text{(B.16)}$$

where the last identity holds provided $C_4 \geq 2\,(C+1)$; see (B.8). Additionally, by (B.8), $P(\mathcal{E}) = 1 - P(\widetilde{\mathcal{E}}) = 1 - O(K^{-1})$. Combining this property, (B.9) and (B.11)–(B.15) we deduce that

$$\big|E\{f(\hat{\boldsymbol{\eta}})\,I(\mathcal{E})\}-f(\boldsymbol{\eta}^0)\big|=O\big(K^{-1}\big)\,. \quad \text{(B.17)}$$

The corollary follows from (B.8) and (B.17).

B.3 Proof of (4.8)

Assume that $\min_{1\le i\le k} n_i(k) \to \infty$ and that (4.5) holds. Then, writing U_{ir} for $\bar{Y}_i^r - Q_{ir}$, and noting that $E(S_r \,|\, \mathcal{X}) = 0$, we have:

$$
\begin{aligned}
\operatorname{cov}(S_r, S_s) &= E\{\operatorname{cov}(S_r, S_s \,|\, \mathcal{X})\} + \operatorname{cov}\{E(S_r \,|\, \mathcal{X})\, E(S_s \,|\, \mathcal{X})\} \\
&= E\{\operatorname{cov}(S_r, S_s \,|\, \mathcal{X})\} = \frac{1}{k^2} \sum_{i=1}^{k} E\{\operatorname{cov}(U_{ir}, U_{is} \,|\, X_i)\} \\
&= \frac{1}{k^2} \sum_{i=1}^{k} E\Big\{\operatorname{cov}\Big(\Big[\lambda_1(X_i, \boldsymbol{\eta}) + \{\bar{Y}_i - \lambda_1(X_i, \boldsymbol{\eta})\}\Big]^r, \\
&\qquad\qquad \Big[\lambda_1(X_i, \boldsymbol{\eta}) + \{\bar{Y}_i - \lambda_1(X_i, \boldsymbol{\eta})\}\Big]^s \;\Big|\; X_i\Big)\Big\} \\
&= \frac{1}{k^2} \sum_{i=1}^{k} E\Big(E\Big[\lambda_1(X_i, \boldsymbol{\eta})^{r+s} + (r+s)\, \lambda_1(X_i, \boldsymbol{\eta})^{r+s-1} \{\bar{Y}_i - \lambda_1(X_i, \boldsymbol{\eta})\} \\
&\quad + \tfrac{1}{2}\,(r+s)\,(r+s-1)\, \lambda_1(X_i, \boldsymbol{\eta})^{r+s-2} \{\bar{Y}_i - \lambda_1(X_i, \boldsymbol{\eta})\}^2 \;\Big|\; X_i\Big] \\
&\quad - \Big[E\Big\{\lambda_1(X_i, \boldsymbol{\eta})^r + r\, \lambda_1(X_i, \boldsymbol{\eta})^{r-1} \{\bar{Y}_i - \lambda_1(X_i, \boldsymbol{\eta})\} \\
&\qquad + \tfrac{1}{2}\, r\,(r-1)\, \lambda_1(X_i, \boldsymbol{\eta})^{r-2} \{\bar{Y}_i - \lambda_1(X_i, \boldsymbol{\eta})\}^2 \;\Big|\; X_i\Big\}\Big] \\
&\quad \times \Big[E\Big\{\lambda_1(X_i, \boldsymbol{\eta})^s + s\, \lambda_1(X_i, \boldsymbol{\eta})^{s-1} \{\bar{Y}_i - \lambda_1(X_i, \boldsymbol{\eta})\} \\
&\qquad + \tfrac{1}{2}\, s\,(s-1)\, \lambda_1(X_i, \boldsymbol{\eta})^{s-2} \{\bar{Y}_i - \lambda_1(X_i, \boldsymbol{\eta})\}^2 \;\Big|\; X_i\Big\}\Big]\Big) + o\big(K^{-1}\big) \\
&= \frac{1}{k^2} \sum_{i=1}^{k} E\Big(\lambda_1(X_i, \boldsymbol{\eta})^{r+s} + \tfrac{1}{2}\,(r+s)\,(r+s-1)\, \lambda_1(X_i, \boldsymbol{\eta})^{r+s-2} \\
&\qquad\qquad \times E\big[\{\bar{Y}_i - \lambda_1(X_i, \boldsymbol{\eta})\}^2 \,\big|\, X_i\big] \\
&\quad - \Big(\lambda_1(X_i, \boldsymbol{\eta})^r + \tfrac{1}{2}\, r\,(r-1)\, \lambda_1(X_i, \boldsymbol{\eta})^{r-2} \\
&\qquad\qquad \times E\big[\{\bar{Y}_i - \lambda_1(X_i, \boldsymbol{\eta})\}^2 \,\big|\, X_i\big]\Big)^2\Big) \\
&\quad \times \Big(\lambda_1(X_i, \boldsymbol{\eta})^s + \tfrac{1}{2}\, s\,(s-1)\, \lambda_1(X_i, \boldsymbol{\eta})^{s-2} \\
&\qquad\qquad \times E\big[\{\bar{Y}_i - \lambda_1(X_i, \boldsymbol{\eta})\}^2 \,\big|\, X_i\big]\Big)^2\Big) + o\big(K^{-1}\big) \\
&= \frac{1}{2\,k^2} \big\{(r+s)\,(r+s-1) - r\,(r-1) - s\,(s-1)\big\} \\
&\quad \times \sum_{i=1}^{k} n_i^{-1}\, E\Big\{\lambda_1(X_i, \boldsymbol{\eta})^{r+s-2} \operatorname{var}(Y_i \,|\, X_i)\Big\} + o\big(K^{-1}\big) \\
&= rs\, K^{-1}\, E\big\{\lambda_1(X, \boldsymbol{\eta})^{r+s-2} \operatorname{var}(Y \,|\, X)\big\} + o\big(K^{-1}\big)\,,
\end{aligned}
$$

