Supplementary Materials to

Semiparametric Analysis of Heterogeneous Data Using Varying-Scale Generalized Linear Models

Minge Xie, Douglas G. Simpson, and Raymond J. Carroll

Rutgers University, University of Illinois and Texas A&M University

The supplementary materials are a full Appendix of the paper. It includes two parts: Appendix A contains an additional theorem on asymptotic expansions of $\widehat{\lambda}_\theta(z)$ and $\widehat{\lambda}_\theta^{(1)}(z)$, two related corollaries, and their proofs. Appendix B contains proofs of the three theorems in the paper.

APPENDIX A: An Additional Theorem

In Appendix A.1, we provide an additional theorem on asymptotic expansions of the local maximum likelihood estimator $\widehat{\lambda}_\theta(z)$, as well as its derivative with respect to $\boldsymbol{\theta}$, $\widehat{\lambda}_\theta^{(1)}(z)$. The asymptotic expansions of $\widehat{w}_\theta(z)$ and $\widehat{w}_\theta^{(1)}(z)$, and their uniform bounds are provided in Corollaries A1 and A2, respectively. Proofs of this theorem and the two corollaries are outlined in Appendix A.2.

A.1 Asymptotic Expansions of $\widehat{\lambda}_\theta$, $\widehat{\lambda}_\theta^{(1)}$ And $\widehat{w}_\theta$, $\widehat{w}_\theta^{(1)}$.

The following theorem holds under some mild conditions.

Theorem A. *Let $K(t)$ be a symmetric kernel function. Suppose $b = O(n^{-\xi})$, $1/6 < \xi < 1/4$ and H is invertible. For any given $\boldsymbol{\beta} \in B_n(r)$, $\boldsymbol{\delta} \in \tilde{B}_n(\tilde{r})$ and z_0, we have the following asymptotic expansions:*

$$\begin{aligned}\mathbf{J}(\widehat{\lambda}_\theta - \lambda) \;=\; & \{\mathbf{H}^{-1} + o(1)\}\Big[\frac{1}{n}\sum_{j=1}^{n}\mathbf{J}^{-1}\mathbf{z}_{j,0}\{y_j - \mu(w_j^{(0)}\eta_j^{(0)} + \tilde{\eta}_j^{(0)})\}\eta_j^{(0)}\tau(w_j^{(0)}\eta_j^{(0)} + \tilde{\eta}_j^{(0)})K_b(z_j - z_0) \\ & + \frac{b^2}{2}w^{(2)}(z_0)f_z(z_0)\gamma(z_0)\begin{pmatrix}\upsilon_2\\0\end{pmatrix} + \begin{pmatrix}1\\0\end{pmatrix}w^{(0)}(z_0)f_z(z_0)\{\gamma_1(z_0)\}^T(\boldsymbol{\beta} - \boldsymbol{\beta}_0) \\ & + \begin{pmatrix}1\\0\end{pmatrix}f_z(z_0)\{\tilde{\gamma}_1(z_0)\}^T(\boldsymbol{\delta} - \boldsymbol{\delta}_0) + o_p(\frac{1}{\sqrt{n}})\Big]\end{aligned}$$

and

$$\begin{aligned}\mathbf{J}\widehat{\lambda}_\theta^{(1)} \;=\; & \{\mathbf{H}^{-1} + o(1)\}\Big[-\frac{1}{n}\sum_{j=1}^{n}\mathbf{J}^{-1}\mathbf{z}_{j,0}\begin{pmatrix}w_j^{(0)}\mathbf{x}_j^T\\ \mathbf{v}_j\end{pmatrix}\eta_j^{(0)}\tau_1(w_j^{(0)}\eta_j^{(0)} + \tilde{\eta}_j^{(0)})K_b(z_j - z_0) \\ & + \frac{1}{n}\sum_{j=1}^{n}\mathbf{J}^{-1}\mathbf{z}_{j,0}\begin{pmatrix}\{w_j^{(0)}\eta_j^{(0)}\tau'(w_j^{(0)}\eta_j^{(0)} + \tilde{\eta}_j^{(0)}) + \tau(w_j^{(0)}\eta_j^{(0)} + \tilde{\eta}_j^{(0)})\}\mathbf{x}_j^T\\ \eta_j^{(0)}\tau'(w_j^{(0)}\eta_j^{(0)} + \tilde{\eta}_j^{(0)})\mathbf{v}_j^T\end{pmatrix}\{y_j \\ & - \mu(w_j^{(0)}\eta_j^{(0)} + \tilde{\eta}_j^{(0)})\}K_b(z_j - z_0) - \frac{b^2}{2}w^{(2)}(z_0)f_z(z_0)\begin{pmatrix}\upsilon_2\\0\end{pmatrix}\{\boldsymbol{\Gamma}_R(z_0)\,\boldsymbol{\beta}_0\}^T\end{aligned}$$

$$-w^{(0)}(z_0)f_z(z_0)\begin{pmatrix}1\\0\end{pmatrix}\{\boldsymbol{\Gamma}(z_0)(\boldsymbol{\beta}-\boldsymbol{\beta}_0)\}^T - f_z(z_0)\begin{pmatrix}1\\0\end{pmatrix}\{\tilde{\boldsymbol{\Gamma}}(z_0)(\boldsymbol{\delta}-\boldsymbol{\delta}_0)\}^T + o_p(\frac{1}{\sqrt{n}})\Big],$$

where $\mathbf{z}_{j,0} = \begin{pmatrix}1\\ z_j - z_0\end{pmatrix}$, *and the* $(p+q)\times p$ *matrices* $\Gamma_R(t)$, $\Gamma(t)$ *and* $\tilde{\Gamma}(t)$ *are defined in the proof of Lemma A1 in Appendix A.2.*

The following corollary suggests that $\widehat{w}_\theta(z_i)$ is a consistent estimator of $w^{(0)}(z_i)$ and $-\widehat{w}_\theta^{(1)}(z_i)$ is a consistent estimator of $\begin{pmatrix}\mathbf{m}(z_i)\\ \tilde{\mathbf{m}}(z_i)\end{pmatrix}$.

Corollary A1. *If* $K(t)$ *is a symmetric kernel function, we have*

$$\begin{aligned}\widehat{w}_\theta(z_i) &= w^{(0)}(z_i) + \frac{1}{n}\sum_{j=1}^n \{y_j - \mu(w_j^{(0)}\eta_j^{(0)} + \tilde{\eta}_j^{(0)})\}q_{j,i}K_b(z_j - z_i) + c_{i,1}b^2 + \mathbf{c}_{i,2}(\boldsymbol{\beta}-\boldsymbol{\beta}_0)\\ &\quad + \tilde{\mathbf{c}}_{i,2}(\boldsymbol{\delta}-\boldsymbol{\delta}_0) + o_p\Big(\frac{1}{\sqrt{n}}\Big). \qquad (A.1)\end{aligned}$$

where $q_{j,i}$ *depends on* $(\mathbf{x}_i, \mathbf{v}_i, z_i)$ *and* $(\mathbf{x}_j, \mathbf{v}_j, z_j)$ *but not on either* y_i *or* y_j, *and* $c_{i,1}$, $p\times 1$ *vector* $\mathbf{c}_{i,2}$ *and* $q\times 1$ *vector* $\tilde{\mathbf{c}}_{i,2}$ *depend only on* z_i*;*

$$\begin{aligned}\widehat{w}_\theta^{(1)}(z_i) &= -\begin{pmatrix}\mathbf{m}(z_i)\\ \tilde{\mathbf{m}}(z_i)\end{pmatrix} + \frac{1}{n}\sum_{j=1}^n \{y_j - \mu(w_j^{(0)}\eta_j^{(0)} + \tilde{\eta}_j^{(0)})\}\mathbf{q}_{j,i}K_b(z_j - z_i) + \mathbf{c}_{i,1}b^2\\ &\quad + \mathbf{C}_{i,2}(\boldsymbol{\beta}-\boldsymbol{\beta}_0) + \tilde{\mathbf{C}}_{i,2}(\boldsymbol{\delta}-\boldsymbol{\delta}_0) + \mathbf{c}_{i,3}(nb)^{-\frac{1}{2}} + o_p\Big(\frac{1}{\sqrt{n}}\Big). \qquad (A.2)\end{aligned}$$

where $(p+q)\times 1$ *vector* $\mathbf{q}_{j,i}$ *depend on* $(\mathbf{x}_i, \mathbf{v}_i, z_i)$ *and* $(\mathbf{x}_j, \mathbf{v}_j, z_j)$ *but not on either* y_i *or* y_j, *and* $(p+q)\times 1$ *vectors* $\mathbf{c}_{i,1}$, $\mathbf{c}_{i,3}$, $(p+q)\times p$ *matrix* $\mathbf{C}_{i,2}$ *and* $(p+q)\times q$ *matrix* $\tilde{\mathbf{C}}_{i,2}$ *depend only on* z_i.

Under some mild conditions, we provide in the next corollary uniform bounds of $\widehat{w}_\theta(z) - w^{(0)}(z)$ and $\widehat{w}_\theta^{(1)}(z) + [\{\mathbf{m}(z)\}^T, \{\tilde{\mathbf{m}}(z)\}^T]^T$.

