Generalized Functional Linear Models with Semiparametric Single-Index Interactions

Web Appendix

Yehua Li
Department of Statistics, University of Georgia, Athens, GA 30602, yehuali@uga.edu
Naisyin Wang
Department of Statistics, University of Michigan, Ann Arbor, MI 48109-1107, nwangaa@umich.edu
Raymond J. Carroll
Department of Statistics, Texas A&M University, TAMU 3143, College Station, TX 77843-3143, carroll@stat.tamu.edu

A. Algorithm Implementation With One-step Updates

To speed up the estimation algorithm, we use a one-step update version of the iterative algorithm in Stages 1 and 2, as follows. The key is to replace the full optimization steps of minimizing (5) within the iterative updating of $\mathcal{S}(\cdot)$ and $\boldsymbol{\Theta}$ by a one-step Newton-Raphson algorithm. This eliminates the computational burden of full iteration within the inner loop when the estimates of $\mathcal{S}(\cdot)$ and $\boldsymbol{\Theta}$ are still far away from the target. As a consequence, it also speeds up the iterative process.

Recall that $q_1(x,y) = \{y - g^{-1}(x)\}\rho_1(x)$, $q_2(x,y) = \{y - g^{-1}(x)\}\rho_1'(x) - \rho_2(x)$, where $\rho_\ell(x) = \{dg^{-1}(x)/dx\}^\ell / V\{g^{-1}(x)\}$, $\ell = 1,2$, and $\boldsymbol{Z}_{ij,1} = \boldsymbol{Z}_{i1} - \boldsymbol{Z}_{j1}$. When we fix the parametric component, $\boldsymbol{\Theta}$, and minimize (5) with respect to (a_{0j}, a_{1j}), we use a Newton-Raphson algorithm to perform the following one-step update:

$$\begin{pmatrix} \widehat{a}_{0j} \\ \widehat{a}_{1j} \end{pmatrix} = \begin{pmatrix} \check{a}_{0j} \\ \check{a}_{1j} \end{pmatrix} - \left\{ \textstyle\sum_{i=1}^n w_{ij} q_2(\check{\eta}_{ij}, Y_i)(\check{\boldsymbol{\alpha}}_2^{\mathrm{T}} \boldsymbol{\xi}_i)^2 \begin{pmatrix} 1 \\ \check{\boldsymbol{\theta}}^{\mathrm{T}} \boldsymbol{Z}_{ij,1} \end{pmatrix} \begin{pmatrix} 1 \\ \check{\boldsymbol{\theta}}^{\mathrm{T}} \boldsymbol{Z}_{ij,1} \end{pmatrix}^{\mathrm{T}} \right\}^{-1}$$
$$\times \left\{ \textstyle\sum_{i=1}^n w_{ij} q_1(\check{\eta}_{ij}, Y_i)(\check{\boldsymbol{\alpha}}_2^{\mathrm{T}} \boldsymbol{\xi}_i) \begin{pmatrix} 1 \\ \check{\boldsymbol{\theta}}^{\mathrm{T}} \boldsymbol{Z}_{ij,1} \end{pmatrix} \right\}, \qquad \text{(A.1)}$$

where $\check{\eta}_{ij} = \{\check{\boldsymbol{\alpha}}_1 + (\check{a}_{0j} + \check{a}_{1j}\check{\boldsymbol{\theta}}^{\mathrm{T}}\boldsymbol{Z}_{ij,1})\check{\boldsymbol{\alpha}}_2\}^{\mathrm{T}}\boldsymbol{\xi}_i + \check{\boldsymbol{\beta}}^{\mathrm{T}}\boldsymbol{Z}_i$: all estimates on the right-hand-side of (A.1), and later (A.2) and (A.3), denote their values before the update. To enforce the constraint that $n^{-1}\sum_{i=1}^n \mathcal{S}(\boldsymbol{\theta}^{\mathrm{T}}\boldsymbol{Z}_{i1}) = 0$, we set $\widehat{a}_{0j} = \widehat{a}_{0j} - n^{-1}\sum_{\ell=1}^n \widehat{a}_{0\ell}$.

Similarly, we update the parametric components other than $\boldsymbol{\theta}$ by

$$\begin{pmatrix} \widehat{\boldsymbol{\alpha}}_1 \\ \widehat{\boldsymbol{\alpha}}_2 \\ \widehat{\boldsymbol{\beta}} \end{pmatrix} = \begin{pmatrix} \check{\boldsymbol{\alpha}}_1 \\ \check{\boldsymbol{\alpha}}_2 \\ \check{\boldsymbol{\beta}} \end{pmatrix} - \left\{ \sum_{j=1}^n \sum_{i=1}^n w_{ij} q_2(\check{\eta}_{ij}, Y_i) \begin{pmatrix} \boldsymbol{\xi}_i \\ \widehat{a}_{0j}\boldsymbol{\xi}_i \\ \boldsymbol{Z}_i \end{pmatrix} \begin{pmatrix} \boldsymbol{\xi}_i \\ \widehat{a}_{0j}\boldsymbol{\xi}_i \\ \boldsymbol{Z}_i \end{pmatrix}^{\mathrm{T}} \right\}^{-1} \times \left\{ \sum_{j=1}^n \sum_{i=1}^n w_{ij} q_1(\check{\eta}_{ij}, Y_i) \begin{pmatrix} \boldsymbol{\xi}_i \\ \widehat{a}_{0j}\boldsymbol{\xi}_i \\ \boldsymbol{Z}_i \end{pmatrix} \right\}. \tag{A.2}$$

To enforce the constraint, we adjust the value of $\boldsymbol{\alpha}_2$ by letting $\widehat{\boldsymbol{\alpha}}_2 = \mathrm{sign}(\widehat{\alpha}_{21}) \times \widehat{\boldsymbol{\alpha}}_2/\|\widehat{\boldsymbol{\alpha}}_2\|$.

Finally, we update $\boldsymbol{\theta}$;

$$\widehat{\boldsymbol{\theta}} = \check{\boldsymbol{\theta}} - \left\{ \sum_{j=1}^n \sum_{i=1}^n w_{ij} q_2(\check{\eta}_{ij}, Y_i)(\widehat{a}_{1j}\widehat{\boldsymbol{\alpha}}_2^{\mathrm{T}}\boldsymbol{\xi}_i)^2 \boldsymbol{Z}_{ij,1}\boldsymbol{Z}_{ij,1}^{\mathrm{T}} \right\}^{-1} \left\{ \sum_{j=1}^n \sum_{i=1}^n w_{ij} q_1(\check{\eta}_{ij}, Y_i)(\widehat{a}_{1j}\widehat{\boldsymbol{\alpha}}_2^{\mathrm{T}}\boldsymbol{\xi}_i)\boldsymbol{Z}_{ij,1} \right\} \tag{A.3}$$

To adjust for the constraint on $\boldsymbol{\theta}$, we let $\widehat{\boldsymbol{\theta}} = \mathrm{sign}(\widehat{\theta}_1) \times \widehat{\boldsymbol{\theta}}/\|\widehat{\boldsymbol{\theta}}\|$.

B. Technical Proofs

We sketch the proofs of our main theoretical results. Throughout, $\eta\{\boldsymbol{\alpha}, \boldsymbol{\beta}, \boldsymbol{\theta}, \mathcal{S}(\cdot)\} = \eta\{\boldsymbol{\Theta}, \mathcal{S}(\cdot)\}$ denotes the expression $\boldsymbol{\alpha}_1^{\mathrm{T}}\boldsymbol{\xi} + \mathcal{S}(\boldsymbol{\theta}^{\mathrm{T}}\boldsymbol{Z}_1)\boldsymbol{\alpha}_2^{\mathrm{T}}\boldsymbol{\xi} + \boldsymbol{\beta}^{\mathrm{T}}\boldsymbol{Z}$, and the subindices "$i$" and "$ij$" of η indicate the replacement of each random variable by the corresponding observations. Provided that the clarity of the presentation is preserved, we also leave out arguments of functions to simplify certain equations.

B.1 Proof of Theorem 1

For the iterative estimation procedure proposed in Section 3.1, we denote the current value of the parameter with a subscript "curr", and the previous values with a subscript "prev".

For the first step, we fix $\widetilde{\boldsymbol{\Theta}}_{\mathrm{prev}}$, and update $\mathcal{S}(\mathbf{z})$ by minimizing (4). We let $a_0 = \mathcal{S}_0(\boldsymbol{\theta}_0^{\mathrm{T}}\mathbf{z})$, $a_1^* = b\mathcal{S}_0'(\boldsymbol{\theta}_0^{\mathrm{T}}\mathbf{z})$, denote $\boldsymbol{Z}_{i0,1}^* = (\boldsymbol{Z}_{i1} - \mathbf{z})/b$, $a_1^* = ba_1$ and define, for any $\mathbf{z} \in \mathbb{R}^{d_1}$, the multivariate kernel weight as $w_i(\mathbf{z}) = [\sum_{\ell=1}^n H\{(\boldsymbol{Z}_{\ell 1} - \mathbf{z})/b\}]^{-1} H\{(\boldsymbol{Z}_{i1} - \mathbf{z})/b\}$, where b is the bandwidth. Recall that $\boldsymbol{Z}_{ij,1} = \boldsymbol{Z}_{i1} - \boldsymbol{Z}_{j1}$. The updated estimate $(\widetilde{a}_{0,\mathrm{curr}}, \widetilde{a}^*_{1,\mathrm{curr}})^{\mathrm{T}}$ satisfies the local estimating equation

$$\begin{aligned} \mathbf{0} &= \textstyle\sum_{i=1}^n w_i(\mathbf{z}) q_1[\eta_i\{\widetilde{\boldsymbol{\alpha}}_{\mathrm{prev}}, \widetilde{\boldsymbol{\beta}}_{\mathrm{prev}}, \widetilde{\boldsymbol{\theta}}_{\mathrm{prev}}, (\widetilde{a}_{0,\mathrm{curr}} + \widetilde{a}^*_{1,\mathrm{curr}} \widetilde{\boldsymbol{\theta}}^{\mathrm{T}}_{\mathrm{prev}} \boldsymbol{Z}^*_{i0,1})\}, Y_i] \\ &\quad \times (\widetilde{\boldsymbol{\alpha}}^{\mathrm{T}}_{2,\mathrm{prev}} \boldsymbol{\xi}_i) \begin{pmatrix} 1 \\ \widetilde{\boldsymbol{\theta}}^{\mathrm{T}}_{\mathrm{prev}} \boldsymbol{Z}^*_{i0,1} \end{pmatrix}. \end{aligned}$$