which establishes (4.8).

B.4 Proof of Theorem 4.2

The version, for the case of centred moments, of the argument leading to (B.3) is identical in essential details to that given before, and shows that $\hat{\boldsymbol{\eta}}$ is the solution in $\boldsymbol{\eta}$ of the equation:

$$\frac{1}{k} \sum_{i=1}^{k} \sum_{r=1}^{q} \Big\{ a'_{isr}(X_i, \boldsymbol{\eta}^0)\, \check{R}_{ir}(\boldsymbol{\eta}^0) + a_{isr}(X_i, \boldsymbol{\eta}^0)\, \check{R}'_{ir}(\boldsymbol{\eta}^0) \Big\}^{\mathrm{T}} (\boldsymbol{\eta} - \boldsymbol{\eta}^0)$$
$$+ \Theta_s(\boldsymbol{\eta}) \left(\left| \bar{Y}^r \right| + 1 \right) \|\boldsymbol{\eta} - \boldsymbol{\eta}^0\|^2 = -\frac{1}{k} \sum_{i=1}^{k} \{ A_i(X_i, \boldsymbol{\eta}^0)\, \check{R}_i(\boldsymbol{\eta}^0) \}_s \,,$$

where $\check{R}'_{ir} = (\partial/\partial\boldsymbol{\eta})\, \check{R}_{ir}$ is a p-vector. Equivalently, in place of (B.4),

$$(\check{M} + \check{L})\, (\boldsymbol{\eta} - \boldsymbol{\eta}^0) + \Theta(\boldsymbol{\eta})\, \|\boldsymbol{\eta} - \boldsymbol{\eta}^0\|^2 = \check{S} \,.$$

The remainder of the proof is virtually equivalent to that of Theorem 4.1.

B.5 Proof of nonsingularity of $A(X_i, \boldsymbol{\eta})$ when defined by (2.15) and (4.6) holds

For brevity we treat the case where $p = q = 2$ and $n_i \equiv n$, although other cases are similar. Writing $\mu_{ir} = E[\{\bar{Y}_i - \lambda_1(X_i, \boldsymbol{\eta})\}^r \mid X_i]$, the first of the following results can be deduced from (2.16):

$$\boldsymbol{\Sigma}_i = \begin{pmatrix} \mu_{i2} & \mu_{i3} \\ \mu_{i3} & \mu_{i4} - \mu_{i2}^2 \end{pmatrix}, \boldsymbol{\Sigma}_i^{-1} = |\boldsymbol{\Sigma}_i|^{-1} \begin{pmatrix} \mu_{i4} - \mu_{i2}^2 & -\mu_{i3} \\ -\mu_{i3} & \mu_{i2} \end{pmatrix},$$