Corollary A2. Suppose $\widehat{w}_\theta(z)$ and $\widehat{w}_\theta^{(1)}(z)$ have asymptotic expansions of forms (A.1) and (A.2). For some small enough ζ, $0<\zeta<\xi/2$, we have

$$\sup{}^*|\widehat{w}_\theta(z) - w^{(0)}(z)| = o_p(n^{-1/2+\zeta}b^{-1/2}) \quad \text{and} \quad \sup{}^*|\widehat{w}_\theta^{(1)}(z) + \begin{pmatrix}\mathbf{m}(z)\\ \tilde{\mathbf{m}}(z)\end{pmatrix}| = o_p(n^{-1/2+\zeta}b^{-1/2}),$$

where $\sup^*$ is the supremum over $\beta\in B_n(r)$, $\delta\in\tilde{B}(\tilde{r})$, $y\in\mathcal{Y}$, $\mathbf{x}\in\mathcal{X}$, $\mathbf{v}\in\mathcal{V}$ and $z\in\mathcal{Z}$. Here $\mathcal{Y}$, $\mathcal{X}$, $\mathcal{V}$ and $\mathcal{Z}$ are admissible sets of response variable y, covariate variables $\mathbf{x}$, $\mathbf{v}$ and z respectively.

A.2 Proofs of Theorem A, Corollaries A1 and A2.

Denote by $\mathbf{u} = (u_1, u_2, \ldots, u_n)^T$, $\eta = (\eta_1, \eta_2, \ldots, \eta_n)^T$ and $\tilde{\eta} = (\tilde{\eta}_1, \tilde{\eta}_2, \ldots, \tilde{\eta}_n)^T$, where $u_j = \lambda^T\mathbf{z}_{j0} = \lambda_0 + \lambda_1(z_j - z_0)$, $\tilde{\eta}_j = \mathbf{v}_j^T\delta$, and without loss of generality $\eta_j = 1 + \mathbf{x}_j^T\beta$. For convenience, following Cai, Fan and Li (2000), we reparameterize λ to $\lambda^* = (\lambda_0, b\lambda_1)^T$. Write $\mathbf{z}_{j0}^* = (1, t_j)^T$, $t_j = (z_j - z_0)/b$. Thus, for a given η and $\tilde{\eta}$, we can write the local likelihood function (6) in terms of λ^* as $\ell_{LO}^*(\lambda^*|\eta,\tilde{\eta}) \overset{\triangle}{=} \ell_{LO}(\lambda|\eta,\tilde{\eta}) = \frac{1}{n}\sum_{j=1}^n \ell\{\mu(u_j\eta_j + \tilde{\eta}_j), y_j\}K_b(z_j - z_0)$, where $u_j = \lambda^{*T}\mathbf{z}_{j0}^* = \lambda^T\mathbf{z}_{j0}$.

Maximizing $\ell_{LO}(\lambda|\eta,\tilde{\eta})$, with respect to λ, is the equivalent of maximizing $\ell^*_{LO}(\lambda^*|\eta,\tilde{\eta})$ with respect to λ^*. We have a score function,

$$\mathbf{S}^*_n(\mathbf{u},\eta,\tilde{\eta}) = \frac{\partial}{\partial\lambda^*}\ell^*_{LO}(\lambda^*|\eta,\tilde{\eta}) = \frac{1}{n}\sum_{j=1}^n \mathbf{z}^*_{j0}\{y_j-\mu(u_j\eta_j+\tilde{\eta}_j)\}\eta_j\tau(u_j\eta_j+\tilde{\eta}_j)K_b(z_j-z_0).$$

The corresponding negative hessian matrix is

$$\begin{aligned}\mathbf{H}^*_n(\mathbf{u},\eta,\tilde{\eta}) &= -\frac{\partial}{\partial\lambda^{*T}}\mathbf{S}^*_n(\mathbf{u},\eta,\tilde{\eta}) = -\frac{\partial^2}{\partial\lambda^*\partial\lambda^{*T}}\ell^*_{LO}(\lambda^*) = \mathbf{H}^*_{n,1}(\mathbf{u},\eta,\tilde{\eta})+\mathbf{H}^*_{n,2}(\mathbf{u},\eta,\tilde{\eta})\\ &= \frac{1}{n}\sum_{j=1}^n \mathbf{z}^*_{j0}\mathbf{z}^{*\,T}_{j0}\eta_j^2\tau_1(u_j\eta_j+\tilde{\eta}_j)K_b(z_j-z_0)\\ &\quad -\frac{1}{n}\sum_{j=1}^n \mathbf{z}^*_j\mathbf{z}^{*T}_j\eta_j^2\tau'(u_j\eta_j+\tilde{\eta}_j)K_b(z_j-z_0)\{y_j-\mu(u_j\eta_j+\tilde{\eta}_j)\}.\end{aligned}$$

Note that $\tau_1(s)=\mu'(s)\tau(s)$. Recall that $\boldsymbol{\theta}=(\boldsymbol{\beta}^T,\boldsymbol{\delta}^T)^T$. We also denote

$$\begin{aligned}\mathbf{R}^*_n(\mathbf{u},\eta,\tilde{\eta}) &= \frac{\partial}{\partial\theta^T}\mathbf{S}^*_n(\mathbf{u},\eta,\tilde{\eta}) = \mathbf{R}^*_{n,1}(\mathbf{u},\eta,\tilde{\eta})+\mathbf{R}^*_{n,2}(\mathbf{u},\eta,\tilde{\eta})\\ &= -\frac{1}{n}\sum_{j=1}^n \mathbf{z}^*_{j0}\begin{pmatrix}u_j\mathbf{x}_j\\ \mathbf{v}_j\end{pmatrix}^T \eta_j\tau_1(u_j\eta_j+\tilde{\eta}_j)K_b(z_j-z_0)\\ &\quad +\frac{1}{n}\sum_{j=1}^n \mathbf{z}^*_{j0}\begin{pmatrix}\{u_j\eta_j\tau'(u_j\eta_j+\tilde{\eta}_j)+\tau(u_j\eta_j+\tilde{\eta}_j)\}\mathbf{x}_j\\ \eta_j\tau'(u_j\eta_j+\tilde{\eta}_j)\mathbf{v}_j\end{pmatrix}^T K_b(z_j-z_0)\{y_j-\mu(u_j\eta_j+\tilde{\eta}_j)\}.\end{aligned}$$

Under some mild smooth conditions on the inverse link function $\mu(\cdot)$, and functions $a(\cdot)$, $f_z(\cdot)$, $\gamma(\cdot)$, $\gamma_1(\cdot)$, $\tilde{\gamma}_1(\cdot)$, $\boldsymbol{\Gamma}_R(\cdot)$, $\boldsymbol{\Gamma}(\cdot)$ and $\tilde{\boldsymbol{\Gamma}}(\cdot)$. We have the following lemma for fixed $\boldsymbol{\beta}$ and $\boldsymbol{\delta}$.

Lemma A1. *Suppose $\beta\in B_n(r)$ and $\delta\in\tilde{B}_n(\tilde{r})$. Under the condition of Theorem A, it follows that*

$$\begin{aligned}\mathbf{S}^*_n(\mathbf{u},\eta,\tilde{\eta}) &= \mathbf{S}^*_n(\mathbf{w}^{(0)},\eta^{(0)},\tilde{\eta}^{(0)}) + w^{(0)}(z_0)f_z(z_0)\{\gamma_1(z_0)\}^T(\boldsymbol{\beta}-\boldsymbol{\beta}_0)\begin{pmatrix}1\\0\end{pmatrix}\\ &\quad +f_z(z_0)\{\tilde{\gamma}_1(z_0)\}^T(\boldsymbol{\delta}-\boldsymbol{\delta}_0)\begin{pmatrix}1\\0\end{pmatrix}+\frac{b^2}{2}w^{(2)}(z_0)f_z(z_0)\gamma(z_0)\begin{pmatrix}\upsilon_2\\0\end{pmatrix}+o_p\Big(\frac{1}{\sqrt{n}}\Big),\\ \mathbf{R}^*_n(\mathbf{u},\eta,\tilde{\eta}) &= \mathbf{R}^*_n(\mathbf{w}^{(0)},\eta^{(0)},\tilde{\eta}^{(0)}) - w^{(0)}(z_0)f_z(z_0)\begin{pmatrix}1\\0\end{pmatrix}\{\boldsymbol{\Gamma}(z_0)(\boldsymbol{\beta}-\boldsymbol{\beta}_0)\}^T\\ &\quad -f_z(z_0)\begin{pmatrix}1\\0\end{pmatrix}\{\tilde{\boldsymbol{\Gamma}}(z_0)(\boldsymbol{\delta}-\boldsymbol{\delta}_0)\}^T-\frac{b^2}{2}w^{(2)}(z_0)f_z(z_0)\begin{pmatrix}\upsilon_2\\0\end{pmatrix}\{\boldsymbol{\Gamma}_R(z_0)\,\boldsymbol{\beta}_0\}^T+o_p\Big(\frac{1}{\sqrt{n}}\Big),\end{aligned}$$

and

$$\mathbf{H}^*_n(\mathbf{u},\eta) = \mathbf{H}+o_p(1).$$

Here $\mathbf{w}^{(0)}=(w_1^{(0)},\ldots,w_n^{(0)})^T$, $w_j^{(0)}=w^{(0)}(z_j)$, and $\eta^{(0)}=(\eta_1^{(0)},\ldots,\eta_n^{(0)})^T$.