By a standard Taylor expansion, we have

$$\begin{aligned} \mathbf{0} &= \textstyle\sum_{i=1}^n w_i(\mathbf{z}) q_1[\eta_i\{\widetilde{\boldsymbol{\alpha}}_{\mathrm{prev}}, \widetilde{\boldsymbol{\beta}}_{\mathrm{prev}}, \widetilde{\boldsymbol{\theta}}_{\mathrm{prev}}, (a_0 + a^*_1 \widetilde{\boldsymbol{\theta}}^{\mathrm{T}}_{\mathrm{prev}} \boldsymbol{Z}^*_{i0,1})\}, Y_i] (\widetilde{\boldsymbol{\alpha}}^{\mathrm{T}}_{2,\mathrm{prev}} \boldsymbol{\xi}_i) \begin{pmatrix} 1 \\ \widetilde{\boldsymbol{\theta}}^{\mathrm{T}}_{\mathrm{prev}} \boldsymbol{Z}^*_{i0,1} \end{pmatrix} \\ &\quad + \Big(\textstyle\sum_{i=1}^n w_i(\mathbf{z}) q_2[\eta_i\{\widetilde{\boldsymbol{\alpha}}_{\mathrm{prev}}, \widetilde{\boldsymbol{\beta}}_{\mathrm{prev}}, \widetilde{\boldsymbol{\theta}}_{\mathrm{prev}}, (a_0 + a^*_1 \widetilde{\boldsymbol{\theta}}^{\mathrm{T}}_{\mathrm{prev}} \boldsymbol{Z}^*_{i0,1})\}, Y_i] (\widetilde{\boldsymbol{\alpha}}^{\mathrm{T}}_{2,\mathrm{prev}} \boldsymbol{\xi}_i)^2 \\ &\qquad \times \begin{pmatrix} 1 \\ \widetilde{\boldsymbol{\theta}}^{\mathrm{T}}_{\mathrm{prev}} \boldsymbol{Z}^*_{i0,1} \end{pmatrix} \begin{pmatrix} 1 \\ \widetilde{\boldsymbol{\theta}}^{\mathrm{T}}_{\mathrm{prev}} \boldsymbol{Z}^*_{i0,1} \end{pmatrix}^{\mathrm{T}} \Big) \begin{pmatrix} \widetilde{a}_{0,\mathrm{curr}} - a_0 \\ \widetilde{a}^*_{1,\mathrm{curr}} - a^*_1 \end{pmatrix} \times \{1 + o_p(1)\}. \end{aligned} \tag{B.1}$$

Denote $\delta^*_n = b^2 + \{\log(n)/(nb^{d_1})\}^{1/2}$ and $\Delta(\widetilde{\boldsymbol{\Theta}}_{\mathrm{prev}}) = \|(\widetilde{\boldsymbol{\alpha}}^{\mathrm{T}}_{\mathrm{prev}}, \widetilde{\boldsymbol{\beta}}^{\mathrm{T}}_{\mathrm{prev}})^{\mathrm{T}} - (\boldsymbol{\alpha}_0^{\mathrm{T}}, \boldsymbol{\beta}_0^{\mathrm{T}})^{\mathrm{T}}\|$. Define

$$\begin{aligned} \boldsymbol{\gamma}_1(\mathbf{z}; \boldsymbol{\Theta}, \mathcal{S}) &= \mathrm{E}[q_1\{\eta_i(\boldsymbol{\Theta}, \mathcal{S}), Y_i\}(\boldsymbol{\alpha}_2^{\mathrm{T}} \boldsymbol{\xi}_i) | \boldsymbol{Z}_{i1} = \mathbf{z}], \\ \boldsymbol{V}_1(\mathbf{z}; \boldsymbol{\Theta}, \mathcal{S}) &= -\mathrm{E}[q_2\{\eta_i(\boldsymbol{\Theta}, \mathcal{S}), Y_i\}(\boldsymbol{\alpha}_2^{\mathrm{T}} \boldsymbol{\xi}_i)^2 | \boldsymbol{Z}_{i1} = \mathbf{z}]. \end{aligned}$$

By (B.1) and some standard derivation, we obtain that, uniformly for $\mathbf{z} \in \mathcal{D}$,

$$\begin{aligned} & \textstyle\sum_{i=1}^n w_i(\mathbf{z}) q_1[\eta_i\{\widetilde{\boldsymbol{\alpha}}_{\mathrm{prev}}, \widetilde{\boldsymbol{\beta}}_{\mathrm{prev}}, \widetilde{\boldsymbol{\theta}}_{\mathrm{prev}}, (a_0 + a^*_1 \widetilde{\boldsymbol{\theta}}^{\mathrm{T}}_{\mathrm{prev}} \boldsymbol{Z}^*_{i0,1})\}, Y_i] (\widetilde{\boldsymbol{\alpha}}^{\mathrm{T}}_{2,\mathrm{prev}} \boldsymbol{\xi}_i) \\ = & \textstyle\sum_{i=1}^n w_i(\mathbf{z}) q_1[\widetilde{\boldsymbol{\alpha}}^{\mathrm{T}}_{1,\mathrm{prev}} \boldsymbol{\xi}_i + (a_0 + a^*_1 \widetilde{\boldsymbol{\theta}}^{\mathrm{T}}_{\mathrm{prev}} \boldsymbol{Z}^*_{i0,1}) \widetilde{\boldsymbol{\alpha}}^{\mathrm{T}}_{2,\mathrm{prev}} \boldsymbol{\xi}_i + \widetilde{\boldsymbol{\beta}}^{\mathrm{T}}_{\mathrm{prev}} \boldsymbol{Z}_i, Y_i] (\widetilde{\boldsymbol{\alpha}}^{\mathrm{T}}_{2,\mathrm{prev}} \boldsymbol{\xi}_i) \\ = & \boldsymbol{\gamma}_1(\mathbf{z}; \widetilde{\boldsymbol{\alpha}}_{\mathrm{prev}}, \widetilde{\boldsymbol{\beta}}_{\mathrm{prev}}, \boldsymbol{\theta}_0, \mathcal{S}_0) + O_p(\delta^*_n), \end{aligned}$$

and

$$\begin{aligned} & \textstyle\sum_{i=1}^n w_i(\mathbf{z}) q_1\{\widetilde{\boldsymbol{\alpha}}^{\mathrm{T}}_{1,\mathrm{prev}} \boldsymbol{\xi}_i + (a_0 + a^*_1 \widetilde{\boldsymbol{\theta}}^{\mathrm{T}}_{\mathrm{prev}} \boldsymbol{Z}^*_{i0,1}) \widetilde{\boldsymbol{\alpha}}^{\mathrm{T}}_{2,\mathrm{prev}} \boldsymbol{\xi}_i + \widetilde{\boldsymbol{\beta}}^{\mathrm{T}}_{\mathrm{prev}} \boldsymbol{Z}_i, Y_i\} (\widetilde{\boldsymbol{\alpha}}^{\mathrm{T}}_{2,\mathrm{prev}} \boldsymbol{\xi}_i) \widetilde{\boldsymbol{\theta}}^{\mathrm{T}}_{\mathrm{prev}} \boldsymbol{Z}^*_{i0,1} \\ = & \widetilde{\boldsymbol{\theta}}^{\mathrm{T}}_{\mathrm{prev}} \textstyle\sum_{i=1}^n w_i(\mathbf{z}) q_1\{\widetilde{\boldsymbol{\alpha}}^{\mathrm{T}}_{1,\mathrm{prev}} \boldsymbol{\xi}_i + \mathcal{S}(\boldsymbol{\theta}_0 \boldsymbol{Z}_{i1}) \widetilde{\boldsymbol{\alpha}}^{\mathrm{T}}_{2,\mathrm{prev}} \boldsymbol{\xi}_i + \widetilde{\boldsymbol{\beta}}^{\mathrm{T}}_{\mathrm{prev}} \boldsymbol{Z}_i, Y_i\} (\widetilde{\boldsymbol{\alpha}}^{\mathrm{T}}_{2,\mathrm{prev}} \boldsymbol{\xi}_i) \boldsymbol{Z}^*_{i0,1} \\ & \quad + \Big[\textstyle\sum_{i=1}^n w_i(\mathbf{z}) \{\mathcal{S}_0(\boldsymbol{\theta}_0^{\mathrm{T}} \mathbf{z}) - \mathcal{S}_0(\boldsymbol{\theta}_0^{\mathrm{T}} \boldsymbol{Z}_{i1}) + b \mathcal{S}'(\boldsymbol{\theta}_0^{\mathrm{T}} \mathbf{z}) \widetilde{\boldsymbol{\theta}}^{\mathrm{T}}_{\mathrm{prev}} \boldsymbol{Z}^*_{i0,1}\} \\ & \qquad \times q_2\{\widetilde{\boldsymbol{\alpha}}^{\mathrm{T}}_{1,\mathrm{prev}} \boldsymbol{\xi}_i + \mathcal{S}(\boldsymbol{\theta}_0 \mathbf{z}) \widetilde{\boldsymbol{\alpha}}^{\mathrm{T}}_{2,\mathrm{prev}} \boldsymbol{\xi}_i + \widetilde{\boldsymbol{\beta}}^{\mathrm{T}}_{\mathrm{prev}} \boldsymbol{Z}_i, Y_i\} (\widetilde{\boldsymbol{\alpha}}^{\mathrm{T}}_{2,\mathrm{prev}} \boldsymbol{\xi}_i)^2 (\widetilde{\boldsymbol{\theta}}^{\mathrm{T}}_{\mathrm{prev}} \boldsymbol{Z}^*_{i0,1}) \Big] \times \{1 + o_p(1)\} \end{aligned}$$

$$
\begin{aligned}
&= b\widetilde{\boldsymbol{\theta}}_{\rm prev}^{\rm T}\{\frac{\partial}{\partial \mathbf{z}}\boldsymbol{\gamma}_1 + \frac{1}{f_{\boldsymbol{Z}_1}(\mathbf{z})}\frac{\partial}{\partial \mathbf{z}} f_{\boldsymbol{Z}_1}(\mathbf{z})\boldsymbol{\gamma}_1\}(\mathbf{z};\widetilde{\boldsymbol{\alpha}}_{\rm prev},\widetilde{\boldsymbol{\beta}}_{\rm prev},\boldsymbol{\theta}_0,\mathcal{S}_0)\\
&\qquad -b(1-\widetilde{\boldsymbol{\theta}}_{\rm prev}^{\rm T}\boldsymbol{\theta}_0)\mathcal{S}'(\boldsymbol{\theta}_0\mathbf{z})\boldsymbol{V}_1(\mathbf{z};\widetilde{\boldsymbol{\alpha}}_{\rm prev},\widetilde{\boldsymbol{\beta}}_{\rm prev},\boldsymbol{\theta}_0,\mathcal{S}_0) + O_p(\delta_n^*)\\
&= -b(1-\widetilde{\boldsymbol{\theta}}_{\rm prev}^{\rm T}\boldsymbol{\theta}_0)\mathcal{S}'(\boldsymbol{\theta}_0\mathbf{z})\boldsymbol{V}_1(\mathbf{z};\widetilde{\boldsymbol{\alpha}}_{\rm prev},\widetilde{\boldsymbol{\beta}}_{\rm prev},\boldsymbol{\theta}_0,\mathcal{S}_0) + O_p\{b\Delta(\widetilde{\boldsymbol{\Theta}}_{\rm prev})\} + O_p(\delta_n^*),
\end{aligned}
$$

