Appendix B

B.1 Proof of Theorem 4.1

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

$$
\begin{align*}
\lambda_s(X_i, \eta) &= \lambda_s(X_i, \eta^0) + \lambda_s'(X_i, \eta^0)^T \Delta + \Theta_{is}(\eta) \|\eta - \eta^0\|^2, \\
a_{isr}(X_i, \eta) &= a_{isr}(X_i, \eta^0) + a_{isr}'(X_i, \eta^0)^T \Delta + \Theta_{isr}(\eta) \|\eta - \eta^0\|^2,
\end{align*}
$$

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

$$
Q_{ir}(\eta) = \sum_{r=1}^r \sum_{s_1, \ldots, s_\ell} \nu_{s_1, \ldots, s_\ell}(r) \left( \prod_{t=1}^\ell W_{isr} \right) \\
\times \left\{ \lambda_{s_1}(X_i, \eta^0) + \lambda_{s_1}'(X_i, \eta^0)^T \Delta + \Theta_{it_1}(\eta) \|\eta - \eta^0\|^2 \right\} \\
\times \ldots \times \left\{ \lambda_{s_\ell}(X_i, \eta^0) + \lambda_{s_\ell}'(X_i, \eta^0)^T \Delta + \Theta_{it_\ell}(\eta) \|\eta - \eta^0\|^2 \right\}
$$

$$
= \sum_{r=1}^r \sum_{s_1, \ldots, s_\ell} \nu_{s_1, \ldots, s_\ell}(r) \left( \prod_{t=1}^\ell W_{isr} \right) \left[ \prod_{t=1}^\ell \lambda_{s_t}(X_i, \eta^0) \right.
\left. + \sum_{t=1}^\ell \left( \prod_{u: u \neq t} \lambda_{s_u}(X_i, \eta^0) \right) \lambda_{s_t}'(X_i, \eta^0)^T \Delta + \Theta_{i}(\eta) \|\eta - \eta^0\|^2 \right]
$$

$$
= Q_{ir}(X_i, \eta^0) + u_{ir}^T \Delta + \Theta_{i}(\eta) \|\eta - \eta^0\|^2. \quad (B.2)
$$
Combining (B.2) and (B.1) we deduce that

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

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

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

or equivalently,

\[
(M + L) (\eta - \eta^0) + \Theta(\eta) \| \eta - \eta^0 \|^2 = S. \quad (B.4)
\]

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

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, \(\ell_{rj}\) has zero mean and variance
of order $k^{-1}$. Therefore Rosenthal’s inequality implies that for all $C_3, C_4 > 1$,

$$P(\|L\| \geq C_3) \leq C_5(C_4) (C_3^{1/2})^{-C_4}, \quad (B.5)$$

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 $\|Qv\|$ over all $p$-vectors $v$ for which $\|v\| = 1$.)

Let $c \in (0, 1)$ denote a lower bound to the least eigenvalue of $M^TM$, and put $C_3 = \frac{1}{2} c$. If $\|\eta - \eta^0\| \leq \min(C_1^{-1} C_3, C_2)$ and $\|L\| \leq C_3$ then

$$\left\|(M + L)(\eta - \eta^0)\right\| \geq (c - \|L\| - C_3) \|\eta - \eta^0\| \geq C_3 \|\eta - \eta^0\|,$$

and therefore equations (2.12) have a solution whenever

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

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 $\hat{\eta}$ of (B.1) satisfies

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

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

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

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

\[
\left\| (\hat{\eta} - \eta^0) - (M + L)^{-1} S \right\| \leq (3/2c) C_1 C_3^{-2} \| S \|^2,
\]

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

\[
(\sup |f|)^{-1} E\{ |f(\hat{\eta})| I(\tilde{\mathcal{E}}) \} \leq P(\tilde{\mathcal{E}}) \leq P(\| S \| > \epsilon) + P(\| L \| > \epsilon) \\
\leq \epsilon^{-2} E\left( \| S \|^2 \right) + \epsilon^{-2(C+1)} E\left( \| L \|^{2(C+1)} \right) \\
= O(K^{-1} + k^{-(C+1)}) = O(K^{-1}). \tag{B.8}
\]