Proof of Lemma A1: Write $\mathbf{S}^*_n(\mathbf{u},\eta,\tilde{\eta})-\mathbf{S}^*_n(\mathbf{w}^{(0)},\eta^{(0)},\tilde{\eta}^{(0)})=(I_{s,1})+(I_{s,2})+(I_{s,3})$, where

$$(I_{s,1}) = \frac{1}{n}\sum_{j=1}^n \mathbf{z}^*_{j0}\{y_j-\mu(w_j^{(0)}\eta_j^{(0)}+\tilde{\eta}_j^{(0)})\}\{\eta_j\tau(u_j\eta_j+\tilde{\eta}_j)-\eta_j^{(0)}\tau(w_j^{(0)}\eta_j^{(0)}+\tilde{\eta}_j^{(0)})\}K_b(z_j-z_0),$$

$$
\begin{aligned}
(I_{s,2}) &= \frac{1}{n}\sum_{j=1}^{n} \mathbf{z}_{j0}^{*}\{\mu(w_j^{(0)}\eta_j^{(0)}+\tilde{\eta}_j^{(0)})-\mu(u_j\eta_j+\tilde{\eta}_j)\}\eta_j^{(0)}\tau(w_j^{(0)}\eta_j^{(0)}+\tilde{\eta}_j^{(0)})K_b(z_j-z_0),\\
(I_{s,3}) &= \frac{1}{n}\sum_{j=1}^{n} \mathbf{z}_{j0}^{*}\{\mu(w_j^{(0)}\eta_j^{(0)}+\tilde{\eta}_j^{(0)})-\mu(u_j\eta_j+\tilde{\eta}_j)\}\{\eta_j\tau(u_j\eta_j+\tilde{\eta}_j)\\
&\quad -\eta_j^{(0)}\tau(w_j^{(0)}\eta_j^{(0)}+\tilde{\eta}_j^{(0)})\}K_b(z_j-z_0).
\end{aligned}
$$

Note that $\eta_j-\eta_j^{(0)}=\mathbf{x}_j^T(\boldsymbol{\beta}-\boldsymbol{\beta}^{(0)})$, $\tilde{\eta}_j-\tilde{\eta}_j^{(0)}=\mathbf{v}_j^T(\boldsymbol{\delta}-\boldsymbol{\delta}^{(0)})$ and $u_j-w_j^{(0)}=\frac{1}{2}w^{(2)}(z_0)t_j^2b^2+O_p(|t_j|^3b^3)$. We can express $\eta_j\tau(u_j\eta_j+\tilde{\eta}_j)-\eta_j^{(0)}\tau(w_j^{(0)}\eta_j^{(0)}+\tilde{\eta}_j)$ as $T_{j,1}\mathbf{x}_j^T(\boldsymbol{\beta}-\boldsymbol{\beta}_0)+\tilde{T}_{j,1}\mathbf{v}_j^T(\boldsymbol{\delta}-\boldsymbol{\delta}_0)+T_{j,2}t_j^2b^2+O_p(|t_j|^3b^3)+o_p(n^{-1/2})$ for some functions $T_{j,1}$, $\tilde{T}_{j,1}$ and $T_{j,2}$ which only depend on z_0, $\eta_j^{(0)}$, $\tilde{\eta}_j^{(0)}$ and the form of $w^{(0)}$. Since $O_p(b^3)=o_p(n^{-1/2})$, it follows that $(I_{s,1})=(I_{s,1a})+(\tilde{I}_{s,1a})+(I_{s,1b})+o_p(n^{-1/2})$, where

$$
\begin{aligned}
(I_{s,1a}) &= \left[\frac{1}{nb}\sum_{j=1}^{n} T_{j,1}\mathbf{z}_{j0}^{*}\mathbf{x}_j^T\{y_j-\mu(w_j^{(0)}\eta_j^{(0)}+\tilde{\eta}_j^{(0)})\}K(t_j)\right](\boldsymbol{\beta}-\boldsymbol{\beta}_0),\\
(\tilde{I}_{s,1a}) &= \left[\frac{1}{nb}\sum_{j=1}^{n} \tilde{T}_{j,1}\mathbf{z}_{j0}^{*}\mathbf{v}_j^T\{y_j-\mu(w_j^{(0)}\eta_j^{(0)}+\tilde{\eta}_j^{(0)})\}K(t_j)\right](\boldsymbol{\delta}-\boldsymbol{\delta}_0),\\
(I_{s,1b}) &= \frac{1}{nb}\sum_{j=1}^{n} T_{j,2}\mathbf{z}_{j0}^{*}t_j^2b^2\{y_j-\mu(w_j^{(0)}\eta_j^{(0)}+\tilde{\eta}_j^{(0)})\}K(t_j).
\end{aligned}
$$

From the standard derivation, we can prove that $E|I_{s,1a}|^2=o_p(n^{-1})$, $E|\tilde{I}_{s,1a}|^2=o_p(n^{-1})$ and $E|I_{s,1b}|=o_p(n^{-1})$. Thus, $(I_{s,1})=o_p(n^{-1/2})$.

Now, by the Taylor expansion, $\mu(w_j^{(0)}\eta_j^{(0)}+\tilde{\eta}_j^{(0)})-\mu(u_j\eta_j+\tilde{\eta}_j)=\mu'(w_j^{(0)}\eta_j^{(0)}+\tilde{\eta}_j^{(0)})\ \{w_j^{(0)}\mathbf{x}_j^T(\beta-\beta_0)+\mathbf{v}_j^T(\delta-\delta_0)+\frac{1}{2}\eta_j^{(0)}w^{(2)}(z_j)t_j^2b^2\}\ +O_p(|t_j|^3b^3)+o_p(n^{-1/2})$. We have

$$
\begin{aligned}
(I_{s,2}) &= \frac{1}{nb}\sum_{j=1}^{n} \mathbf{z}_{j0}^{*}\tau_1(w_j^{(0)}\eta_j^{(0)}+\tilde{\eta}_j^{(0)})\eta_j^{(0)}\{w^{(0)}(z_j)\mathbf{x}_j^T(\boldsymbol{\beta}-\boldsymbol{\beta}_0)+\mathbf{v}_j^T(\boldsymbol{\delta}-\boldsymbol{\delta}_0)\}K(t_j)\\
&\quad +\frac{1}{2nb}\sum_{j=1}^{n} \mathbf{z}_{j0}^{*}w^{(2)}(z_j)\tau_1(w_j^{(0)}\eta_j^{(0)}+\tilde{\eta}_j^{(0)})(\eta_j^{(0)})^2t_j^2b^2K(t_j)+o_p(n^{-1/2})\\
&= \int\begin{pmatrix}1\\ t\end{pmatrix}K(t)\Big[w^{(0)}(z_0+bt)\{\gamma_1(z_0+bt)\}^T(\boldsymbol{\beta}-\boldsymbol{\beta}_0)+\{\tilde{\gamma}_1(z_0+bt)\}^T(\boldsymbol{\delta}-\boldsymbol{\delta}_0)\Big]f_z(z_0+bt)dt\\
&\quad +\frac{b^2}{2}\int\begin{pmatrix}1\\ t\end{pmatrix}t^2\gamma(z_0+bt)w^{(2)}(z_0+bt)K(t)f_z(z_0+bt)dt+o_p(n^{-\frac{1}{2}})\\
&= f_z(z_0)\Big[w^{(0)}(z_0)\{\gamma_1(z_0)\}^T(\boldsymbol{\beta}-\boldsymbol{\beta}_0)+\{\tilde{\gamma}_1(z_0)\}^T(\boldsymbol{\delta}-\boldsymbol{\delta}_0)\Big]\begin{pmatrix}1\\ 0\end{pmatrix}\\
&\quad +\frac{b^2}{2}w^{(2)}(z_0)f_z(z_0)\gamma(z_0)\begin{pmatrix}\upsilon_2\\ 0\end{pmatrix}+o_p(n^{-1/2}).
\end{aligned}
$$

Note that the kernel function $K(t)$ is symmetric. Also, it is easy to see that $(I_{s,3})=o_p(n^{-1/2})$. Thus, the first statement of this lemma holds.

For the second statement on $\mathbf{R}_n^*$, under some mild conditions, one can write $\mathbf{R}_{n,1}^*(\mathbf{u},\eta,\tilde{\eta})-\mathbf{R}_{n,1}^*(\mathbf{w}^{(0)},\eta^{(0)},\tilde{\eta}^{(0)})=(I_{R,1})$ and $\mathbf{R}_{n,2}^*(\mathbf{u},\eta,\tilde{\eta})-\mathbf{R}_{n,2}^*(\mathbf{w}^{(0)},\eta^{(0)},\tilde{\eta}^{(0)})=(I_{R,2a})+(I_{R,2b})$, where

$$
(I_{R,1}) = -\frac{1}{n}\sum_{j=1}^{n}\mathbf{z}_{j0}^{*}\{\begin{pmatrix}u_j\mathbf{x}_j\\ \mathbf{v}_j\end{pmatrix}^T\eta_j\tau_{1,j}-\begin{pmatrix}w_j^{(0)}\mathbf{x}_j\\ \mathbf{v}_j\end{pmatrix}^T\eta_j^{(0)}\tau_{1,j}^{(0)}\}K_b(z_j-z_0),
$$

$$
\begin{aligned}
(I_{R,2a}) &= \frac{1}{n}\sum_{j=1}^n \mathbf{z}_{j0}^* \Big[\begin{pmatrix} \{u_j\eta_j\tau_j' + \tau_j\}\mathbf{x}_j \\ \eta_j\tau_j'\mathbf{v}_j \end{pmatrix}^T - \begin{pmatrix} \{w_j^{(0)}\eta_j^{(0)}\tau_j'^{(0)} + \tau_j^{(0)}\}\mathbf{x}_j \\ \eta_j^{(0)}\tau_j'^{(0)}\mathbf{v}_j \end{pmatrix}^T \Big] \{y_j \\
&\quad -\mu(w_j^{(0)}\eta_j^{(0)} + \tilde{\eta}_j^{(0)})\} K_b(z_j - z_0), \\
(I_{R,2b}) &= \frac{1}{n}\sum_{j=1}^n \mathbf{z}_{j0}^* \begin{pmatrix} \{u_j\eta_j\tau_j' + \tau_j\}\mathbf{x}_j \\ \eta_j\tau_j'\mathbf{v}_j \end{pmatrix}^T \{\mu(w_j^{(0)}\eta_j^{(0)} + \tilde{\eta}_j^{(0)}) - \mu(u_j\eta_j + \tilde{\eta}_j)\} K_b(z_j - z_0).
\end{aligned}
$$