By similar calculations, we can derive from (B.1) that, uniformly for $\mathbf{z} \in \mathcal{D}$,

$$
\begin{aligned}
\begin{pmatrix} \widetilde{a}_{0,\rm curr} - a_0 \\ \widetilde{a}_{1,\rm curr}^* - a_1^* \end{pmatrix} &= \left\{ \begin{array}{cc} \boldsymbol{V}_1(\mathbf{z};\widetilde{\boldsymbol{\alpha}}_{\rm prev},\widetilde{\boldsymbol{\beta}}_{\rm prev},\boldsymbol{\theta}_0,\mathcal{S}_0) & \boldsymbol{V}_{12} \\ \boldsymbol{V}_{12} & \boldsymbol{V}_1(\mathbf{z};\widetilde{\boldsymbol{\alpha}}_{\rm prev},\widetilde{\boldsymbol{\beta}}_{\rm prev},\boldsymbol{\theta}_0,\mathcal{S}_0) \end{array} \right\}^{-1} \\
&\quad \times \left\{ \begin{array}{c} \boldsymbol{\gamma}_1(\mathbf{z};\widetilde{\boldsymbol{\alpha}}_{\rm prev},\widetilde{\boldsymbol{\beta}}_{\rm prev},\boldsymbol{\theta}_0,\mathcal{S}_0) \\ -b(1-\widetilde{\boldsymbol{\theta}}_{\rm prev}^{\rm T}\boldsymbol{\theta}_0)\mathcal{S}'(\boldsymbol{\theta}_0\mathbf{z})\boldsymbol{V}_1(\mathbf{z};\widetilde{\boldsymbol{\alpha}}_{\rm prev},\widetilde{\boldsymbol{\beta}}_{\rm prev},\boldsymbol{\theta}_0,\mathcal{S}_0) + O_p\{b\Delta(\widetilde{\boldsymbol{\Theta}}_{\rm prev})\} \end{array} \right\} \\
&\quad + O_p(\delta_n^*),
\end{aligned}
$$

where

$$
\boldsymbol{V}_{12} = b\widetilde{\boldsymbol{\theta}}_{\rm prev}^{T}\{\frac{\partial}{\partial \mathbf{z}}\boldsymbol{V}_2 + \frac{1}{f_{\boldsymbol{Z}_1}(\mathbf{z})}\frac{\partial}{\partial \mathbf{z}} f_{\boldsymbol{Z}_1}(\mathbf{z})\boldsymbol{V}_2\}(\mathbf{z},\widetilde{\boldsymbol{\alpha}}_{\rm prev},\widetilde{\boldsymbol{\beta}}_{\rm prev},\boldsymbol{\theta}_0,\mathcal{S}_0) = O_p(b)
$$

is an negligible term.

Define $\boldsymbol{V}_2(\mathbf{z};\boldsymbol{\Theta},\mathcal{S})$ as

$$
-\mathrm{E}\left[q_2\{\eta_i(\boldsymbol{\Theta},\mathcal{S}),Y_i\}(\boldsymbol{\alpha}_2^{\rm T}\boldsymbol{\xi}_i) \left\{ \begin{array}{c} \boldsymbol{\xi}_i \\ \mathcal{S}(\boldsymbol{\theta}^{\rm T}\boldsymbol{Z}_{i1})\boldsymbol{\xi}_i \\ \boldsymbol{Z}_i \end{array} \right\} \middle| \boldsymbol{Z}_{i1} = \mathbf{z} \right].
$$

By the fact that $\boldsymbol{\gamma}_1(\mathbf{z};\boldsymbol{\Theta}_0,\mathcal{S}_0) = 0$ for all $\mathbf{z}$, a Taylor expansion of $\boldsymbol{\gamma}_1$ on $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$, and further derivations, we obtain

$$
\begin{aligned}
\widetilde{a}_{0,\rm curr} - a_0 &= -\boldsymbol{V}_1^{-1}(\mathbf{z};\widetilde{\boldsymbol{\alpha}}_{\rm prev},\widetilde{\boldsymbol{\beta}}_{\rm prev},\boldsymbol{\theta}_0,\mathcal{S}_0) \\
&\quad \times \left\{ \boldsymbol{V}_2^{\rm T}(\mathbf{z};\boldsymbol{\Theta}_0,\mathcal{S}_0) \begin{pmatrix} \widetilde{\boldsymbol{\alpha}}_{1,\rm prev} - \boldsymbol{\alpha}_{1,0} \\ \widetilde{\boldsymbol{\alpha}}_{2,\rm prev} - \boldsymbol{\alpha}_{2,0} \\ \widetilde{\boldsymbol{\beta}}_{\rm prev} - \boldsymbol{\beta}_0 \end{pmatrix} \right\} \times \{1 + o_p(1)\} + O_p(\delta_n^*), \quad \text{(B.2)}
\end{aligned}
$$

and

$$
\widetilde{a}_{1,\rm curr}^* = b\mathcal{S}'(\boldsymbol{\theta}_0^{\rm T}\mathbf{z})\widetilde{\boldsymbol{\theta}}_{\rm prev}^{\rm T}\boldsymbol{\theta}_0 + O_p\{b\Delta(\widetilde{\boldsymbol{\Theta}}_{\rm prev})\} + O_p(\delta_n^*). \qquad \text{(B.3)}
$$

Next, holding the link function and $\boldsymbol{\theta}$ fixed at their current values, we update $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$. At convergence, $\widetilde{\boldsymbol{\alpha}}_{\text{curr}}$ and $\widetilde{\boldsymbol{\beta}}_{\text{curr}}$ satisfy the estimating equations (B.4),

$$
\begin{aligned}
\mathbf{0} &= \textstyle\sum_{j=1}^{n}\sum_{i=1}^{n} w_{ij} q_1[\eta_i\{\widetilde{\boldsymbol{\alpha}}_{\text{curr}}, \widetilde{\boldsymbol{\beta}}_{\text{curr}}, \widetilde{\boldsymbol{\theta}}_{\text{prev}}, (\widetilde{a}_{0j,\text{curr}} + \widetilde{a}^*_{1j,\text{curr}} \widetilde{\boldsymbol{\theta}}^{\mathrm{T}}_{\text{prev}} \boldsymbol{Z}^*_{ij,1})\}, Y_i] \\
&\qquad \times \left\{ \begin{array}{c} \boldsymbol{\xi}_i \\ (\widetilde{a}_{0j} + \widetilde{a}^*_{1j} \widetilde{\boldsymbol{\theta}}^{\mathrm{T}}_{\text{prev}} \boldsymbol{Z}^*_{ij,1}) \boldsymbol{\xi}_i \\ \boldsymbol{Z}_i \end{array} \right\} \\
&= \textstyle\sum_{j=1}^{n}\sum_{i=1}^{n} w_{ij} q_1[\eta\{\boldsymbol{\alpha}_0, \boldsymbol{\beta}_0, \widetilde{\boldsymbol{\theta}}_{\text{prev}}, (\widetilde{a}_{0j,\text{curr}} + \widetilde{a}^*_{1j,\text{curr}} \widetilde{\boldsymbol{\theta}}^{\mathrm{T}}_{\text{prev}} \boldsymbol{Z}^*_{ij,1})\}, Y_i] \begin{pmatrix} \boldsymbol{\xi}_i \\ \widetilde{a}_{0j,\text{curr}} \boldsymbol{\xi}_i \\ \boldsymbol{Z}_i \end{pmatrix} \\
&\quad + \Big(\textstyle\sum_{j=1}^{n}\sum_{i=1}^{n} w_{ij} q_2[\eta_{ij}\{\boldsymbol{\alpha}_0, \boldsymbol{\beta}_0, \widetilde{\boldsymbol{\theta}}_{\text{prev}}, (\widetilde{a}_{0j,\text{curr}} + \widetilde{a}^*_{1j,\text{curr}} \widetilde{\boldsymbol{\theta}}^{\mathrm{T}}_{\text{prev}} \boldsymbol{Z}^*_{ij,1})\}, Y_i] \\
&\qquad \times \begin{pmatrix} \boldsymbol{\xi}_i \\ \widetilde{a}_{0j,\text{curr}} \boldsymbol{\xi}_i \\ \boldsymbol{Z}_i \end{pmatrix}^{\otimes 2} \begin{pmatrix} \widetilde{\boldsymbol{\alpha}}_{1,\text{curr}} - \boldsymbol{\alpha}_{1,0} \\ \widetilde{\boldsymbol{\alpha}}_{2,\text{curr}} - \boldsymbol{\alpha}_{2,0} \\ \widetilde{\boldsymbol{\beta}}_{\text{curr}} - \boldsymbol{\beta}_0 \end{pmatrix} \Bigg) \times \{1 + o_p(1)\} + O_p(b^2),
\end{aligned} \tag{B.4}
$$

where $d^{\otimes 2} = dd^{\mathrm{T}}$. Define

$$
\begin{aligned}
\boldsymbol{\gamma}_2(\mathbf{z}; \boldsymbol{\Theta}, \mathcal{S}) &= \mathrm{E}\left[q_1\{\eta_i(\boldsymbol{\Theta}, \mathcal{S}), Y_i\} \left\{ \begin{array}{c} \boldsymbol{\xi}_i \\ \mathcal{S}(\boldsymbol{\theta}^{\mathrm{T}} \boldsymbol{Z}_{i1}) \boldsymbol{\xi}_i \\ \boldsymbol{Z}_i \end{array} \right\} \Big| \boldsymbol{Z}_{i1} = \mathbf{z} \right], \\
\boldsymbol{V}_3(\mathbf{z}; \boldsymbol{\Theta}, \mathcal{S}) &= -\mathrm{E}\left[q_2\{\eta_i(\boldsymbol{\Theta}, \mathcal{S}), Y_i\} \left\{ \begin{array}{c} \boldsymbol{\xi}_i \\ \mathcal{S}(\boldsymbol{\theta}^{\mathrm{T}} \boldsymbol{Z}_{i1}) \boldsymbol{\xi}_i \\ \boldsymbol{Z}_i \end{array} \right\}^{\otimes 2} \Big| \boldsymbol{Z}_{i1} = \mathbf{z} \right].
\end{aligned}
$$