Assume that \( f \) has two bounded derivatives within radius \( C_7 \) of \( \eta^0 \), 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 \( M^T M \), 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 \( \| \hat{\eta} - \eta^0 \| \leq \frac{1}{2} C_7 \), \( \| (M + L)^{-1} S \| \leq \frac{1}{2} C_7 \) and \( D_4 \| S \|^2 \leq \frac{1}{2} C_7 \), and that the least eigenvalues of \( (M + L)^T (M + L) \) and \( (I + LM^{-1})^T (I + LM^{-1}) \) exceed \( \frac{1}{2} c \). Hence, by Taylor expansion of \( f(\hat{\eta}) \) about \( \eta^0 \), we deduce from Theorem 4.1 that there exists a constant \( C_8 > 0 \) such that, provided \( \mathcal{E} \) obtains,

\[
| f(\hat{\eta}) - f(\eta^0) - \{ (M + L)^{-1} S \}^T f'(\eta^0) | \leq C_8 \| S \|^2.
\]
Therefore,

\[
\left| E\{f(\hat{\eta})I(\mathcal{E})\} - f(\eta^0)P(\mathcal{E}) - E\left[\{(M + L)^{-1}S\}^T I(\mathcal{E})\right] f'(\eta^0) \right| = O(E\|S\|^2) = O(K^{-1}).
\] (B.9)

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

\[
E\{(M + L)^{-1}SI(\mathcal{E})\} = M^{-1} E\{(I + LM^{-1})^{-1}SI(\mathcal{E})\}
\]

\[
= M^{-1} \sum_{j=0}^{j_0} E\{(-LM)^jSI(\mathcal{E})\} + O\left[ E\{\|L\|^{j_0+1} \|S\|I(\mathcal{E})\} \right]
\]

\[
= M^{-1} \sum_{j=0}^{j_0} \left[ E\{(-LM)^jS\} - E\{(-LM)^jS|\|S\| > \epsilon\} \right]
\]

\[
+ O\left\{ \sum_{j=0}^{j_0} \left[ E\{\|L\|^{2j} I(\|L\| > \epsilon)\}\right]^{1/2} \left( E\|S\|^2\right)^{1/2} + E\{\|L\|^{j_0+1} \|S\|I(\mathcal{E})\} \right\},
\] (B.10)

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 \( \epsilon \) 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:

\[
E\{(M + L_1)^{-1} S I(\mathcal{E})\}
= M^{-1} \sum_{j=0}^{j_0} \left[ E\{(- L_1 M^{-1})^j S\} - E\{(- L_1 M^{-1})^j S I(\|S\| > \epsilon)\} \right] \\
+ O\left\{ \sum_{j=0}^{j_0} \left[ E\{\|L_1\|^2 I(\|L_1\| > \frac{1}{2} \epsilon)\} \right]^{1/2} (E\|S\|^2)^{1/2} \\
+ E\{\|L_1\|^{j_0+1} \|S\| I(\mathcal{E})\} \right\} + O(K^{-1} + k^{-(j_0+1)} K^{-1/2}). \tag{B.11}
\]

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

\[
E\{\|L_1\|^{j_0+1} \|S\| I(\mathcal{E})\} \leq \left[ E(\|L_1\|^{2(j_0+1)}) E(\|S\|^2) \right]^{1/2} \\
= O\left\{ (E\|L_1\|^{2(j_0+1)})^{1/2} K^{-1/2} \right\}.
\]

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

\[
E(\|L_1\|^{2(j_0+1)}) = O(k^{-(j_0+1)}) = O(K^{-1}), \tag{B.12}
\]

and therefore

\[
E\{\|L_1\|^{j_0+1} \|S\| I(\mathcal{E})\} = O(K^{-1}). \tag{B.13}
\]

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

\[
E\{- L_1 M^{-1})^j S\} = E\{(- L_1 M^{-1})^j E(S \mid \mathcal{F}_X)\} = 0. \tag{B.14}
\]
Also, for $j \geq 1$,
\[
E\{(-L_1M^{-1})^j S I(\|S\| > \epsilon)\} = E\left[(-L_1M^{-1})^j E\left\{ S I(\|S\| > \epsilon) \mid \mathcal{F}_X \right\} \right] = O\left[ E\{\|L_1\|^j E(\|S\|^2 \mid \mathcal{F}_X)\} \right] = O(K^{-1}). \tag{B.15}
\]