Here, $\tau_j = \tau(u_j\eta_j + \tilde{\eta}_j)$, $\tau_j^{(0)} = \tau(w_j^{(0)}\eta_j^{(0)} + \tilde{\eta}_j^{(0)})$, $\tau_{1,j} = \tau_1(u_j\eta_j + \tilde{\eta}_j)$, $\tau_{1,j}^{(0)} = \tau_1(w_j^{(0)}\eta_j^{(0)} + \tilde{\eta}_j^{(0)})$, $\tau_j' = \tau'(u_j\eta_j + \tilde{\eta}_j)$, $\tau_j'^{(0)} = \tau'(w_j^{(0)}\eta_j^{(0)} + \tilde{\eta}_j^{(0)})$

Note that, similar to $(I_{s,1})$, we can prove that $(I_{R,2a}) = o_p(n^{-1/2})$. The remaining is to figure out the other two terms $(I_{R,1})$ and $(I_{R,2b})$.

Write $\mu_j'^{(0)} = \mu'(w_j^{(0)}\eta_j^{(0)} + \tilde{\eta}_j^{(0)})$, $\tau_{11,j}^{(0)} = \tau_{11}(w_j^{(0)}\eta_j^{(0)} + \tilde{\eta}_j^{(0)})\}$ and $\tau_{11}(s) = \mu'(s)\tau'(s)$, $\tau_{2,j}^{(0)} = \tau_2(w_j^{(0)}\eta_j^{(0)} + \tilde{\eta}_j^{(0)})\}$ and $\tau_2(s) = \mu''(s)\tau(s) = \mu'(s)\mu''(s)/a''\{u(s)\}$. Also, denote

$$
\begin{aligned}
\mathbf{\Gamma}_1(t) &= \begin{pmatrix} \mathrm{E}[\{w^{(0)}(z)\eta^{(0)}\tau_{2c}(w^{(0)}(z)\eta^{(0)} + \tilde{\eta}^{(0)}) + \tau_1(w^{(0)}(z)\eta^{(0)} + \tilde{\eta}^{(0)})\}\mathbf{x}_j \otimes \mathbf{x}_j^T | z = t] \\ \mathrm{E}[\{\eta^{(0)}\tau_{2c}(w^{(0)}(z)\eta^{(0)} + \tilde{\eta}^{(0)}) + \tau_1(w^{(0)}(z)\eta^{(0)} + \tilde{\eta}^{(0)})\}\mathbf{v}_j \otimes \mathbf{x}_j^T | z = t] \end{pmatrix} \\
\tilde{\mathbf{\Gamma}}_1(t) &= \begin{pmatrix} w^{(0)}(t)\mathrm{E}\{\eta^{(0)}\tau_{2c}(w^{(0)}(z)\eta^{(0)} + \tilde{\eta}^{(0)})\mathbf{x}_j \otimes \mathbf{v}_j^T | z = t\} \\ \mathrm{E}\{\eta^{(0)}\tau_{2c}(w^{(0)}(z)\eta^{(0)} + \tilde{\eta}^{(0)})\mathbf{v}_j \otimes \mathbf{v}_j^T | z = t\} \end{pmatrix} \\
\mathbf{\Gamma}_{R,1}(t) &= \begin{pmatrix} \mathrm{E}[\{w^{(0)}(z)\eta^{(0)}\tau_{2c}(w^{(0)}(z)\eta^{(0)} + \tilde{\eta}^{(0)}) + \tau_1(w^{(0)}(z)\eta^{(0)} + \tilde{\eta}^{(0)})\}\eta^{(0)}\mathbf{x}_j \otimes \mathbf{x}_j^T | z = t] \\ \mathrm{E}[\{\eta^{(0)}\}^2\tau_{2c}(w^{(0)}(z)\eta^{(0)} + \tilde{\eta}^{(0)})\mathbf{v}_j \otimes \mathbf{x}_j^T | z = t] \end{pmatrix} \\
\mathbf{\Gamma}_2 &= \begin{pmatrix} \mathrm{E}[\{w^{(0)}(z)\eta^{(0)}\tau_{11}(w^{(0)}(z)\eta^{(0)} + \tilde{\eta}^{(0)}) + \tau_1(w^{(0)}(z)\eta^{(0)} + \tilde{\eta}^{(0)})\}\mathbf{x}_j \otimes \mathbf{x}_j^T | z = t] \\ \mathrm{E}\{\eta^{(0)}\tau_{11}(w^{(0)}(z)\eta^{(0)} + \tilde{\eta}^{(0)})\mathbf{v}_j \otimes \mathbf{x}_j^T | z = t\} \end{pmatrix} \\
\tilde{\mathbf{\Gamma}}_2(t) &= \begin{pmatrix} \mathrm{E}[\{w^{(0)}(z)\eta^{(0)}\tau_{11}(w^{(0)}(z)\eta^{(0)} + \tilde{\eta}^{(0)}) + \tau_1(w^{(0)}(z)\eta^{(0)} + \tilde{\eta}^{(0)})\}\mathbf{x}_j \otimes \mathbf{v}_j^T | z = t] \\ \mathrm{E}\{\eta^{(0)}\tau_{11}(w^{(0)}(z)\eta^{(0)} + \tilde{\eta}^{(0)})\mathbf{v}_j \otimes \mathbf{v}_j^T | z = t\} \end{pmatrix} \\
\mathbf{\Gamma}_{R,2}(t) &= \begin{pmatrix} \mathrm{E}[\{w^{(0)}(z)\eta^{(0)}\tau_{11}(w^{(0)}(z)\eta^{(0)} + \tilde{\eta}^{(0)}) + \tau_1(w^{(0)}(z)\eta^{(0)} + \tilde{\eta}^{(0)})\}\eta^{(0)}\mathbf{x}_j \otimes \mathbf{x}_j^T | z = t] \\ \mathrm{E}[\{\eta^{(0)}\}^2\tau_{11}(w^{(0)}(z)\eta^{(0)} + \tilde{\eta}^{(0)})\mathbf{v}_j \otimes \mathbf{x}_j^T | z = t \end{pmatrix},
\end{aligned}
$$

where the function $\tau_{2c}(s) = \tau_{11}(s) + \tau_2(s)$. We can prove that

$$
\begin{aligned}
(I_{R,1}) &= -\frac{1}{n}\sum_{j=1}^n \mathbf{z}_{j0}^* \begin{pmatrix} \{u_j\eta_j - w_j^{(0)}\eta_j^{(0)}\}\mathbf{x}_j \\ (\eta_j - \eta_j^{(0)})\mathbf{v}_j \end{pmatrix}^T \tau_{1,j}^{(0)} K_b(z_j - z_0) \\
&\quad -\frac{1}{n}\sum_{j=1}^n \mathbf{z}_{j0}^* \begin{pmatrix} w_j^{(0)}\mathbf{x}_j \\ \mathbf{v}_j \end{pmatrix}^T \eta_j^{(0)}\{\tau_{11,j}^{(0)} + \tau_{2,j}^{(0)}\}\{u_j\eta_j - w_j^{(0)}\eta^{(0)} + \tilde{\eta}_j - \tilde{\eta}^{(0)}\} K_b(z_j - z_0) + o_p(\frac{1}{\sqrt{n}}) \\
&= -\frac{1}{n}\sum_{j=1}^n \mathbf{z}_{j0}^* \begin{pmatrix} [w_j^{(0)}\eta_j^{(0)}\{\tau_{11,j}^{(0)} + \tau_{2,j}^{(0)}\} + \tau_{1,j}^{(0)}]\mathbf{x}_j \\ [\eta^{(0)}\{\tau_{11,j}^{(0)} + \tau_{2,j}^{(0)}\} + \tau_{1,j}^{(0)}]\mathbf{v}_j \end{pmatrix}^T w_j^{(0)}\mathbf{x}_j^T(\boldsymbol{\beta} - \boldsymbol{\beta}_0) K_b(z_j - z_0) \\
&\quad -\frac{1}{n}\sum_{j=1}^n \mathbf{z}_{j0}^* \begin{pmatrix} w_j^{(0)}\eta_j^{(0)}\{\tau_{11,j}^{(0)} + \tau_{2,j}^{(0)}\}\mathbf{x}_j \\ \eta_j^{(0)}\{\tau_{11,j}^{(0)} + \tau_{2,j}^{(0)}\}\mathbf{v}_j \end{pmatrix}^T \mathbf{v}_j^T(\boldsymbol{\delta} - \boldsymbol{\delta}_0) K_b(z_j - z_0) + o_p(\frac{1}{\sqrt{n}}) \\
&\quad -\frac{1}{2n}\sum_{j=1}^n \mathbf{z}_{j0}^* \begin{pmatrix} [w_j^{(0)}\eta^{(0)}\{\tau_{11,j}^{(0)} + \tau_{2,j}^{(0)}\} + \tau_{1,j}^{(0)}]\mathbf{x}_j \\ \eta^{(0)}\{\tau_{11,j}^{(0)} + \tau_{2,j}^{(0)}\}\mathbf{v}_j \end{pmatrix}^T \eta_j^{(0)} w^{(2)}(z_0) t_j^2 b^2 K_b(z_j - z_0) \\
&= -\int \begin{pmatrix} 1 \\ t \end{pmatrix} w^{(0)}(z_0 + bt) K(t) \{\mathbf{\Gamma}_1(z_0 + bt)(\boldsymbol{\beta} - \boldsymbol{\beta}_0)\}^T f_z(z_0 + bt) dt \\
&\quad -\int \begin{pmatrix} 1 \\ t \end{pmatrix} K(t) \{\tilde{\mathbf{\Gamma}}_1(z_0 + bt)(\boldsymbol{\delta} - \boldsymbol{\delta}_0)\}^T f_z(z_0 + bt) dt \\
&\quad -\frac{1}{2} w^{(2)}(z_0) \int \begin{pmatrix} t \\ t^3 \end{pmatrix} K(t) \{\mathbf{\Gamma}_{R,1}(z_0 + bt)\, \boldsymbol{\beta}_0\}^T f_z(z_0 + bt) dt + o_p\Big(\frac{1}{\sqrt{n}}\Big)
\end{aligned}
$$