By (B.2), (B.4), the fact $\boldsymbol{\gamma}_2(\mathbf{z}; \boldsymbol{\Theta}_0, \mathcal{S}_0) = 0$ for all $\mathbf{z}$, the expansion $\boldsymbol{\gamma}_2(\boldsymbol{Z}_{j1}; \boldsymbol{\Theta}_0, \widetilde{\mathcal{S}}_{\text{curr}}) = -\boldsymbol{V}_2(\boldsymbol{Z}_{j1}; \boldsymbol{\Theta}_0, \mathcal{S}_0)(\widetilde{a}_{0j,\text{curr}} - a_{0j}) \times \{1 + o_p(1)\}$, we obtain

$$
\begin{aligned}
\begin{pmatrix} \widetilde{\boldsymbol{\alpha}}_{1,\text{curr}} - \boldsymbol{\alpha}_{1,0} \\ \widetilde{\boldsymbol{\alpha}}_{2,\text{curr}} - \boldsymbol{\alpha}_{2,0} \\ \widetilde{\boldsymbol{\beta}}_{\text{curr}} - \boldsymbol{\beta}_0 \end{pmatrix} &= \Big\{ n^{-1} \sum_j \boldsymbol{V}_3(\boldsymbol{Z}_{j1}; \boldsymbol{\Theta}_0, \mathcal{S}_0) \Big\}^{-1} \Big\{ n^{-1} \sum_j (\boldsymbol{V}_2 \boldsymbol{V}_2^{\mathrm{T}} / \boldsymbol{V}_1)(\boldsymbol{Z}_{j1}; \boldsymbol{\Theta}_0, \mathcal{S}_0) \Big\} \\
&\qquad \times \begin{pmatrix} \widetilde{\boldsymbol{\alpha}}_{1,\text{prev}} - \boldsymbol{\alpha}_{1,0} \\ \widetilde{\boldsymbol{\alpha}}_{2,\text{prev}} - \boldsymbol{\alpha}_{2,0} \\ \widetilde{\boldsymbol{\beta}}_{\text{prev}} - \boldsymbol{\beta}_0 \end{pmatrix} \times \{1 + o_p(1)\} + O_p(\delta^*_n).
\end{aligned} \tag{B.5}
$$

One source of additional variation that could be easily missed is the $\widetilde{\boldsymbol{\Theta}}_{\text{prev}}$ embedded within the expressions of $\widehat{a}_{\rm curr}$, as given by (B.2) and (B.3). This is carefully taken into account in our derivations.

Asymptotically, the iterations converge if the distances between the current and previous estimates of $\boldsymbol{\alpha}$'s and $\boldsymbol{\beta}$'s go to zero as the iteration number goes to ∞. By (B.5) and for a large n, this occurs when $\mathrm{E}\{\boldsymbol{V}_3(\boldsymbol{Z}_1;\boldsymbol{\Theta}_0,\mathcal{S}_0)\}^{-1}\mathrm{E}\{(\boldsymbol{V}_2\boldsymbol{V}_2^{\rm T}/\boldsymbol{V}_1)(\boldsymbol{Z}_1;\boldsymbol{\Theta}_0,\mathcal{S}_0)\}$ has eigenvalues strictly less than 1. By the Cauchy-Schwartz inequality, we can show that, for all $\mathbf{z}\in\mathcal{D}$, $(\boldsymbol{V}_2\boldsymbol{V}_2^{\rm T})(\mathbf{z};\boldsymbol{\Theta},\mathcal{S})\leq(\boldsymbol{V}_1\times\boldsymbol{V}_3)(\mathbf{z};\boldsymbol{\Theta},\mathcal{S})$, where equality holds only if the order of the products of nonlinear functions and the conditional expectations are exchangeable, which does not hold in our model. Consequently, this establishes the asymptotic convergence property for the iteration procedure. As a result, at the limit,

$$\left(\widetilde{\boldsymbol{\alpha}}_{1,\rm curr}-\boldsymbol{\alpha}_{1,0},\widetilde{\boldsymbol{\alpha}}_{2,\rm curr}-\widetilde{\boldsymbol{\alpha}}_{2,0},\widetilde{\boldsymbol{\beta}}_{\rm curr}-\boldsymbol{\beta}_0\right)=O_p(\delta_n^*). \tag{B.6}$$

The constraints on $\mathcal{S}(\cdot)$ and $\boldsymbol{\alpha}_2$ guarantee that the estimators converge to the uniquely defined parameters.

Finally, using the $(\widetilde{\boldsymbol{\alpha}},\widetilde{\boldsymbol{\beta}})$ at convergence, we update $\boldsymbol{\theta}$ by solving the local estimating equation

$$\mathbf{0}=n^{-1}\textstyle\sum_{j=1}^n\sum_{i=1}^n w_{ij}q_1[\eta_{ij}\{\widetilde{\boldsymbol{\alpha}},\widetilde{\boldsymbol{\beta}},\widetilde{\boldsymbol{\theta}}_{\rm curr},(\widetilde{a}_{0j}+\widetilde{a}_{1j}^*\widetilde{\boldsymbol{\theta}}_{\rm curr}^{\rm T}\boldsymbol{Z}_{ij,1}^*)\},Y_i](\widetilde{\boldsymbol{\alpha}}_2^{\rm T}\boldsymbol{\xi}_i)\widetilde{a}_{1j}^*\boldsymbol{Z}_{ij,1}^*.$$

We expand the right hand side of the equation as before,

$$\begin{aligned}
\mathbf{0} &= n^{-1}\textstyle\sum_{j=1}^n\sum_{i=1}^n w_{ij}q_1\{\widetilde{\boldsymbol{\alpha}}_1^{\rm T}\boldsymbol{\xi}_i+(\widetilde{a}_{0j}+\widetilde{a}_{1j}^*\widetilde{\boldsymbol{\theta}}_{\rm curr}^{\rm T}\boldsymbol{Z}_{ij,1}^*)\widetilde{\boldsymbol{\alpha}}_2^{\rm T}\boldsymbol{\xi}_i+\widetilde{\boldsymbol{\beta}}^{\rm T}\boldsymbol{Z}_i,Y_i\}(\widetilde{\boldsymbol{\alpha}}_2^{\rm T}\boldsymbol{\xi}_i)\widetilde{a}_{1j}^*\boldsymbol{Z}_{ij,1}^*\\
&= n^{-1}\textstyle\sum_{j=1}^n\sum_{i=1}^n w_{ij}q_1\{\widetilde{\boldsymbol{\alpha}}_1^{\rm T}\boldsymbol{\xi}_i+\mathcal{S}(\boldsymbol{\theta}_0^{\rm T}\boldsymbol{Z}_{i1})\widetilde{\boldsymbol{\alpha}}_2^{\rm T}\boldsymbol{\xi}_i+\widetilde{\boldsymbol{\beta}}^{\rm T}\boldsymbol{Z}_i,Y_i\}(\widetilde{\boldsymbol{\alpha}}_2^{\rm T}\boldsymbol{\xi}_i)\widetilde{a}_{1j}^*\boldsymbol{Z}_{ij,1}^*\\
&\quad+n^{-1}\textstyle\sum_{j=1}^n\sum_{i=1}^n w_{ij}q_2\{\widetilde{\boldsymbol{\alpha}}_1^{\rm T}\boldsymbol{\xi}_i+\mathcal{S}(\boldsymbol{\theta}_0^{\rm T}\boldsymbol{Z}_{i1})\widetilde{\boldsymbol{\alpha}}_2^{\rm T}\boldsymbol{\xi}_i+\widetilde{\boldsymbol{\beta}}^{\rm T}\boldsymbol{Z}_i,Y_i\}(\widetilde{\boldsymbol{\alpha}}_2^{\rm T}\boldsymbol{\xi}_i)^2\widetilde{a}_{1j}^*\boldsymbol{Z}_{ij,1}^*\\
&\qquad\qquad\times(\widetilde{a}_{0j}-a_{0j}+a_{0j}-a_{0i}+\widetilde{a}_{1j}^*\widetilde{\boldsymbol{\theta}}_{\rm curr}^{\rm T}\boldsymbol{Z}_{ij,1}^*)\times\{1+o_p(1)\}\\
&= n^{-1}\textstyle\sum_{j=1}^n b\widetilde{a}_{1j}^*\{\frac{\partial}{\partial\mathbf{z}}+\frac{1}{f_{\boldsymbol{Z}_1}(\boldsymbol{Z}_{j1})}\frac{\partial}{\partial\mathbf{z}}f_{\boldsymbol{Z}_1}(\boldsymbol{Z}_{j1})\}\gamma_1(\boldsymbol{Z}_{j1};\widetilde{\boldsymbol{\alpha}},\widetilde{\boldsymbol{\beta}},\boldsymbol{\theta}_0,\mathcal{S}_0)+O_p(b\delta_n^*)\\
&\quad+n^{-1}\textstyle\sum_{j=1}^n b\widetilde{a}_{1j}^*(\widetilde{a}_{0j}-a_{0j})\{\frac{\partial}{\partial\mathbf{z}}+\frac{1}{f_{\boldsymbol{Z}_1}(\boldsymbol{Z}_{j1})}\frac{\partial}{\partial\mathbf{z}}f_{\boldsymbol{Z}_1}(\boldsymbol{Z}_{j1})\}\boldsymbol{V}_1(\boldsymbol{Z}_{j1};\widetilde{\boldsymbol{\alpha}},\widetilde{\boldsymbol{\beta}},\boldsymbol{\theta}_0,\mathcal{S}_0)\\
&\quad-n^{-1}\textstyle\sum_{j=1}^n b\widetilde{a}_{1j}^*\{-\mathcal{S}_0'(\boldsymbol{\theta}_0^{\rm T}\boldsymbol{Z}_{j1})\}\boldsymbol{V}_1(\boldsymbol{Z}_{j1};\widetilde{\boldsymbol{\alpha}},\widetilde{\boldsymbol{\beta}},\boldsymbol{\theta}_0,\mathcal{S}_0)\boldsymbol{\theta}_0\\
&\quad-n^{-1}\textstyle\sum_{j=1}^n(\widetilde{a}_{1j}^*)^2\boldsymbol{V}_1(\boldsymbol{Z}_{j1};\widetilde{\boldsymbol{\alpha}},\widetilde{\boldsymbol{\beta}},\boldsymbol{\theta}_0,\mathcal{S}_0)\widetilde{\boldsymbol{\theta}}_{\rm curr}+o_p(b^2\delta_n^*).
\end{aligned}$$