Furthermore, using (B.5) we deduce that
\[
E\{\|L_1\|^{2j} I(\|L_1\| > \frac{1}{2} \epsilon)\} \leq (E\|L_1\|^{4j})^{1/2} P(\|L_1\| > \frac{1}{2} \epsilon)^{1/2} = O(k^{-2j} k^{-C_4/2}), \tag{B.16}
\]
where the last identity holds provided $C_4 \geq 2(C + 1)$; see (B.8). Additionally, by (B.8),
\[
P(\mathcal{E}) = 1 - P(\tilde{\mathcal{E}}) = 1 - O(K^{-1}).
\]
Combining this property, (B.9) and (B.11)–(B.15) we deduce that
\[
|E\{f(\hat{\eta}) I(\mathcal{E})\} - f(\eta^0)| = O(K^{-1}). \tag{B.17}
\]
The corollary follows from (B.8) and (B.17).
B.3 Proof of (4.8)

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

\[
\text{cov}(S_r, S_s) = E\left\{ \text{cov}(S_r, S_s \mid \mathcal{X}) \right\} + \text{cov}\left\{ E(S_r \mid \mathcal{X}) E(S_s \mid \mathcal{X}) \right\} \\
= E\left\{ \text{cov}(S_r, S_s \mid \mathcal{X}) \right\} = \frac{1}{k^2} \sum_{i=1}^k E\left\{ \text{cov}(U_{ir}, U_{is} \mid X_i) \right\} \\
= \frac{1}{k^2} \sum_{i=1}^k E\left\{ \text{cov}\left( \left[ \lambda_1(X_i, \eta) + \{\bar{Y}_i - \lambda_1(X_i, \eta)\} \right]^r, \left[ \lambda_1(X_i, \eta) + \{\bar{Y}_i - \lambda_1(X_i, \eta)\} \right]^s \mid X_i \right) \right\} \\
= \frac{1}{k^2} \sum_{i=1}^k E\left( E\left[ \lambda_1(X_i, \eta)^{r+s} + (r + s) \lambda_1(X_i, \eta)^{r+s-1} \{\bar{Y}_i - \lambda_1(X_i, \eta)\} \right] \mid X_i \right) \\
+ \frac{1}{2} (r + s) (r + s - 1) \lambda_1(X_i, \eta)^{r+s-2} \{\bar{Y}_i - \lambda_1(X_i, \eta)\}^2 \mid X_i \\
- \left[ E\left\{ \lambda_1(X_i, \eta)^{r} + r \lambda_1(X_i, \eta)^{r-1} \{\bar{Y}_i - \lambda_1(X_i, \eta)\} \right\} \mid X_i \right] \\
+ \frac{1}{2} r (r - 1) \lambda_1(X_i, \eta)^{r-2} \{\bar{Y}_i - \lambda_1(X_i, \eta)\}^2 \mid X_i \right) \\
\times E\left[ \lambda_1(X_i, \eta)^{s} + s \lambda_1(X_i, \eta)^{s-1} \{\bar{Y}_i - \lambda_1(X_i, \eta)\} \right] \\
+ \frac{1}{2} s (s - 1) \lambda_1(X_i, \eta)^{s-2} \{\bar{Y}_i - \lambda_1(X_i, \eta)\}^2 \mid X_i \right) \right] + o(K^{-1}) \\
= \frac{1}{k^2} \sum_{i=1}^k E\left( \lambda_1(X_i, \eta)^{r+s} + \frac{1}{2} (r + s) (r + s - 1) \lambda_1(X_i, \eta)^{r+s-2} \\
\times E\left[ \{\bar{Y}_i - \lambda_1(X_i, \eta)\}^2 \mid X_i \right] \\
- \left( \lambda_1(X_i, \eta)^{r} + \frac{1}{2} r (r - 1) \lambda_1(X_i, \eta)^{r-2} \right) \\
\times E\left[ \{\bar{Y}_i - \lambda_1(X_i, \eta)\}^2 \mid X_i \right] \right)^2 \right) \\
\times \left( \lambda_1(X_i, \eta)^{s} + \frac{1}{2} s (s - 1) \lambda_1(X_i, \eta)^{s-2} \right) \\
\times E\left[ \{\bar{Y}_i - \lambda_1(X_i, \eta)\}^2 \mid X_i \right] \right)^2 \right) + o(K^{-1}) \\
= \frac{1}{2k^2} \left\{ (r + s) (r + s - 1) - r (r - 1) - s (s - 1) \right\} \\
\times \sum_{i=1}^k n_i^{-1} E\left\{ \lambda_1(X_i, \eta)^{r+s-2} \text{var}(Y_i \mid X_i) \right\} + o(K^{-1}) \\
= rs K^{-1} E\left\{ \lambda_1(X, \eta)^{r+s-2} \text{var}(Y \mid X) \right\} + o(K^{-1}) ,
\]
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 \( \hat{\eta} \) is the solution in \( \eta \) of the equation:

\[
\frac{1}{k} \sum_{i=1}^{k} \sum_{r=1}^{q} \left\{ a_{irs}^\prime (X_i, \eta^0) \hat{R}_{ir}(\eta^0) + a_{irs}(X_i, \eta^0) \hat{R}_{ir}^\prime (\eta^0) \right\}^T (\eta - \eta^0) \\
+ \Theta_s(\eta) (|\bar{Y}| + 1) \| \eta - \eta^0 \|^2 = -\frac{1}{k} \sum_{i=1}^{k} \{ A_i(X_i, \eta^0) \hat{R}_i(\eta^0) \}_s,
\]

where \( \hat{R}_{ir}^\prime = (\partial / \partial \eta) \hat{R}_{ir} \) is a \( p \)-vector. Equivalently, in place of (B.4),

\[
(\hat{M} + \hat{L}) (\eta - \eta^0) + \Theta(\eta) \| \eta - \eta^0 \|^2 = \hat{S}.
\]

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

**B.5 Proof of nonsingularity of \( A(X_i, \eta) \) 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 \( \mu_{ir} = E[\{ \bar{Y}_i - \lambda_1(X_i, \eta) \}^r | X_i] \), the first of the following results can be deduced from (2.16):

\[
\Sigma_i = \begin{pmatrix}
\mu_{i2} & \mu_{i3} \\
\mu_{i3} & \mu_{i4} - \mu_{i2}^2
\end{pmatrix}, \quad \Sigma_i^{-1} = |\Sigma_i|^{-1} \begin{pmatrix}
\mu_{i4} - \mu_{i2}^2 & -\mu_{i3} \\
-\mu_{i3} & \mu_{i2}
\end{pmatrix},
\]
\[
\lambda'_r(X_i, \eta) = \begin{pmatrix}
\lambda'_{11}(X_i, \eta) \\
\lambda'_{12}(X_i, \eta)
\end{pmatrix}, \quad
\dot{D}_i = \begin{pmatrix}
\lambda'_{11}(X_i, \eta) & n^{-1} \lambda'_{21}(X_i, \eta) \\
\lambda'_{12}(X_i, \eta) & n^{-1} \lambda'_{22}(X_i, \eta)
\end{pmatrix},
\]

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

\[
\mu_{i4} - \mu_{i2}^2 = n^{-3} \lambda_4(X_i, \eta) + 3 \left(n^{-1} \lambda_2(X_i, \eta)\right)^2 - \left(n^{-1} \lambda_2(X_i, \eta)\right)^2 \\
= n^{-3} \lambda_4(X_i, \eta) + 2 \left(n^{-1} \lambda_2(X_i, \eta)\right)^2,
\]

\[
|\Sigma_i| = \mu_{i2} \mu_{i4} - \mu_{i2}^3 - \mu_{i3}^2 = 2 \mu_{i2}^3 + O_p(n^{-4}) = 2 n^{-3} \lambda_2(X_i, \eta)^3 + O_p(n^{-4}).
\]

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

\[
\Sigma_i^{-1} = \frac{1 + O_p(n^{-1})}{2 n^{-3} \lambda_2^3} \begin{pmatrix}
-2 n^{-2} \lambda_2^2 & n^{-2} \lambda_3 \\
-n^{-2} \lambda_3 & n^{-1} \lambda_2
\end{pmatrix}.
\]

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

\[
n^{-1} \dot{D}_i \Sigma_i^{-1} = \frac{1 + O_p(n^{-1})}{2 \lambda_2^3} \begin{pmatrix}
2 \lambda_2^2 \lambda'_{11} + n^{-1} (\lambda_4 \lambda'_{11} - \lambda_3 \lambda'_{21}) & \lambda_2 \lambda'_{21} - \lambda_3 \lambda'_{11} \\
2 \lambda_2^2 \lambda'_{12} + n^{-1} (\lambda_4 \lambda'_{12} - \lambda_3 \lambda'_{22}) & \lambda_2 \lambda'_{22} - \lambda_3 \lambda'_{12}
\end{pmatrix} \begin{pmatrix}
c_{11} \\
c_{12}
\end{pmatrix} = \begin{pmatrix}
c_{21} \\
c_{22}
\end{pmatrix},
\]