$$\lambda_r'(X_i, \boldsymbol{\eta}) = \begin{pmatrix} \lambda_{r1}'(X_i, \boldsymbol{\eta}) \\ \lambda_{r2}'(X_i, \boldsymbol{\eta}) \end{pmatrix}, \check{\boldsymbol{D}}_i = \begin{pmatrix} \lambda_{11}'(X_i, \boldsymbol{\eta}) & n^{-1}\, \lambda_{21}'(X_i, \boldsymbol{\eta}) \\ \lambda_{12}'(X_i, \boldsymbol{\eta}) & n^{-1}\, \lambda_{22}'(X_i, \boldsymbol{\eta}) \end{pmatrix},$$

where $\lambda_r'(X_i, \boldsymbol{\eta})$ is the 2-vector of derivatives of $\lambda_r(X_i, \boldsymbol{\eta})$ with respect to $\boldsymbol{\eta}$, $\lambda_{rj}' = \partial \lambda_r / \partial \boldsymbol{\eta}_j$ and we have used the fact that $W_{ir} = n^{1-r}$. Similarly, $\mu_{i2} = n^{-1}\, \lambda_2(X_i, \boldsymbol{\eta})$, $\mu_{i3} = n^{-2}\, \lambda_3(X_i, \boldsymbol{\eta})$,

$$\begin{aligned} \mu_{i4} - \mu_{i2}^2 &= n^{-3}\, \lambda_4(X_i, \boldsymbol{\eta}) + 3\, \{n^{-1}\, \lambda_2(X_i, \boldsymbol{\eta})\}^2 - \{n^{-1}\, \lambda_2(X_i, \boldsymbol{\eta})\}^2 \\ &= n^{-3}\, \lambda_4(X_i, \boldsymbol{\eta}) + 2\, \{n^{-1}\, \lambda_2(X_i, \boldsymbol{\eta})\}^2 , \end{aligned}$$

$$|\boldsymbol{\Sigma}_i| = \mu_{i2}\, \mu_{i4} - \mu_{i2}^3 - \mu_{i3}^2 = 2\, \mu_{i2}^3 + O_p\big(n^{-4}\big) = 2\, n^{-3}\, \lambda_2(X_i, \boldsymbol{\eta})^3 + O_p\big(n^{-4}\big) .$$

Therefore, provided that $\lambda_2(x, \boldsymbol{\eta})$ is bounded away from zero uniformly in x in the support of the distribution of X and in a neighbourhood of $\boldsymbol{\eta}^0$, we have:

$$\boldsymbol{\Sigma}_i^{-1} = \frac{1 + O_p(n^{-1})}{2\, n^{-3}\, \lambda_2^3} \begin{pmatrix} n^{-3}\, \lambda_4 + 2\, n^{-2}\, \lambda_2^2 & -n^{-2}\, \lambda_3 \\ -n^{-2}\, \lambda_3 & n^{-1}\, \lambda_2 \end{pmatrix} .$$

Here and below we write simply λ and λ_{rj}' for $\lambda(X_i, \boldsymbol{\eta})$ and $\lambda_{rj}'(X_i, \boldsymbol{\eta})$, respectively. Hence, defining $\lambda_r(X_i, \boldsymbol{\eta}) = (\lambda_{r1}(X_i, \boldsymbol{\eta}), \lambda_{r2}(X_i, \boldsymbol{\eta}))^{\mathrm{T}}$, a 2-vector, we deduce that

$$n^{-1}\, \check{\boldsymbol{D}}_i\, \boldsymbol{\Sigma}_i^{-1} = \frac{1 + O_p(n^{-1})}{2\, \lambda_2^3} \begin{pmatrix} 2\, \lambda_2^2\, \lambda_{11}' + n^{-1}\, (\lambda_4\, \lambda_{11}' - \lambda_3\, \lambda_{21}') & \lambda_2\, \lambda_{21}' - \lambda_3\, \lambda_{11}' \\ 2\, \lambda_2^2\, \lambda_{12}' + n^{-1}\, (\lambda_4\, \lambda_{12}' - \lambda_3\, \lambda_{22}') & \lambda_2\, \lambda_{22}' - \lambda_3\, \lambda_{12}' \end{pmatrix} = \begin{pmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{pmatrix},$$

say, where λ_r and λ_{rj}' denote $\lambda_r(X_i, \boldsymbol{\eta})$ and $\lambda_{rj}'(X_i, \boldsymbol{\eta})$, respectively, and we have suppressed the dependence of these quantities and the coefficients c_{st} on i.