By (B.3), we obtain that $\widetilde{\boldsymbol{\theta}}_{\text{curr}} = (\widetilde{\boldsymbol{\theta}}_{\text{prev}}^{\text{T}}\boldsymbol{\theta}_0)^{-1}\boldsymbol{\theta}_0 + O_p(b^{-1}\delta_n^*)$. By standardizing $\widetilde{\boldsymbol{\theta}}_{\text{curr}}$ and correcting the sign of the first entry, we have

$$\widetilde{\boldsymbol{\theta}}_{\text{curr}} - \boldsymbol{\theta}_0 = O_p(b^{-1}\delta_n^*), \tag{B.7}$$

which converges to 0 by Condition (C2.1).

B.2 Proof of Theorem 2

Following similar notation as in Carroll et al. (1997), we denote $\mathcal{U}_i = \boldsymbol{\theta}_0^{\text{T}}\boldsymbol{Z}_{i1}$, $\widehat{\mathcal{U}}_i = \widehat{\boldsymbol{\theta}}^{\text{T}}\boldsymbol{Z}_{i1}$ For the refined estimator, the kernel weight is defined as $w_i(u) = K\{(\widehat{\mathcal{U}}_i - u)/h\}/\sum_{\ell=1}^n K\{(\widehat{\mathcal{U}}_\ell - u)/h\}$ for $u \in \mathbb{R}$. Denote $\delta_n = \{\log(n)/(nh)\}^{1/2}$, $\Delta(\widehat{\boldsymbol{\Theta}}) = \|\widehat{\boldsymbol{\Theta}} - \boldsymbol{\Theta}_0\|$.

First, we fix the value of $\boldsymbol{\Theta}$ at the value from the previous iteration, denoted by $\widehat{\boldsymbol{\Theta}}_{\text{prev}}$. For any $\mathbf{z} \in \mathcal{D} \subset \mathbb{R}^{d_1}$, as in the previous subsection, we let $\boldsymbol{Z}_{i0,1}^* = (\boldsymbol{Z}_{i1} - \mathbf{z})/h$ and $(\widehat{a}_{0,\text{curr}}, \widehat{a}_{1,\text{curr}}^*)$ be the updated estimators for a_0 and a_1^* which solves the local estimation equation (B.8),

$$\begin{aligned}
\mathbf{0} &= \textstyle\sum_{i=1}^n w_i(\widehat{\boldsymbol{\theta}}_{\text{prev}}^{\text{T}}\mathbf{z}) q_1\Big[\eta_i\{\widehat{\boldsymbol{\alpha}}_{\text{prev}}, \widehat{\boldsymbol{\beta}}_{\text{prev}}, \widehat{\boldsymbol{\theta}}_{\text{prev}}, (\widehat{a}_{0,\text{curr}} + \widehat{a}_{1,\text{curr}}^*\widehat{\boldsymbol{\theta}}_{\text{prev}}^{\text{T}}\boldsymbol{Z}_{i0,1}^*)\}, Y_i\Big] \\
&\qquad \times(\widehat{\boldsymbol{\alpha}}_{2,\text{prev}}^{\text{T}}\boldsymbol{\xi}_i)\begin{pmatrix} 1 \\ \widehat{\boldsymbol{\theta}}_{\text{prev}}^{\text{T}}\boldsymbol{Z}_{i0,1}^* \end{pmatrix} \\
&= \textstyle\sum_{i=1}^n w_i(\widehat{\boldsymbol{\theta}}_{\text{prev}}^{\text{T}}\mathbf{z}) q_1(\eta_i, Y_i)(\boldsymbol{\alpha}_{2,0}^{\text{T}}\boldsymbol{\xi}_i)\begin{pmatrix} 1 \\ \widehat{\boldsymbol{\theta}}_{\text{prev}}^{\text{T}}\boldsymbol{Z}_{i0,1}^* \end{pmatrix} \times \{1 + o_p(1)\} \\
&+ \Bigg[\textstyle\sum_{i=1}^n w_i(\widehat{\boldsymbol{\theta}}_{\text{prev}}^{\text{T}}\mathbf{z}) q_2(\eta_i, Y_i)(\boldsymbol{\alpha}_{2,0}^{\text{T}}\boldsymbol{\xi}_i)^2\begin{pmatrix} 1 \\ \widehat{\boldsymbol{\theta}}_{\text{prev}}^{\text{T}}\boldsymbol{Z}_{i0,1}^* \end{pmatrix}\{\widehat{a}_{0,\text{curr}} + \widehat{a}_{1,\text{curr}}^*\widehat{\boldsymbol{\theta}}_{\text{prev}}^{\text{T}}\boldsymbol{Z}_{i0,1}^* - \mathcal{S}(\boldsymbol{\theta}_0^{\text{T}}\boldsymbol{Z}_{i1})\} \\
&+ \textstyle\sum_{i=1}^n w_i(\widehat{\boldsymbol{\theta}}_{\text{prev}}^{\text{T}}\mathbf{z}) q_2(\eta_i, Y_i)(\boldsymbol{\alpha}_{2,0}^{\text{T}}\boldsymbol{\xi}_i)\begin{pmatrix} 1 \\ \widehat{\boldsymbol{\theta}}_{\text{prev}}^{\text{T}}\boldsymbol{Z}_{i0,1}^* \end{pmatrix}\begin{Bmatrix} \boldsymbol{\xi}_i \\ \mathcal{S}(\boldsymbol{\theta}_0^{\text{T}}\boldsymbol{Z}_{i1})\boldsymbol{\xi}_i \\ \boldsymbol{Z}_i \end{Bmatrix}^{\text{T}}\begin{pmatrix} \widehat{\boldsymbol{\alpha}}_{1,\text{prev}} - \boldsymbol{\alpha}_{1,0} \\ \widehat{\boldsymbol{\alpha}}_{2,\text{prev}} - \boldsymbol{\alpha}_{2,0} \\ \widehat{\boldsymbol{\beta}}_{\text{prev}} - \boldsymbol{\beta}_0 \end{pmatrix}\Bigg] \\
&\quad \times\{1 + o_p(1)\}.
\end{aligned} \tag{B.8}$$

For the observations with $|\widehat{\boldsymbol{\theta}}_{\text{prev}}^{\text{T}}\boldsymbol{Z}_{i1} - u| \leq h$, we obtain

$$\begin{aligned}
&\widehat{a}_{0,\text{curr}} + \widehat{a}_{1,\text{curr}}^*\widehat{\boldsymbol{\theta}}_{\text{prev}}^{\text{T}}\boldsymbol{Z}_{i0,1}^* - \mathcal{S}(\boldsymbol{\theta}_0^{\text{T}}\boldsymbol{Z}_{i1}) \\
&= \begin{pmatrix} 1 \\ \widehat{\boldsymbol{\theta}}_{\text{prev}}^{\text{T}}\boldsymbol{Z}_{i0,1}^* \end{pmatrix}^{\text{T}}\begin{pmatrix} \widehat{a}_{0,\text{curr}} - a_0 \\ \widehat{a}_{1,\text{curr}}^* - a_1^* \end{pmatrix} + \mathcal{S}'(\boldsymbol{\theta}_0^{\text{T}}\mathbf{z})\boldsymbol{Z}_{i0,1}^{\text{T}}(\widehat{\boldsymbol{\theta}}_{\text{prev}} - \boldsymbol{\theta}_0) \\
&\qquad -\frac{1}{2}\mathcal{S}^{(2)}(\widehat{\boldsymbol{\theta}}_0\mathbf{z})(\widehat{\boldsymbol{\theta}}_{\text{prev}}^{\text{T}}\boldsymbol{Z}_{i0,1})^2 + O(h^3) + o_p(\|\widehat{\boldsymbol{\theta}}_{\text{prev}} - \boldsymbol{\theta}_0\|).
\end{aligned}$$