say, where \( \lambda_r \) and \( \lambda'_{rj} \) denote \( \lambda_r(X_i, \eta) \) and \( \lambda'_{rj}(X_i, \eta) \), respectively, and we have suppressed the dependence of these quantities and the coefficients \( c_{st} \) on \( i \).
Since $\hat{R}_{ir} = \{\bar{Y}_i - \lambda_1(X_i, \eta)\}^r - \hat{Q}_{ir}$ then

$$-\hat{R}_r' = -\partial_\eta \hat{R}_{ir} = r \{\bar{Y}_i - \lambda_1(X_i, \eta)\}^{r-1} \lambda_1' + \partial_\eta \hat{Q}_{ir}$$

$$= r \{\bar{Y}_i - \lambda_1(X_i, \eta)\}^{r-1} \left( \begin{array}{c} \lambda_1' \\ \lambda_1' \\ \end{array} \right) + \begin{cases} 0 & \text{if } r = 1 \\ n^{-1} \left( \begin{array}{c} \lambda_1' \\ \lambda_1' \\ \end{array} \right) & \text{if } r = 2, \end{cases}$$

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

$$\hat{R}_{ir}(\eta) = \hat{R}_{ir}(\eta^0) + \hat{R}_r'(\eta^0)^T (\eta - \eta^0) + O_p(\|\Delta\|^2)$$

$$= \hat{R}_{ir}(\eta^0) - r \{\bar{Y}_i - \lambda_1(X_i, \eta)\}^{r-1} (\lambda_1' \Delta_1 + \lambda_1' \Delta_2)$$

$$+ O_p(\|\Delta\|^2) - \begin{cases} 0 & \text{if } r = 1 \\ n^{-1} (\lambda_2^1 \Delta_1 + \lambda_2^1 \Delta_2) & \text{if } r = 2. \end{cases}$$

Hence,

$$n^{-1} \sum_{i=1}^k \tilde{D}_i \Sigma_i^{-1} \hat{R}_i$$

$$= \sum_{i=1}^k \left( \begin{array}{cc} c_{11} & c_{12} \\ c_{21} & c_{22} \end{array} \right) \left\{ \hat{R}_i(\eta^0) - \left( \begin{array}{c} \lambda_1' \Delta_1 + \lambda_1' \Delta_2 \\ 2 \{\bar{Y}_i - \lambda_1(X_i, \eta)\} (\lambda_1' \Delta_1 + \lambda_1' \Delta_2) + n^{-1} (\lambda_2^1 \Delta_1 + \lambda_2^1 \Delta_2) \end{array} \right) \right\}$$

$$+ O_p(k \|\Delta\|^2)$$

$$= \sum_{i=1}^k \left\{ \left( \begin{array}{cc} c_{11} & c_{12} \\ c_{21} & c_{22} \end{array} \right) \hat{R}_i(\eta^0) - \left( \begin{array}{cc} c_{11} (\lambda_1' \Delta_1 + \lambda_1' \Delta_2) + c_{12} \delta \\ c_{21} (\lambda_1' \Delta_1 + \lambda_1' \Delta_2) + c_{22} \delta \end{array} \right) \right\} + (\text{negligible}),$$

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

Now,

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

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

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

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

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

and so

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

Therefore,

\[
n^{-1} \sum_{i=1}^{k} \tilde{D}_i \Sigma_i^{-1} \tilde{R}_i = \sum_{i=1}^{k} \left\{ \begin{pmatrix} c_{11}(i) \{ \bar{Y}_i - \lambda_1(X_i, \eta) \} \\
               c_{21}(i) \{ \bar{Y}_i - \lambda_1(X_i, \eta) \} \end{pmatrix} \begin{pmatrix} a_1 & a_2 \\
                                       a_2 & a_3 \end{pmatrix} \begin{pmatrix} \Delta_1 \\
                                       \Delta_2 \end{pmatrix} \right\} + \text{(negligible)}. 
\]
The condition on $M^0 M^0$ 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:

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

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

where

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

We can solve this for $\Delta_1$ and $\Delta_2$, obtaining the result that $\Delta_1$ and $\Delta_2$ both equal linear combinations in $k^{-1} \sum_i \tilde{R}_{i1}$ and $n k^{-1} \sum_i \tilde{R}_{i2}$. The second of these is only $O(k^{-1})$. 

13