Since $\check{R}_{ir} = \{\bar{Y}_i - \lambda_1(X_i, \boldsymbol{\eta})\}^r - \check{Q}_{ir}$ then

$$
\begin{aligned}
-\check{R}_{ir}' &= -\partial \frac{\partial}{\boldsymbol{\eta}}\, \check{R}_{ir} = r\,\{\bar{Y}_i - \lambda_1(X_i, \boldsymbol{\eta})\}^{r-1}\, \lambda_1' + \partial \frac{\partial}{\boldsymbol{\eta}}\, \check{Q}_{ir} \\
&= r\,\{\bar{Y}_i - \lambda_1(X_i, \boldsymbol{\eta})\}^{r-1} \begin{pmatrix} \lambda_{11}' \\ \lambda_{12}' \end{pmatrix} + \begin{cases} 0 & \text{if } r = 1 \\ n^{-1} \begin{pmatrix} \lambda_{21} \\ \lambda_{22} \end{pmatrix} & \text{if } r = 2, \end{cases}
\end{aligned}
$$

because $\check{Q}_{i1} \equiv 0$. Therefore, writing $\Delta = (\Delta_1, \Delta_2)^{\mathrm{T}} = \boldsymbol{\eta} - \boldsymbol{\eta}^0$, we have:

$$
\begin{aligned}
\check{R}_{ir}(\boldsymbol{\eta}) &= \check{R}_{ir}(\boldsymbol{\eta}^0) + \check{R}_{ir}'(\boldsymbol{\eta}^0)^{\mathrm{T}}\,(\boldsymbol{\eta} - \boldsymbol{\eta}^0) + O_p\big(\|\Delta\|^2\big) \\
&= \check{R}_{ir}(\boldsymbol{\eta}^0) - r\,\{\bar{Y}_i - \lambda_1(X_i, \boldsymbol{\eta})\}^{r-1}\,(\lambda_{11}'\,\Delta_1 + \lambda_{12}'\,\Delta_2) \\
&\quad + O_p(\|\Delta\|^2) - \begin{cases} 0 & \text{if } r = 1 \\ n^{-1} \quad (\lambda_{21}'\Delta_1 + \lambda_{22}'\Delta_2) & \text{if} r = 2. \end{cases}
\end{aligned}
$$

Hence,

$$
\begin{aligned}
& n^{-1} \sum_{i=1}^{k} \check{\boldsymbol{D}}_i\, \boldsymbol{\Sigma}_i^{-1}\, \check{R}_i \\
& = \sum_{i=1}^{k} \begin{pmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{pmatrix} \left\{ \check{R}_i(\boldsymbol{\eta}^0) - \begin{pmatrix} \lambda_{11}'\,\Delta_1 + \lambda_{12}'\,\Delta_2 \\ 2\,\{\bar{Y}_i - \lambda_1(X_i, \boldsymbol{\eta})\}\,(\lambda_{11}'\,\Delta_1 + \lambda_{12}'\,\Delta_2) + n^{-1}\,(\lambda_{21}'\,\Delta_1 + \lambda_{22}'\,\Delta_2) \end{pmatrix} \right\} \\
& + O_p\left(k\,\|\Delta\|^2\right) \\
& = \sum_{i=1}^{k} \left\{ \begin{pmatrix} c_{11} & c_{12} \\ c_{21} & c_{22} \end{pmatrix} \check{R}_i(\boldsymbol{\eta}^0) - \begin{pmatrix} c_{11}\,(\lambda_{11}'\,\Delta_1 + \lambda_{12}'\,\Delta_2) + c_{12}\,\delta \\ c_{21}\,(\lambda_{11}'\,\Delta_1 + \lambda_{12}'\,\Delta_2) + c_{22}\,\delta \end{pmatrix} \right\} + (\text{negligible}),
\end{aligned}
$$

where

$$\delta = 2\,\{\bar{Y}_i - \lambda_1(X_i, \boldsymbol{\eta})\}\,(\lambda_{11}'\,\Delta_1 + \lambda_{12}'\,\Delta_2) + n^{-1}\,(\lambda_{21}'\,\Delta_1 + \lambda_{22}'\,\Delta_2)\,.$$