Define

$$\mathcal{V}_1(\mathbf{z};\boldsymbol{\theta}) = \mathrm{E}\{\rho_2(\eta_i)(\boldsymbol{\alpha}_{2,0}^{\mathrm{T}}\boldsymbol{\xi}_i)^2|\boldsymbol{\theta}^{\mathrm{T}}\boldsymbol{Z}_{i1} = \boldsymbol{\theta}^{\mathrm{T}}\mathbf{z}\}, \tag{B.9}$$

$$\boldsymbol{\pi}_{i1} = \rho_2(\eta_i)(\boldsymbol{\alpha}_{2,0}^{\mathrm{T}}\boldsymbol{\xi}_i)\left\{\begin{array}{c}\boldsymbol{\xi}_i\\ \mathcal{S}(\boldsymbol{\theta}_0^{\mathrm{T}}\boldsymbol{Z}_{i1})\boldsymbol{\xi}_i\\ \boldsymbol{Z}_i\end{array}\right\}, \quad \boldsymbol{\pi}_{i2} = \rho_2(\eta_i)(\boldsymbol{\alpha}_{2,0}^{\mathrm{T}}\boldsymbol{\xi}_i)^2\boldsymbol{Z}_{i1},$$

$$\widehat{\boldsymbol{\pi}}_1(\mathbf{z};\boldsymbol{\theta}) = \mathrm{E}(\boldsymbol{\pi}_{i1}|\boldsymbol{\theta}^{\mathrm{T}}\boldsymbol{Z}_{i1} = \boldsymbol{\theta}^{\mathrm{T}}\mathbf{z}), \quad \widehat{\boldsymbol{\pi}}_2(\mathbf{z};\boldsymbol{\theta}) = \mathrm{E}(\boldsymbol{\pi}_{i2}|\boldsymbol{\theta}^{\mathrm{T}}\boldsymbol{Z}_{i1} = \boldsymbol{\theta}^{\mathrm{T}}\mathbf{z}),$$

$$\mathcal{V}_2(\mathbf{z};\boldsymbol{\theta}) = \left[\begin{array}{c}\widehat{\boldsymbol{\pi}}_1(\mathbf{z};\boldsymbol{\theta})\\ \mathcal{S}^{(1)}(\boldsymbol{\theta}^{\mathrm{T}}\mathbf{z})\{\widehat{\boldsymbol{\pi}}_2(\mathbf{z};\boldsymbol{\theta}) - \mathcal{V}_1(\mathbf{z};\boldsymbol{\theta})\mathbf{z}\}\end{array}\right].$$

By solving the local estimating equation in (B.8), we obtain the following results, uniformly for all $u = \boldsymbol{\theta}_0^{\mathrm{T}}\mathbf{z}$ with $\mathbf{z} \in \mathcal{D}$:

$$\begin{aligned}\widehat{a}_{0,\mathrm{curr}} - a_0 &= -(\mathcal{V}_1^{-1}\mathcal{V}_2^{\mathrm{T}})(\mathbf{z};\boldsymbol{\theta}_0) \times (\widehat{\boldsymbol{\Theta}}_{\mathrm{prev}} - \boldsymbol{\Theta}_0) + R_{0,n}(\mathbf{z}) \\ &\qquad + \frac{\sigma_K^2}{2}\mathcal{S}^{(2)}(\boldsymbol{\theta}_0^{\mathrm{T}}\mathbf{z})h^2 + O_p(h^3) + o_p\{\Delta(\widehat{\boldsymbol{\Theta}}_{\mathrm{prev}})\}, \\ \widehat{a}_{1,\mathrm{curr}}^* - a_1^* &= O_p\{(h+\delta_n) \times \Delta(\widehat{\boldsymbol{\Theta}}_{\mathrm{prev}})\} + O_p(h^3 + \delta_n),\end{aligned} \tag{B.10}$$

where, by letting $\epsilon_i = q_1\{\eta_i, Y_i\}$,

$$R_{0,n}(\mathbf{z}) = \mathcal{V}_1^{-1}(\mathbf{z};\widehat{\boldsymbol{\theta}}_{\mathrm{prev}})\textstyle\sum_{i=1}^n w_i(\widehat{\boldsymbol{\theta}}_{\mathrm{prev}}^{\mathrm{T}}\mathbf{z})(\boldsymbol{\alpha}_{2,0}^{\mathrm{T}}\boldsymbol{\xi}_i)\epsilon_i. \tag{B.11}$$

Again, the expressions in (B.10) allow us to account for the $\widehat{\boldsymbol{\Theta}}_{\mathrm{prev}}$ embedded with the estimated a's in equation (B.13) below. Next, we fix $\widehat{\mathcal{S}}(\cdot)$ and $\widehat{\mathcal{S}}^{(1)}(\cdot)$ and update $\widehat{\boldsymbol{\Theta}}$ by solving the estimating equation (B.12),

$$\begin{aligned}\mathbf{0} &= n^{-1}\textstyle\sum_{j=1}^n\sum_{i=1}^n w_{ij}q_1[\eta_i\{\widehat{\boldsymbol{\alpha}}_{\mathrm{curr}}, \widehat{\boldsymbol{\beta}}_{\mathrm{curr}}, \widehat{\boldsymbol{\theta}}_{\mathrm{curr}}, (\widehat{a}_{0j,\mathrm{curr}} + \widehat{a}_{1j,\mathrm{curr}}\widehat{\boldsymbol{\theta}}_{\mathrm{curr}}^{\mathrm{T}}\boldsymbol{Z}_{ij,1})\}, Y_i] \\ &\qquad \times \left\{\begin{array}{c}\boldsymbol{\xi}_i\\ (\widehat{a}_{0j,\mathrm{curr}} + \widehat{a}_{1j,\mathrm{curr}}\widehat{\boldsymbol{\theta}}_{\mathrm{curr}}^{\mathrm{T}}\boldsymbol{Z}_{ij,1})\boldsymbol{\xi}_i\\ \boldsymbol{Z}_i\\ (\widehat{\boldsymbol{\alpha}}_{2,\mathrm{curr}}^{\mathrm{T}}\boldsymbol{\xi}_i)\widehat{a}_{1j,\mathrm{curr}}\boldsymbol{Z}_{ij,1}\end{array}\right\}.\end{aligned} \tag{B.12}$$

For observations with $|\widehat{\boldsymbol{\theta}}_{\mathrm{curr}}^{\mathrm{T}}\boldsymbol{Z}_{ij,1}| \leq h$, we have

$$\begin{aligned}&\widehat{a}_{0j,\mathrm{curr}} + \widehat{a}_{1j,\mathrm{curr}}\widehat{\boldsymbol{\theta}}_{\mathrm{curr}}^{\mathrm{T}}\boldsymbol{Z}_{ij,1} - \mathcal{S}(\boldsymbol{\theta}_0\boldsymbol{Z}_{i1}) \\ &= \mathcal{S}^{(1)}(\boldsymbol{\theta}_0\boldsymbol{Z}_{j1})\boldsymbol{Z}_{ij,1}^{\mathrm{T}}(\widehat{\boldsymbol{\theta}}_{\mathrm{curr}} - \boldsymbol{\theta}_0) + \left(\begin{array}{c}1\\ \widehat{\boldsymbol{\theta}}_{\mathrm{curr}}^{\mathrm{T}}\boldsymbol{Z}_{ij,1}\end{array}\right)\left(\begin{array}{c}\widehat{a}_{0j,\mathrm{curr}} - a_{0j}\\ \widehat{a}_{1j,\mathrm{curr}} - a_{1j}\end{array}\right) \\ &\qquad -\frac{1}{2}\mathcal{S}^{(2)}(\boldsymbol{\theta}_0^{\mathrm{T}}\boldsymbol{Z}_{j1})(\widehat{\boldsymbol{\theta}}_{\mathrm{prev}}^{\mathrm{T}}\boldsymbol{Z}_{ij,1})^2 + O_p(h^3) + o_p(\|\widehat{\boldsymbol{\theta}}_{\mathrm{curr}} - \boldsymbol{\theta}_0\|).\end{aligned}$$

Define $\nu_{\mathbf{Z}_1}(\mathbf{z};\boldsymbol{\theta}) = \mathrm{E}(\boldsymbol{Z}_1|\boldsymbol{\theta}^{\mathrm{T}}\boldsymbol{Z}_1 = \boldsymbol{\theta}^{\mathrm{T}}\mathbf{z})$. The right hand side of equation (B.12) can be rewritten as

$$\begin{aligned}
\mathbf{0} &= n^{-1}\textstyle\sum_{i=1}^{n}\epsilon_i \begin{bmatrix} \boldsymbol{\xi}_i \\ \mathcal{S}(\boldsymbol{\theta}_0^{\mathrm{T}}\boldsymbol{Z}_{i1})\boldsymbol{\xi}_i \\ \boldsymbol{Z}_i \\ (\boldsymbol{\alpha}_{2,0}^{\mathrm{T}}\boldsymbol{\xi}_i)\mathcal{S}^{(1)}(\boldsymbol{\theta}_0^{\mathrm{T}}\boldsymbol{Z}_{i1})\{\boldsymbol{Z}_{i1} - \nu_{\mathbf{Z}_1}(\boldsymbol{Z}_{i1},\boldsymbol{\theta}_0)\} \end{bmatrix} \times \{1 + o_p(1)\} \\
&\quad + n^{-1}\textstyle\sum_{j=1}^{n}\Big\{ -\mathcal{V}_3(\boldsymbol{Z}_{j1};\boldsymbol{\theta}_0)(\widehat{\boldsymbol{\Theta}}_{\mathrm{curr}} - \boldsymbol{\Theta}_0) - \mathcal{V}_2(\boldsymbol{Z}_{j1};\boldsymbol{\theta}_0)(\widehat{a}_{0j,\mathrm{curr}} - a_{0j}) \\
&\quad + \frac{\sigma_K^2 h^2}{2}\mathcal{S}^{(2)}(\boldsymbol{\theta}_0^{\mathrm{T}}\boldsymbol{Z}_{j1})\mathcal{V}_2(\boldsymbol{Z}_{j1};\boldsymbol{\theta}_0)\Big\}\{1 + o_p(1)\} + o_p(n^{-1/2}) + o_p\{\Delta(\widehat{\boldsymbol{\Theta}}_{\mathrm{prev}})\}, \quad \text{(B.13)}
\end{aligned}$$

where $\mathcal{V}_3(\mathbf{z};\boldsymbol{\theta})$ is a symmetric matrix whose $(\ell_1,\ell_2)^{\mathrm{th}}$ block is denoted by $\mathcal{V}_3^{[\ell_1,\ell_2]}$ for $\ell_1,\ell_2 = 1, 2$, and