$$
\begin{aligned}
= \quad & -w^{(0)}(z_0)f_z(z_0)\begin{pmatrix}1\\0\end{pmatrix}\{\boldsymbol{\Gamma}_1(z_0)(\boldsymbol{\beta}-\boldsymbol{\beta}_0)\}^T - f_z(z_0)\begin{pmatrix}1\\0\end{pmatrix}\{\tilde{\boldsymbol{\Gamma}}_1(z_0)(\boldsymbol{\delta}-\boldsymbol{\delta}_0)\}^T \\
& -\frac{1}{2}b^2w^{(2)}(z_0)f_z(z_0)\begin{pmatrix}\upsilon_2\\0\end{pmatrix}\{\boldsymbol{\Gamma}_{R,1}(z_0)\,\boldsymbol{\beta}_0\}^T + o_p\Big(\frac{1}{\sqrt{n}}\Big)
\end{aligned}
$$

and

$$
\begin{aligned}
(I_{R,2b}) \quad = \quad & -\frac{1}{n}\sum_{j=1}^n \mathbf{z}_{j0}^*\begin{pmatrix}\{w_j^{(0)}\eta_j^{(0)}\tau_{11,j}^{(0)}+\tau_{1,j}^{(0)}\}\mathbf{x}_j\\ \eta_j^{(0)}\tau_{11,j}^{(0)}\mathbf{v}_j\end{pmatrix}^T\{w_j^{(0)}\mathbf{x}_j^T(\beta-\beta_0)+\mathbf{v}_j^T(\delta-\delta_0) \\
& +\frac{1}{2}\eta_j^{(0)}w^{(2)}(z_0)t_j^2b^2\}K_b(z_j-z_0)+o_p(\frac{1}{\sqrt{n}}) \\
= \quad & -\int\begin{pmatrix}1\\t\end{pmatrix}w^{(0)}(z_0+bt)K(t)\{\boldsymbol{\Gamma}_2(z_0+bt)(\boldsymbol{\beta}-\boldsymbol{\beta}_0)\}^Tf_z(z_0+bt)dt \\
& -\int\begin{pmatrix}1\\t\end{pmatrix}K(t)\{\tilde{\boldsymbol{\Gamma}}_2(z_0+bt)(\boldsymbol{\delta}-\boldsymbol{\delta}_0)\}^Tf_z(z_0+bt)dt \\
& -\frac{1}{2}w^{(2)}(z_0)\int\begin{pmatrix}1\\t\end{pmatrix}K(t)\{\boldsymbol{\Gamma}_{R,2}(z_0+bt)(\beta-\beta_0)\}^Tf_z(z_0+bt)dt+o_p\Big(\frac{1}{\sqrt{n}}\Big) \\
= \quad & -w^{(0)}(z_0)f_z(z_0)\begin{pmatrix}1\\0\end{pmatrix}\{\boldsymbol{\Gamma}_2(z_0)(\boldsymbol{\beta}-\boldsymbol{\beta}_0)\}^T - f_z(z_0)\begin{pmatrix}1\\0\end{pmatrix}\{\tilde{\boldsymbol{\Gamma}}_2(z_0)(\boldsymbol{\delta}-\boldsymbol{\delta}_0)\}^T \\
& -\frac{1}{2}b^2w^{(2)}(z_0)f_z(z_0)\begin{pmatrix}\upsilon_2\\0\end{pmatrix}\{\boldsymbol{\Gamma}_{R,2}(z_0)\,\boldsymbol{\beta}_0\}^T + o_p\Big(\frac{1}{\sqrt{n}}\Big)
\end{aligned}
$$

Denote that $\boldsymbol{\Gamma}(t)=\boldsymbol{\Gamma}_1(1)+\boldsymbol{\Gamma}_2(t)$, $\tilde{\boldsymbol{\Gamma}}(t)=\tilde{\boldsymbol{\Gamma}}_1(1)+\tilde{\boldsymbol{\Gamma}}_2(t)$, $\boldsymbol{\Gamma}_R(t)=\boldsymbol{\Gamma}_{R,1}(1)+\boldsymbol{\Gamma}_{R,2}(t)$, and put the above results together, we have the asymptotic expansion of $R_n^*(\mathbf{u},\eta)$ in the lemma.

For the last statement on the negative hessian matrix $\mathbf{H}_n^*$, under some mild conditions, one has $\mathbf{H}_{n,2}^*=O_p\{(nb)^{-1/2}+b^2\}$. Thus,

$$
\begin{aligned}
\mathbf{H}_n^*(\mathbf{u},\eta,\tilde{\eta}) \quad = \quad & \int\begin{pmatrix}1 & t\\ t & t^2\end{pmatrix}\gamma(z_0+bt)K(t)f_z(z_0+bt)dt+O_p(\frac{1}{\sqrt{nb}}+b^2) \\
= \quad & f_z(z_0)\gamma(z_0)\begin{pmatrix}1 & 0\\ 0 & \upsilon_2\end{pmatrix}+O_p(\frac{1}{\sqrt{nb}}+b^2)=\mathbf{H}+O_p(\frac{1}{\sqrt{nb}}+b^2).
\end{aligned}
$$

Proof of Theorem A. The first equation in Theorem A follows immediately from Lemma A1, and the Taylor expansion equation

$$
\mathbf{S}_n^*(\hat{\mathbf{u}},\eta,\tilde{\eta})-\mathbf{S}_n^*(\mathbf{u},\eta,\tilde{\eta})=\mathbf{H}_n^*(\mathbf{u},\eta,\tilde{\eta})(\widehat{\lambda}^*-\lambda^*)+o_p\Big(\frac{1}{\sqrt{n}}\Big),
$$

where where $\widehat{\mathbf{u}}_\theta=(\widehat{u}_{1,\theta},\widehat{u}_{2\theta},\ldots,\widehat{u}_{n,\theta})^T$ and $\widehat{u}_{j,\theta}=\mathbf{z}_{j,0}^T\widehat{\lambda}_\theta=\widehat{\lambda}_{0,\theta}+\widehat{\lambda}_{1,\theta}(z_j-z_0)$.

Let us move to the asymptotic expression of $\mathbf{R}_n^*(\mathbf{u},\eta,\tilde{\eta})$. We note that, for every given z_0 and θ, $\widehat{\mathbf{u}}_\theta$ solves equation $\mathbf{S}_n^*(\widehat{\mathbf{u}}_\theta,\eta,\tilde{\eta})=0$. Thus, taking derivative with respect to θ on both sides of the equation, we can get

$$
\mathbf{H}_n^*(\widehat{\mathbf{u}}_\theta,\eta,\tilde{\eta})\Big\{\frac{\partial}{\partial\theta^T}\widehat{\lambda}_\theta\Big\}=\mathbf{R}_n^*(\widehat{\mathbf{u}}_\theta,\eta,\tilde{\eta}).
$$

The second equation in Theorem A follows immediately from Lemma A1 and this above equation.

Proof of Corollary A1. The first result is obvious from Theorem A. For the second result, noting that the kernel function $K(t)$ is symmetric, we have from the standard result of kernel estimation

$$\frac{1}{n}\sum_{j=1}^{n}\mathbf{J}^{-1}\mathbf{z}_{j,0}\begin{pmatrix} w_j^{(0)}\mathbf{x}_j \\ \mathbf{v}_j \end{pmatrix}^T \eta_j^{(0)}\tau_{1,j}^{(0)}K_b(z_j-z_0) = f_z(z_0)\begin{pmatrix} 1 \\ 0 \end{pmatrix}\begin{pmatrix} w^{(0)}(z_0)\,\gamma_1(z_0) \\ \tilde{\gamma}_1(z_0) \end{pmatrix}^T + O_p\Big(\frac{1}{\sqrt{nb}}+b^2\Big)$$

The first row of the main term on the right hand side is just $\mathbf{H}[\{\mathbf{m}(z)\}^T,\{\tilde{\mathbf{m}}(z)\}]^T$. So, the second result follows immediately from Theorem A.