Now,

$$\frac{1}{k}\sum_{i=1}^{k}(c_{j1}\,\check{R}_{i1} + c_{j2}\,\check{R}_{i2}) \\ = \frac{1}{k}\sum_{i=1}^{k}\Big(c_{j1}(i)\,\{\bar{Y}_i - \lambda_1(X_i, \boldsymbol{\eta})\} + c_{j2}(i)\,\big[\{\bar{Y}_i - \lambda_1(X_i, \boldsymbol{\eta})\}^2 - n^{-1}\,\lambda_2(X_i, \boldsymbol{\eta})\big]\Big)\,.$$

The two terms here give contributions of sizes $(kn)^{-1/2}$ and $k^{-1/2}n^{-1}$, respectively, so the second is negligible. Now,

$$a_1 = a_1(i) \equiv c_{11}\,\lambda_{11}' = \lambda_2^{-1}\,{\lambda_{11}'}^2 + O\big(n^{-1}\big),$$

$$a_2 = a_2(i) \equiv c_{11}\,\lambda_{12}' = \lambda_2^{-1}\,\lambda_{11}'\,\lambda_{12}' + O\big(n^{-1}\big),$$

$$c_{21}\,\lambda_{11}' = \lambda_2^{-1}\,\lambda_{11}'\,\lambda_{12}' + O\big(n^{-1}\big)\,, \quad a_3 = a_3(i) \equiv c_{21}\,\lambda_{12}' = \lambda_2^{-1}\,{\lambda_{12}'}^2 + O\big(n^{-1}\big),$$

and so

$$a_2 = a_2(i) = c_{11}\,\lambda_{12}' = c_{21}\,\lambda_{11}' + O\big(n^{-1}\big)\,.$$

Therefore,

$$n^{-1}\sum_{i=1}^{k}\check{\boldsymbol{D}}_i\,\boldsymbol{\Sigma}_i^{-1}\,\check{R}_i = \sum_{i=1}^{k}\left\{\begin{pmatrix} c_{11}(i)\,\{\bar{Y}_i - \lambda_1(X_i, \boldsymbol{\eta})\} \\ c_{21}(i)\,\{\bar{Y}_i - \lambda_1(X_i, \boldsymbol{\eta})\} \end{pmatrix} - \begin{pmatrix} a_1 & a_2 \\ a_2 & a_3 \end{pmatrix}\begin{pmatrix} \Delta_1 \\ \Delta_2 \end{pmatrix}\right\} + (\text{negligible}).$$

The condition on $M^{0\mathrm{T}} M^0$ being nonsingular reduces to asymptotic nonsingularity of the matrix

$$\frac{1}{k} \sum_{i=1}^{k} \begin{pmatrix} a_1 & a_2 \\ a_2 & a_3 \end{pmatrix}$$

Therefore: the first of the following holds:

$$\sum_{i=1}^{k} (c_{11}\, \check{R}_{i1} + n\, c_{12}\, \check{R}_{i2}) = \sum_{i=1}^{k} \Big\{ c_{11}\, (\lambda_{11}\, \Delta_1 + \lambda_{12}\, \Delta_2) + c_{12}\, (\lambda_{21}\, \Delta_1 + \lambda_{22}\, \Delta_2) \Big\} + (\text{negligible}) \,,$$
$$\sum_{i=1}^{k} (c_{21}\, \check{R}_{i1} + n\, c_{22}\, \check{R}_{i2}) = \sum_{i=1}^{k} \Big\{ c_{21}\, (\lambda_{11}\, \Delta_1 + \lambda_{12}\, \Delta_2) + c_{22}\, (\lambda_{21}\, \Delta_1 + \lambda_{22}\, \Delta_2) \Big\} + (\text{negligible}) \,,$$

where

$$\check{R}_{i1} = \bar{Y}_i - \lambda_1(X_i, \boldsymbol{\eta}) \,, \quad \check{R}_{i2} = \{\bar{Y}_i - \lambda_1(X_i, \boldsymbol{\eta})\}^2 - E\big[\{\bar{Y}_i - \lambda_1(X_i, \boldsymbol{\eta})\}^2 \,|\, X_i\big] \,.$$

We can solve this for Δ_1 and Δ_2, obtaining the result that Δ_1 and Δ_2 both equal linear combinations in $k^{-1} \sum_i \check{R}_{i1}$ and $n\, k^{-1} \sum_i \check{R}_{i2}$. The second of these is only $O(k^{-1})$.