$$\begin{aligned}
\mathcal{V}_3^{[1,1]}(\mathbf{z};\boldsymbol{\theta}) &= \mathrm{E}\left[\rho_2(\eta_i)\begin{Bmatrix} \boldsymbol{\xi}_i \\ \mathcal{S}(\boldsymbol{\theta}_0^{\mathrm{T}}\boldsymbol{Z}_{i1})\boldsymbol{\xi}_i \\ \boldsymbol{Z}_i \end{Bmatrix}^{\otimes 2} \middle| \boldsymbol{\theta}^{\mathrm{T}}\boldsymbol{Z}_{i1} = \boldsymbol{\theta}^{\mathrm{T}}\mathbf{z}\right], \\
\mathcal{V}_3^{[1,2]}(\mathbf{z};\boldsymbol{\theta}) &= \mathcal{S}^{(1)}(\boldsymbol{\theta}_0^{\mathrm{T}}\mathbf{z})\{\mathrm{E}(\boldsymbol{\pi}_{i1}\boldsymbol{Z}_{i1}^{\mathrm{T}}|\boldsymbol{\theta}^{\mathrm{T}}\boldsymbol{Z}_{i1} = \boldsymbol{\theta}^{\mathrm{T}}\mathbf{z}) - \widehat{\boldsymbol{\pi}}_1(\mathbf{z};\boldsymbol{\theta})\mathbf{z}^{\mathrm{T}}\}, \\
\mathcal{V}_3^{[2,1]}(\mathbf{z};\boldsymbol{\theta}) &= \{\mathcal{V}_3^{[1,2]}(\mathbf{z};\boldsymbol{\theta})\}^{\mathrm{T}}, \\
\mathcal{V}_3^{[2,2]}(\mathbf{z};\boldsymbol{\theta}) &= \{\mathcal{S}^{(1)}(\boldsymbol{\theta}_0^{\mathrm{T}}\mathbf{z})\}^2\Big[\mathrm{E}\{\rho_2(\eta_i)(\boldsymbol{\alpha}_{2,0}^{\mathrm{T}}\boldsymbol{\xi}_i)^2\boldsymbol{Z}_{i1}\boldsymbol{Z}_{i1}^{\mathrm{T}}|\boldsymbol{\theta}^{\mathrm{T}}\boldsymbol{Z}_{i1} = \boldsymbol{\theta}^{\mathrm{T}}\mathbf{z}\} - \widehat{\boldsymbol{\pi}}_2(\mathbf{z};\boldsymbol{\theta})\mathbf{z}^{\mathrm{T}} \\
&\qquad - \mathbf{z}\widehat{\boldsymbol{\pi}}_2^{\mathrm{T}}(\mathbf{z};\boldsymbol{\theta}) + \mathcal{V}_1(\mathbf{z};\boldsymbol{\theta})\mathbf{z}\mathbf{z}^{\mathrm{T}}\Big].
\end{aligned}$$

By plugging (B.10) into (B.13) we have

$$\widehat{\boldsymbol{\Theta}}_{\mathrm{curr}} - \boldsymbol{\Theta}_0 = \mathcal{A}^{-}\mathcal{N}_n + \mathcal{A}^{-}\mathcal{C}(\widehat{\boldsymbol{\Theta}}_{\mathrm{prev}} - \boldsymbol{\Theta}_0) + o_p(n^{-1/2}) + o_p(\|\widehat{\boldsymbol{\Theta}}_{\mathrm{prev}} - \boldsymbol{\Theta}_0\|), \quad \text{(B.14)}$$

where

$$\begin{aligned}
\mathcal{N}_n &= n^{-1}\textstyle\sum_{i=1}^{n}\epsilon_i \begin{bmatrix} \boldsymbol{\xi}_i - \mathcal{V}_1^{-1}(\boldsymbol{Z}_{i1};\boldsymbol{\theta}_0)(\boldsymbol{\alpha}_{2,0}^{\mathrm{T}}\boldsymbol{\xi}_i)\mathrm{E}\{\rho_2(\eta)(\boldsymbol{\alpha}_{2,0}^{\mathrm{T}}\boldsymbol{\xi})\boldsymbol{\xi}|\boldsymbol{\theta}_0^{\mathrm{T}}\boldsymbol{Z}_1 = \boldsymbol{\theta}_0^{\mathrm{T}}\boldsymbol{Z}_{i1}\} \\ \mathcal{S}(\boldsymbol{\theta}_0^{\mathrm{T}}\boldsymbol{Z}_{i1})\boldsymbol{\xi}_i - \mathcal{V}_1^{-1}(\boldsymbol{Z}_{i1};\boldsymbol{\theta}_0)(\boldsymbol{\alpha}_{2,0}^{\mathrm{T}}\boldsymbol{\xi}_i)\mathrm{E}\{\rho_2(\eta)(\boldsymbol{\alpha}_{2,0}^{\mathrm{T}}\boldsymbol{\xi})\mathcal{S}(\boldsymbol{\theta}_0^{\mathrm{T}}\boldsymbol{Z}_1)\boldsymbol{\xi}|\boldsymbol{\theta}_0^{\mathrm{T}}\boldsymbol{Z}_1 = \boldsymbol{\theta}_0^{\mathrm{T}}\boldsymbol{Z}_{i1}\} \\ \boldsymbol{Z}_i - \mathcal{V}_1^{-1}(\boldsymbol{Z}_{i1};\boldsymbol{\theta}_0)(\boldsymbol{\alpha}_{2,0}^{\mathrm{T}}\boldsymbol{\xi}_i)\mathrm{E}\{\rho_2(\eta)(\boldsymbol{\alpha}_{2,0}^{\mathrm{T}}\boldsymbol{\xi})\boldsymbol{Z}|\boldsymbol{\theta}_0^{\mathrm{T}}\boldsymbol{Z}_1 = \boldsymbol{\theta}_0^{\mathrm{T}}\boldsymbol{Z}_{i1}\} \\ (\boldsymbol{\alpha}_{2,0}^{\mathrm{T}}\boldsymbol{\xi}_i)\mathcal{S}^{(1)}(\boldsymbol{\theta}_0^{\mathrm{T}}\boldsymbol{Z}_{i1})\{\boldsymbol{Z}_{i1} - \mathcal{V}_1^{-1}(\boldsymbol{Z}_{i1;\boldsymbol{\theta}_0})\widehat{\boldsymbol{\pi}}_2(\boldsymbol{Z}_{i1},\boldsymbol{\theta}_0)\} \end{bmatrix}, \\
\mathcal{C} &= \mathrm{E}\left[\mathcal{V}_1^{-1}\begin{Bmatrix} \widehat{\boldsymbol{\pi}}_1\widehat{\boldsymbol{\pi}}_1^{\mathrm{T}} & \mathcal{S}^{(1)}(\boldsymbol{\theta}_0\boldsymbol{Z}_1)\widehat{\boldsymbol{\pi}}_1(\widehat{\boldsymbol{\pi}}_2 - \mathcal{V}_1\boldsymbol{Z}_1)^{\mathrm{T}} \\ \mathcal{S}^{(1)}(\boldsymbol{\theta}_0\boldsymbol{Z}_1)(\widehat{\boldsymbol{\pi}}_2 - \mathcal{V}_1\boldsymbol{Z}_1)\widehat{\boldsymbol{\pi}}_1^{\mathrm{T}} & \{\mathcal{S}^{(1)}(\boldsymbol{\theta}_0\boldsymbol{Z}_1)\}^2(\widehat{\boldsymbol{\pi}}_2 - \mathcal{V}_1\boldsymbol{Z}_1)(\widehat{\boldsymbol{\pi}}_2 - \mathcal{V}_1\boldsymbol{Z}_1)^{\mathrm{T}} \end{Bmatrix}\right], \\
\mathcal{A} &= \mathrm{E}\{\mathcal{V}_3(\boldsymbol{Z}_1;\boldsymbol{\theta}_0)\}. \quad \text{(B.15)}
\end{aligned}$$

Define

$$\mathbf{g}_1 = \rho_2^{1/2}(\eta)\left[\boldsymbol{\xi}^{\mathrm{T}}, \{\mathcal{S}(\boldsymbol{\theta}_0^{\mathrm{T}}\boldsymbol{Z}_1)\boldsymbol{\xi}\}^{\mathrm{T}}, \boldsymbol{Z}^{\mathrm{T}}\right]^{\mathrm{T}}, \quad \mathbf{g}_2 = \mathcal{S}^{(1)}(\boldsymbol{\theta}_0^{\mathrm{T}}\boldsymbol{Z}_1)\rho_2^{1/2}(\eta)(\boldsymbol{\alpha}_{2,0}^{\mathrm{T}}\boldsymbol{\xi})\boldsymbol{Z}_1,$$
$$\mathbf{g} = (\mathbf{g}_1^{\mathrm{T}}, \mathbf{g}_2^{\mathrm{T}})^{\mathrm{T}}, \quad \boldsymbol{\varrho} = \{\widehat{\boldsymbol{\pi}}_1^{\mathrm{T}}(\boldsymbol{Z}_1;\boldsymbol{\theta}_0), \mathcal{S}^{(1)}(\boldsymbol{\theta}_0^{\mathrm{T}}\boldsymbol{Z}_1)\widehat{\boldsymbol{\pi}}_2^{\mathrm{T}}(\boldsymbol{Z}_1;\boldsymbol{\theta}_0)\}^{\mathrm{T}},$$

and let $\boldsymbol{D} = \mathcal{A} - \mathcal{C} = \mathrm{E}(\mathbf{g}\mathbf{g}^{\mathrm{T}}) - \mathrm{E}(\mathcal{V}_1^{-1}\boldsymbol{\varrho}\boldsymbol{\varrho}^{\mathrm{T}})$. It can be shown that both $\mathcal{A}$ and $\mathcal{C}$ are positive semi-definite matrices with rank $2p + 2d - 1$. As in the previous subsection, an asymptotic convergence property is achieved provided that the eigenvalues of $\mathcal{A}^{-}\mathcal{C}$ are strictly less than 1. This can be established by the Cauchy-Schwartz inequality so that $\boldsymbol{D}$ is positive semi-definite with the same rank as $\mathcal{A}$ and $\mathcal{C}$, and the eigenvalues of $\mathcal{A}^{-}\mathcal{C}$ are strictly less than 1. Consequently, at the end of iterations, $\widehat{\boldsymbol{\Theta}}_{\mathrm{curr}} - \boldsymbol{\Theta}_0 = \mathcal{A}^{-}\mathcal{N}_n + o_p(n^{-1/2})$. By the Central Limit Theorem, $\mathcal{N}_n$ is asymptotically normal. Defining

$$\mathcal{B} = n \times \mathrm{cov}(\mathcal{N}_n), \tag{B.16}$$

the asymptotic normality result of $\widehat{\boldsymbol{\Theta}}$ follows immediately. Finally, the asymptotic normal distribution of $\widehat{\mathcal{S}}(u)$ follows from (B.10).