Proof Outline of Corollary A2. To save space, we only provide here a sketch outline on how to prove this lemma, omitting a detailed proof. The key technique in obtaining the uniform bounds are (a) establishing pointwise mean square bounds, and (b) using Rosenthal's inequality to establish the uniform bounds for $\boldsymbol{\beta}\in B_n(r)$, $\boldsymbol{\delta}\in\tilde{B}_n(\tilde{r})$, $\mathbf{x}\in\mathcal{X}$, $\mathbf{v}\in\mathcal{V}$ and $z\in\mathcal{Z}$, where the uniform bounds increase the point-wise bounds by n^{ζ}, for any $\zeta>0$. The establishment of pointwise bounds in part (a) follows from the standard argument in kernel estimation. The argument for part (b), after establishing part (a), follows (almost) exactly those discussions in "Step (ii) of Section 4" of Härdle, Hall and Ichimura (1993). Also, see the appendix of Weisberg and Welsh (1994), who also used the same approach.

APPENDIX B: Proofs of Theorems 1-3

B.3 Proof of Theorem 1.

Proof of Theorem 1. Based on the asymptotic expansion of $\widehat{\lambda}_\theta$ in Theorem A, one only needs to prove

$$\frac{1}{n}\sum_{j=1}^{n}\mathbf{J}^{-1}\mathbf{z}_{j,0}\{y_j-\mu(w_j^{(0)}\eta_j^{(0)}+\tilde{\eta}_j^{(0)})\}\eta_j^{(0)}\tau(w_j^{(0)}\eta_j^{(0)}+\tilde{\eta}_j^{(0)})K_b(z_j-z_0) \tag{A.3}$$

is asymptotically normally distributed. The term (A.3) is a summation of a series independent random variables of mean zero. Under some mild assumptions, by verifying the Lindeberg or Lyapounov condition, one can prove that this term is asymptotically normally distributed.

B.4 Proof of Theorem 2.

Denote by $\widehat{w}_0(z)=\widehat{w}_\theta(z)|_{\theta=\theta^{(0)}}$ and $\widehat{w}_0^{(1)}(z)=\widehat{w}_\theta^{(1)}(z)|_{\theta=\theta^{(0)}}$. We first prove the following lemmas.

Lemma B1. *Under the conditions of Theorem 2, it follows that*

$$\sup_{\beta\in B_n(r),\delta\in\tilde{B}_n(\tilde{r})}\Big|n^{-1/2}\sum_{i=1}^{n}[y_i-\mu\{\widehat{w}_0(z_i)\eta_i^{(0)}+\tilde{\eta}_i^{(0)}\}]\tau\{\widehat{w}_0(z_i)\eta_i+\tilde{\eta}_i\}\{\begin{pmatrix} \widehat{w}_0(z_i)\mathbf{x}_i \\ \mathbf{v}_i \end{pmatrix}+\eta_i^{(0)}\widehat{w}_0^{(1)}(z_i)\}$$
$$-n^{-1/2}\sum_{i=1}^{n}\{y_i-\mu(w_i^{(0)}\eta_i^{(0)}+\tilde{\eta}_i^{(0)})\}\tau(w_i^{(0)}\eta_i^{(0)}+\tilde{\eta}_i^{(0)})\{\begin{pmatrix} w_i^{(0)}\mathbf{x}_i \\ \mathbf{v}_i \end{pmatrix}-\eta_i^{(0)}\begin{pmatrix} \mathbf{m}(z_i) \\ \tilde{\mathbf{m}}(z_i) \end{pmatrix}\}\Big| = o_p(1).$$

Proof of Lemma B1. The term inside the absolute bars equals to $(I_1)+(I_2)+(I_3)$, where

$$\begin{aligned}
(I_1) &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\{y_i-\mu(w_i^{(0)}\eta_i^{(0)}+\tilde{\eta}_i^{(0)})\}\Big[\tau\{\widehat{w}_0(z_i)\eta_i+\tilde{\eta}_i\}\{\begin{pmatrix}\widehat{w}_0(z_i)\mathbf{x}_i\\ \mathbf{v}_i\end{pmatrix}+\eta_i^{(0)}\widehat{w}_0^{(1)}(z_i)\}- \\
&\qquad \tau(w_i^{(0)}\eta_i^{(0)}+\tilde{\eta}_i^{(0)})\{\begin{pmatrix}w_i^{(0)}\mathbf{x}_i\\ \mathbf{v}_i\end{pmatrix}-\eta_i^{(0)}\begin{pmatrix}\mathbf{m}(z_i)\\ \tilde{\mathbf{m}}(z_i)\end{pmatrix}\}\Big] \\
(I_2) &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}[\mu(w_i^{(0)}\eta_i^{(0)}+\tilde{\eta}_i^{(0)})-\mu\{\widehat{w}_0(z_i)\eta_i^{(0)}+\tilde{\eta}_i^{(0)}\}]\Big[\tau\{\widehat{w}_0(z_i)\eta_i+\tilde{\eta}_i\}\{\begin{pmatrix}\widehat{w}_0(z_i)\mathbf{x}_i\\ \mathbf{v}_i\end{pmatrix}+ \\
&\qquad \eta_i^{(0)}\widehat{w}_0^{(1)}(z_i)\}-\tau(w_i^{(0)}\eta_i^{(0)}+\tilde{\eta}_i^{(0)})\{\begin{pmatrix}w_i^{(0)}\mathbf{x}_i\\ \mathbf{v}_i\end{pmatrix}-\eta_i^{(0)}\begin{pmatrix}\mathbf{m}(z_i)\\ \tilde{\mathbf{m}}(z_i)\end{pmatrix}\}\Big] \\
(I_3) &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}[\mu(w_i^{(0)}\eta_i^{(0)}+\tilde{\eta}_i^{(0)})-\mu\{\widehat{w}_0(z_i)\eta_i^{(0)}+\tilde{\eta}_i^{(0)}\}]\tau(w_i^{(0)}\eta_i^{(0)}+\tilde{\eta}_i^{(0)})\{\begin{pmatrix}w_i^{(0)}\mathbf{x}_i\\ \mathbf{v}_i\end{pmatrix}-\eta_i^{(0)}\begin{pmatrix}\mathbf{m}(z_i)\\ \tilde{\mathbf{m}}(z_i)\end{pmatrix}\}.
\end{aligned}$$

Based on Corollary A2, each summation term in (I_2) is at most of order $o_p(n^{-1+2\zeta}b^{-1})$. Thus, $\sup_{\beta\in B_n(r),\delta\in\tilde{B}_n(\tilde{r})}|(I_2)|=o_p(1)$. It remains to prove that both (I_1) and (I_3) are $o_p(1)$.

From the asymptotic expression of $\widehat{w}_0(z_i)$ and $\widehat{w}_0^{(1)}(z_i)$ in (A.1) and (A.2), we have

$$\begin{aligned}
(I_1) &= n^{-3/2}\sum_{i=1}^{n}\sum_{j=1}^{n}\{y_i-\mu(w_i^{(0)}\eta_i^{(0)}+\tilde{\eta}_i^{(0)})\}\{y_j-\mu(w_j^{(0)}\eta_j^{(0)}+\tilde{\eta}_j^{(0)})\}\bar{\mathbf{q}}_{j,i}K_b(z_j-z_i)+o_p(1) \\
&= n^{-3/2}\sum_{i\neq j}\{y_i-\mu(w_i^{(0)}\eta_i^{(0)}+\tilde{\eta}_i^{(0)})\}\{y_j-\mu(w_j^{(0)}\eta_j^{(0)}+\tilde{\eta}_j^{(0)})\}\bar{\mathbf{q}}_{j,i}K_b(z_j-z_i) \\
&\quad +n^{-3/2}\sum_{i=1}^{n}\{y_i-\mu(w_i^{(0)}\eta_i^{(0)}+\tilde{\eta}_i^{(0)})\}^2\bar{\mathbf{q}}_{j,i}K_b(0)+o_p(1).
\end{aligned}$$

Here $\bar{\mathbf{q}}_{j,i}=\tau(w_i^{(0)}\eta_i^{(0)}+\tilde{\eta}_i^{(0)})\{(q_{j,i}\mathbf{x}_i^T,0)^T+\eta_i^{(0)}\mathbf{q}_{j,i}\}$. Clearly, the second term at the right hand side is $o_p(1)$ and the first term forms a martingale with respect to $\mathcal{F}_n=\sigma\{(y_1,\mathbf{x}_1,\mathbf{v}_1,z_1),\ldots,(y_n,\mathbf{x}_n,\mathbf{v}_n,z_n)\}$. Thus,

$$\begin{aligned}
& E|n^{-3/2}\sum_{i\neq j}\{y_i-\mu(w_i^{(0)}\eta_i^{(0)}+\tilde{\eta}_i^{(0)})\}\{y_j-\mu(w_j^{(0)}\eta_j^{(0)}+\tilde{\eta}_j^{(0)})\}\bar{\mathbf{q}}_{j,i}K_b(z_j-z_i)|^2 \\
=\ & n^{-3}\frac{n(n-1)}{2b^2}E|\{y_i-\mu(w_i^{(0)}\eta_i^{(0)}+\tilde{\eta}_i^{(0)})\}\{y_j-\mu(w_j^{(0)}\eta_j^{(0)}+\tilde{\eta}_j^{(0)})\}\bar{\mathbf{q}}_{j,i}K\Big(\frac{z_j-z_i}{b}\Big) \\
& +\{y_j-\mu(w_j^{(0)}\eta_j^{(0)}+\tilde{\eta}_i^{(0)})\}\{y_i-\mu(w_i^{(0)}\eta_i^{(0)}+\tilde{\eta}_i^{(0)})\}\bar{\mathbf{q}}_{j,i}K\Big(\frac{z_i-z_j}{b}\Big)|^2=O_p(n^{-1}b^{-2}),
\end{aligned}$$

and it follows that $\sup_{\beta\in B_n(r),\delta\in\tilde{B}_n(\tilde{r})}|(I_1)|=o_p(1)$. A similar argument can be found in Weisberg and Welsh (1994) (proof of $\sup_{\beta\in B_n(r)}|T_5|=o_p(1)$, page 1698).

Since $\mu(w_i^{(0)}\eta_i^{(0)}+\tilde{\eta}_i^{(0)})-\mu(\widehat{w}_0(z_i)\eta_i^{(0)}+\tilde{\eta}_i^{(0)})=-\mu'(w_i^{(0)}\eta_i^{(0)}+\tilde{\eta}_i^{(0)})\eta_i^{(0)}\{\widehat{w}_0(z_i)-w_i^{(0)}\}+o_p(n^{-1/2+\zeta}b^{-1/2})$, we have $(I_3)=(I_{3a})+o_p(1)$, where

$$\begin{aligned}
(I_{3a}) &= \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\tau_1(w_i^{(0)}\eta_i^{(0)}+\tilde{\eta}_i^{(0)})\eta_i^{(0)}\{\begin{pmatrix}w_i^{(0)}\mathbf{x}_i\\ \mathbf{v}_i\end{pmatrix}-\eta_i^{(0)}\begin{pmatrix}\mathbf{m}(z_i)\\ \tilde{\mathbf{m}}(z_i)\end{pmatrix}\}\{\widehat{w}_0(z_i)-w_i^{(0)}\} \\
&= \frac{1}{n^{3/2}}\sum_{i=1}^{n}\sum_{j=1}^{n}[\tau_1(w_i^{(0)}\eta_i^{(0)}+\tilde{\eta}_i^{(0)})\eta_i^{(0)}\{\begin{pmatrix}w_i^{(0)}\mathbf{x}_i\\ \mathbf{v}_i\end{pmatrix}-\eta_i^{(0)}\begin{pmatrix}\mathbf{m}(z_i)\\ \tilde{\mathbf{m}}(z_i)\end{pmatrix}\}]\{y_j- \\
&\qquad \mu(w_j^{(0)}\eta_j^{(0)}+\tilde{\eta}_j^{(0)})\}q_{j,i}K_b(z_j-z_i)+o_p(1).
\end{aligned}$$

We have a key equation that $\mathrm{E}[\tau_1(w_i^{(0)}\eta_i^{(0)}+\tilde{\eta}_i^{(0)})\eta_i^{(0)}\{\begin{pmatrix} w_i^{(0)}\mathbf{x}_i \\ \mathbf{v}_i \end{pmatrix} - \eta_i^{(0)}\begin{pmatrix} \mathbf{m}(z_i) \\ \tilde{\mathbf{m}}(z_i) \end{pmatrix}\}] = 0$. Following the same argument as in (I_1), one can prove that $\sup_{\beta\in B_n(r),\delta\in\tilde{B}_n(\tilde{r})}|(I_{3a})| = o_p(1)$. The lemma follows.

Lemma B2. *Under the conditions of Theorem 2, it follows that*

$$\sup_{\beta\in B_n(r),\delta\in\tilde{B}_n(\tilde{r})}\Bigg| n^{-1/2}\sum_{i=1}^n [y_i - \mu\{\widehat{w}_\theta(z_i)\eta_i+\tilde{\eta}_i\}]\tau\{\widehat{w}_\theta(z_i)\eta_i+\tilde{\eta}_i\}\{\begin{pmatrix} \widehat{w}_\theta(z_i)\mathbf{x}_i \\ \mathbf{v}_i \end{pmatrix} + \eta_i\widehat{w}_\theta^{(1)}(z_i)\}$$
$$-n^{-1/2}\sum_{i=1}^n [y_i - \mu\{\widehat{w}_0(z_i)\eta_i^{(0)}+\tilde{\eta}_i^{(0)}\}]\tau\{\widehat{w}_0(z_i)\eta_i+\tilde{\eta}_i\}\{\begin{pmatrix} \widehat{w}_0(z_i)\mathbf{x}_i \\ \mathbf{v}_i \end{pmatrix} + \eta_i^{(0)}\widehat{w}_0^{(1)}(z_i)\}$$
$$+\mathbf{A}^{-1}n^{1/2}(\boldsymbol{\theta}-\boldsymbol{\theta}_0)\Bigg| = o_p(1).$$

Proof of Lemma B2. The term inside the absolute bars equals to $(II_1)+(II_2)+(II_3)+(II_4)+(II_5)$, where

$$\begin{aligned}
(II_1) &= -\frac{1}{\sqrt{n}}\sum_{i=1}^n [\mu\{\widehat{w}_\theta(z_i)\eta_i+\tilde{\eta}_i\} - \mu\{\widehat{w}_0(z_i)\eta_i^{(0)}+\tilde{\eta}_i^{(0)}\}]\tau(\widehat{w}_0(z_i)\eta_i+\tilde{\eta}_i)\{\begin{pmatrix} \widehat{w}_0(z_i)\mathbf{x}_i \\ \mathbf{v}_i \end{pmatrix} \\
&\qquad +\eta_i^{(0)}\widehat{w}_0^{(1)}(z_i)\} + \mathbf{A}^{-1}n^{1/2}(\boldsymbol{\theta}-\boldsymbol{\theta}^{(0)}), \\
(II_2) &= -\frac{1}{\sqrt{n}}\sum_{i=1}^n [\mu\{\widehat{w}_\theta(z_i)\eta_i+\tilde{\eta}_i\} - \mu\{\widehat{w}_0(z_i)\eta_i^{(0)}+\tilde{\eta}_i^{(0)}\}]\Big[\tau\{\widehat{w}_\theta(z_i)\eta_i+\tilde{\eta}_i\}\{\begin{pmatrix} \widehat{w}_\theta(z_i)\mathbf{x}_i \\ \mathbf{v}_i \end{pmatrix} \\
&\qquad +\eta_i\widehat{w}_\theta^{(1)}(z_i)\} - \tau(\widehat{w}_0(z_i)\eta_i+\tilde{\eta}_i)\{\begin{pmatrix} \widehat{w}_0(z_i)\mathbf{x}_i \\ \mathbf{v}_i \end{pmatrix} + \eta_i^{(0)}\widehat{w}_0^{(1)}(z_i)\}\Big], \\
(II_3) &= -\frac{1}{\sqrt{n}}\sum_{i=1}^n [\mu\{w_i^{(0)}\eta_i^{(0)}+\tilde{\eta}_i^{(0)}\} - \mu\{\widehat{w}_0(z_i)\eta_i^{(0)}+\tilde{\eta}_i^{(0)}\}]\Big[\tau\{\widehat{w}_\theta(z_i)\eta_i+\tilde{\eta}_i\}\{\begin{pmatrix} \widehat{w}_\theta(z_i)\mathbf{x}_i \\ \mathbf{v}_i \end{pmatrix} \\
&\qquad +\eta_i\widehat{w}_\theta^{(1)}(z_i)\} - \tau\{\widehat{w}_0(z_i)\eta_i^{(0)}+\tilde{\eta}_i^{(0)}\}\{\begin{pmatrix} \widehat{w}_0(z_i)\mathbf{x}_i \\ \mathbf{v}_i \end{pmatrix} + \eta_i^{(0)}\widehat{w}_0^{(1)}(z_i)\}\Big], \\
(II_4) &= \frac{1}{\sqrt{n}}\sum_{i=1}^n [y_i - \mu\{w_i^{(0)}\eta_i^{(0)}+\tilde{\eta}_i^{(0)}\}]\Big[\tau\{\widehat{w}_0(z_i)\eta_i^{(0)}+\tilde{\eta}_i^{(0)}\}\{\begin{pmatrix} \widehat{w}_0(z_i)\mathbf{x}_i \\ \mathbf{v}_i \end{pmatrix} + \eta_i^{(0)}\widehat{w}_0^{(1)}(z_i)\} \\
&\qquad -\tau(w_i^{(0)}\eta_i^{(0)}+\tilde{\eta}_i^{(0)})\{\begin{pmatrix} w_i^{(0)}\mathbf{x}_i \\ \mathbf{v}_i \end{pmatrix} - \eta_i^{(0)}\begin{pmatrix} \mathbf{m}(z_i) \\ \tilde{\mathbf{m}}(z_i) \end{pmatrix}\}\Big], \\
(II_5) &= -\frac{1}{\sqrt{n}}\sum_{i=1}^n [y_i - \mu\{w_i^{(0)}\eta_i^{(0)}+\tilde{\eta}_i^{(0)}\}]\Big[\tau\{\widehat{w}_\theta(z_i)\eta_i+\tilde{\eta}_i\}\{\begin{pmatrix} \widehat{w}_\theta(z_i)\mathbf{x}_i \\ \mathbf{v}_i \end{pmatrix} + \eta_i\widehat{w}_\theta^{(1)}(z_i)\} - \\
&\qquad \tau(w_i^{(0)}\eta_i^{(0)}+\tilde{\eta}_i^{(0)})\{\begin{pmatrix} w_i^{(0)}\mathbf{x}_i \\ \mathbf{v}_i \end{pmatrix} - \eta_i^{(0)}\begin{pmatrix} \mathbf{m}(z_i) \\ \tilde{\mathbf{m}}(z_i) \end{pmatrix}\}\Big],
\end{aligned}$$