B.3 Proof of Theorem 3

Proof of Lemma 1: The first result in Lemma 1 that $\|\widehat{\psi}_k - \psi_k\| = O_p\{h_\varpi^2 + (nh_\varpi)^{-1/2}\}$ was established by Hall, et al. (2006) for the case of fixed m. With the smoothing parameter h_ϖ in the range defined in (C4.4), we have $\|\widehat{\psi}_k - \psi_k\| = O_p(n^{-1/3})$. It is easily shown that

$$\widehat{\xi}_{ik} - \xi_{ik} = \zeta_{1,ik} + \zeta_{2,ik} + \zeta_{3,ik}, \tag{B.17}$$

where $\zeta_{1,ik} = \{(b-a)/m\}\sum_{j=1}^m X_i(t_{ij})\psi_k(t_{ij}) - \int_a^b X_i(t)\psi_k(t)dt$, $\zeta_{2,ik} = \{(b-a)/m\}\sum_{j=1}^m U_{ij}\psi_k(t_{ij})$, and $\zeta_{3,ik} = \{(b-a)/m\}\sum_{j=1}^m W_{ij}\{\widehat{\psi}_k(t_{ij}) - \psi_k(t_{ij})\}$. It can then be established that, with $X_i(t)$ and $\psi_k(t)$ being continuously differentiable, $\zeta_{1,ik}$, representing the numerical integration error, is of order $O_p(m^{-1})$, that $\zeta_{2,ik} = O_p(m^{-1/2})$. Lemma 1 now follows because $\zeta_{3,ik} = O_p(n^{-1/3})$ by the Cauchy-Schwartz inequality, since

$$\begin{aligned} |\zeta_{3,ik}| &\le \left(\frac{b-a}{m}\sum_{j=1}^m W_{ij}^2\right)^{1/2}\left[\frac{b-a}{m}\sum_{j=1}^m\{\widehat{\psi}_k(t_{ij}) - \psi_k(t_{ij})\}^2\right]^{1/2} \\ &= O_p(\|\widehat{\psi}_k - \psi_k\|) = O_p(n^{-1/3}). \end{aligned}$$

Proof of Theorem 3: The proof of Theorem 3 essentially follows the same lines as those for Theorem 1 but with extra effort on keeping track of additional terms due to $\widehat{\boldsymbol{\xi}}_i - \boldsymbol{\xi}_i$, as given by equation (B.17). It can be shown that, for equations (B.2), (B.3) and (B.6), the effect due to $\boldsymbol{\xi}_i$ being estimated is an additional term of order $O_p(m^{-1} + n^{-1/3})$ on the right hand side of the equations. For equation (B.7), this results in an additional term of order $O_p(b^{-1}m^{-1} + b^{-1}n^{-1/3})$. Under the bandwidth assumption (C4.4), the estimator $\widetilde{\boldsymbol{\Theta}}$ remains consistent.

B.4 Proof of Theorem 4

The proof of Theorem 4 follows the same lines as those for Theorem 2, except that the term ϵ_i in equations (B.11)-(B.13) should be replaced by

$$\begin{aligned}\widetilde{\epsilon}_i &= q_1\{\boldsymbol{\alpha}_{1,0}^{\mathrm{T}}\widehat{\boldsymbol{\xi}}_i + \mathcal{S}_0(\boldsymbol{\theta}_0^{\mathrm{T}}\boldsymbol{Z}_{i1})\boldsymbol{\alpha}_{2,0}^{\mathrm{T}}\widehat{\boldsymbol{\xi}}_i + \boldsymbol{\beta}_0^{\mathrm{T}}\boldsymbol{Z}_i, Y_i\} \\ &= \epsilon_i + q_2(\eta_i, Y_i)\{\boldsymbol{\alpha}_{1,0} + \mathcal{S}_0(\boldsymbol{\theta}_0^{\mathrm{T}}\boldsymbol{Z}_{i1})\boldsymbol{\alpha}_{2,0}\}^{\mathrm{T}}(\widehat{\boldsymbol{\xi}}_i - \boldsymbol{\xi}_i).\end{aligned}$$

Let $\mathcal{A}$ and $\mathcal{C}$ be as defined in the Proof of Theorem 2. Equation (B.14) becomes

$$\widehat{\boldsymbol{\Theta}}_{\text{curr}} - \boldsymbol{\Theta}_0 = \mathcal{A}^{-}\widetilde{\mathcal{N}}_n + \mathcal{A}^{-}\mathcal{C}(\widehat{\boldsymbol{\Theta}}_{\text{prev}} - \boldsymbol{\Theta}_0) + o_p(n^{-1/2}) + o_p(\|\widehat{\boldsymbol{\Theta}}_{\text{prev}} - \boldsymbol{\Theta}_0\|),$$

where $\widetilde{\mathcal{N}}_n = \mathcal{N}_n + \mathcal{N}_{1,n}$, $\mathcal{N}_{1,n} = n^{-1}\sum_{i=1}^n \mathcal{E}_i(\widehat{\boldsymbol{\xi}}_i - \boldsymbol{\xi}_i)$ and

$$\mathcal{E}_i = \begin{bmatrix} \boldsymbol{\xi}_i - \mathcal{V}_1^{-1}(\boldsymbol{Z}_{i1};\boldsymbol{\theta}_0)(\boldsymbol{\alpha}_{2,0}^{\mathrm{T}}\boldsymbol{\xi}_i)\mathrm{E}\{\rho_2(\eta)(\boldsymbol{\alpha}_{2,0}^{\mathrm{T}}\boldsymbol{\xi})\boldsymbol{\xi}|\boldsymbol{\theta}_0^{\mathrm{T}}\boldsymbol{Z}_1 = \boldsymbol{\theta}_0^{\mathrm{T}}\boldsymbol{Z}_{i1}\} \\ \mathcal{S}(\boldsymbol{\theta}_0^{\mathrm{T}}\boldsymbol{Z}_{i1})\boldsymbol{\xi}_i - \mathcal{V}_1^{-1}(\boldsymbol{Z}_{i1};\boldsymbol{\theta}_0)(\boldsymbol{\alpha}_{2,0}^{\mathrm{T}}\boldsymbol{\xi}_i)\mathrm{E}\{\rho_2(\eta)(\boldsymbol{\alpha}_{2,0}^{\mathrm{T}}\boldsymbol{\xi})\mathcal{S}(\boldsymbol{\theta}_0^{\mathrm{T}}\boldsymbol{Z}_1)\boldsymbol{\xi}|\boldsymbol{\theta}_0^{\mathrm{T}}\boldsymbol{Z}_1 = \boldsymbol{\theta}_0^{\mathrm{T}}\boldsymbol{Z}_{i1}\} \\ \boldsymbol{Z}_i - \mathcal{V}_1^{-1}(\boldsymbol{Z}_{i1};\boldsymbol{\theta}_0)(\boldsymbol{\alpha}_{2,0}^{\mathrm{T}}\boldsymbol{\xi}_i)\mathrm{E}\{\rho_2(\eta)(\boldsymbol{\alpha}_{2,0}^{\mathrm{T}}\boldsymbol{\xi})\boldsymbol{Z}|\boldsymbol{\theta}_0^{\mathrm{T}}\boldsymbol{Z}_1 = \boldsymbol{\theta}_0^{\mathrm{T}}\boldsymbol{Z}_{i1}\} \\ (\boldsymbol{\alpha}_{2,0}^{\mathrm{T}}\boldsymbol{\xi}_i)\mathcal{S}^{(1)}(\boldsymbol{\theta}_0^{\mathrm{T}}\boldsymbol{Z}_{i1})\{\boldsymbol{Z}_{i1} - \mathcal{V}_1^{-1}(\boldsymbol{Z}_{i1;\boldsymbol{\theta}_0})\widehat{\boldsymbol{\pi}}_2(\boldsymbol{Z}_{i1},\boldsymbol{\theta}_0)\} \end{bmatrix}$$
$$\times q_2(\eta_i, Y_i)\{\boldsymbol{\alpha}_{1,0} + \mathcal{S}_0(\boldsymbol{\theta}_0^{\mathrm{T}}\boldsymbol{Z}_{i1})\boldsymbol{\alpha}_{2,0}\}^{\mathrm{T}}.$$

Define $\Xi(t) = \mathrm{E}\{X_i(t)\mathcal{E}_i\}$, which is a $(2p+d+d_1)\times p$ matrix of functions. Denote $\boldsymbol{\psi}(t) = (\psi_1, \cdots, \psi_p)^{\mathrm{T}}(t)$, then $\mathcal{N}_{1,n} = \langle \Xi, \widehat{\boldsymbol{\psi}} - \boldsymbol{\psi}\rangle_{L^2} + o_p(n^{-1/2})$. By Lemma 2, the j^{th} element in $\mathcal{N}_{1,n}$ is

$$\begin{aligned}\mathcal{N}_{1,n}^{[j]} &= \sum_{\ell=1}^p \langle \Xi^{[j,\ell]}(t), \widehat{\psi}_\ell(t) - \psi_\ell(t)\rangle_{L^2} + o_p(n^{-1/2}) \\ &= n^{-1/2}\sum_{\ell=1}^p\sum_{k\neq\ell}\frac{\langle \Xi^{[j,\ell]}, \psi_k\rangle_{L^2}}{\omega_\ell - \omega_k}\int \mathcal{Z}(s,t)\psi_\ell(s)\psi_k(t)dsdt + o_p(n^{-1/2}).\end{aligned}$$

Since $\mathcal{Z}$ is a Gaussian random field, $n^{1/2}\mathcal{N}_{1,n}$ converges to a Gaussian random vector. It can be further verified that $\mathcal{N}_n$ and $\mathcal{N}_{1,n}$ are independent. Define

$$\mathcal{B}_1 = \lim_{n\to\infty} n \times \mathrm{cov}(\mathcal{N}_{1,n}). \tag{B.18}$$

The asymptotic normality of $\widehat{\boldsymbol{\Theta}}$ now follows from similar arguments to those for Theorem 2.