Similar to (I_2), we can proved that $\sup_{\beta\in B_n(r),\delta\in\tilde{B}_n(\tilde{r})}(II_2) = o_p(1)$ and $\sup_{\beta\in B_n(r),\delta\in\tilde{B}_n(\tilde{r})}(II_3) = o_p(1)$. Similar to (I_1), we can prove that $\sup_{\beta\in B_n(r),\delta\in\tilde{B}_n(\tilde{r})}(II_4) = o_p(1)$ and $\sup_{\beta\in B_n(r),\delta\in\tilde{B}_n(\tilde{r})}(II_5) = o_p(1)$. It remains to prove that (II_1) is $o_p(1)$. To this end, note that $\mu\{\widehat{w}_\beta(z_i)\eta_i+\tilde{\eta}_i\} - \mu\{\widehat{w}_0(z_i)\eta_i^{(0)}+\tilde{\eta}_i^{(0)}\} = \mu'\{\widehat{w}_0(z_i)\eta_i^{(0)}+\tilde{\eta}_i^{(0)}\}\{\begin{pmatrix} w_i^{(0)}\mathbf{x}_i \\ \mathbf{v}_i \end{pmatrix} + \eta_i^{(0)}\widehat{w}_0^{(1)}(z_i)\}^T(\boldsymbol{\theta}-\boldsymbol{\theta}_0) + o_p(n^{-1/2})$. Also, $\mathbf{A}_n^{-1}(\boldsymbol{\theta})$ is continuous in $\boldsymbol{\theta}$, one can directly verify that $\sup_{\beta\in B_n(r),\delta\in\tilde{B}_n(\tilde{r})}|\mathbf{A}_n^{-1}(\boldsymbol{\theta}) - \mathbf{A}^{-1}| = o_p(1)$. It follows that $\sup_{\beta\in B_n(r),\delta\in B_n(\tilde{r})}(II_1) = o_p(1)$. Thus the lemma hold.

From Lemmas B1 and B2, we have immediately Lemma B3.

Lemma B3. *Under the condition of Theorem 2, it follows that, for a fixed $r > 0$,*

$$\sup_{\beta\in B_n(r),\delta\in\tilde{B}_n(\tilde{r})}\Bigg| n^{-1/2}\sum_{i=1}^n [y_i - \mu\{\widehat{w}_\theta(z_i)\eta_i+\tilde{\eta}_i\}]\tau\{\widehat{w}_\theta(z_i)\eta_i+\tilde{\eta}_i\}\{\begin{pmatrix} \widehat{w}_\theta(z_i)\mathbf{x}_i \\ \mathbf{v}_i \end{pmatrix} + \eta_i\widehat{w}_\theta^{(1)}(z_i)\}$$

$$-n^{-1/2}\sum_{i=1}^{n}\{y_i-\mu(w_i^{(0)}\eta_i^{(0)}+\tilde{\eta}_i^{(0)})\}\tau(w_i^{(0)}\eta_i^{(0)}+\tilde{\eta}_i^{(0)})\{\begin{pmatrix} w_i^{(0)}\mathbf{x}_i \\ \mathbf{v}_i \end{pmatrix}-\eta_i^{(0)}\begin{pmatrix} \mathbf{m}(z_i) \\ \tilde{\mathbf{m}}(z_i) \end{pmatrix}\}$$
$$+\mathbf{A}^{-1}n^{1/2}(\boldsymbol{\theta}-\boldsymbol{\theta}_0)\Big| = o_p(1)$$

Sketch Proof of Theorem 2. From Lemma B3, it follows that

$$\sum_{i=1}^{n}(\boldsymbol{\theta}-\boldsymbol{\theta}_0)^T[y_i-\mu\{\widehat{w}_\theta(z_i)\eta_i+\tilde{\eta}_i\}]\tau\{\widehat{w}_\theta(z_i)\eta_i+\tilde{\eta}_i\}\{\begin{pmatrix} \widehat{w}_\theta(z_i)\mathbf{x}_i \\ \mathbf{v}_i \end{pmatrix}+\eta_i\widehat{w}_\theta^{(1)}(z_i)\}<0$$

for $\boldsymbol{\beta}\notin B_n(r)$, $\boldsymbol{\delta}\notin \tilde{B}_n(\tilde{r})$ and some big enough but finite r and $\tilde{r}$, in probability. Note that the left hand side of the above equation is continuous in $\boldsymbol{\theta}=(\boldsymbol{\beta}^T,\boldsymbol{\delta}^T)^T$. The existence and $\sqrt{n}$-consistency results follow from (6.3.4) of Ortega and Rheinboldt (1973, page 163).

Also, from Lemma B3,

$$\sqrt{n}(\widehat{\boldsymbol{\theta}}^{\text{new}}-\boldsymbol{\theta}_0)=\mathbf{A}\frac{1}{\sqrt{n}}\sum_{i=1}^{n}\{y_i-\mu(w_i^{(0)}\eta_i^{(0)}+\tilde{\eta}_i^{(0)})\}\tau(w_i^{(0)}\eta_i^{(0)}+\tilde{\eta}_i^{(0)})\{\begin{pmatrix} w_i^{(0)}\mathbf{x}_i \\ \mathbf{v}_i \end{pmatrix}-\eta_i^{(0)}\begin{pmatrix} \mathbf{m}(z_i) \\ \tilde{\mathbf{m}}(z_i) \end{pmatrix}\}+o_p(1).$$

By checking the Lindeberg conditions, one can conclude the asymptotic normality result.

As mentioned in the proof of Lemma B2, $\sup_{\beta\in B_n(r),\delta\in\tilde{B}_n(\tilde{r})}|\mathbf{A}_n^{-1}(\theta)-\mathbf{A}^{-1}|=o_p(1)$. The final result on the consistent estimate of A is obvious.

B.5 Proof of Theorem 3.

Construction of Efficient Score Function $\mathcal{S}^*$. Let $\mathrm{g}(\mathbf{x},\mathbf{v},z)$ be the joint density of ($\mathbf{x}$, $\mathbf{v}$, z). The joint density of $(y,\mathbf{x},\mathbf{v},z)$ is

$$f(y,\mathbf{x},\mathbf{v},z)=\exp\left[\{y\psi-a(\psi)\}/\phi+b(y,\phi)\right]\mathrm{g}(\mathbf{x},\mathbf{v},z), \tag{A.4}$$

where $\psi=\mathrm{u}(w^{(0)}(z)\eta^{(0)}+\widetilde{\eta}^{(0)})$. Define $P=\Big\{$Model (A.4) with given $\boldsymbol{\theta}_0=(\boldsymbol{\beta}_0^T,\boldsymbol{\delta}_0^T)^T$ and density function $\mathrm{g}(\cdot)\Big\}$. Then, by the standard argument (for example, Bickel et al., 1993), the tangent space of the nonparametric model P is $\Big\{[y-\mu\{w^{(0)}(z)\eta^{(0)}+\widetilde{\eta}^{(0)}\}]\tau\{w^{(0)}(z)\eta^{(0)}+\widetilde{\eta}^{(0)}\}\eta^{(0)}a^*(z)|$ for all $a^*\in L_2\Big\}$. Thus, the efficient score function is $\mathcal{S}^*=\mathcal{S}-\{$Projection of $\mathcal{S}$ onto $P\}$, where $\mathcal{S}=[y-\mu\{w^{(0)}(z)\eta^{(0)}+\widetilde{\eta}^{(0)}\}]\tau\{w^{(0)}(z)\eta^{(0)}+\widetilde{\eta}^{(0)}\}\begin{pmatrix} w^{(0)}(z)\mathbf{x} \\ \mathbf{v} \end{pmatrix}$ is the score function for $\boldsymbol{\theta}_0$. Note that the mean square error $\mathrm{E}|\{y-\mu(w^{(0)}(z)\eta^{(0)}+\widetilde{\eta}^{(0)})\}\ \tau\{w^{(0)}(z)\eta^{(0)}+\widetilde{\eta}^{(0)}\}\{\begin{pmatrix} w^{(0)}(z)\mathbf{x} \\ \mathbf{v} \end{pmatrix}-\eta^{(0)}a^*(z)\}|^2$ achieves its minimum when $a^*(z)=(\{\mathbf{m}(z)\}^T,\{\widetilde{\mathbf{m}}(z)\}^T)^T$. Thus, the efficient score

$$\mathcal{S}^*=[y-\mu\{w^{(0)}(z)\eta^{(0)}+\widetilde{\eta}^{(0)}\}]\tau\{w^{(0)}(z)\eta^{(0)}+\widetilde{\eta}^{(0)}\}\Big\{\begin{pmatrix} w^{(0)}(z)\mathbf{x} \\ \mathbf{v} \end{pmatrix}-\eta^{(0)}\begin{pmatrix} \mathbf{m}(z) \\ \widetilde{\mathbf{m}}(z) \end{pmatrix}\Big\}. \tag{A.5}$$

For each $\boldsymbol{\beta}$ and $\boldsymbol{\delta}$, replacing $w^{(0)}(\cdot)$, $\mathbf{m}(\cdot)$ and $\widetilde{\mathbf{m}}(\cdot)$ by their estimators in $\mathcal{S}^*$, leads to estimating equations (7) in the paper.

Proof of Theorem 3. The form of the efficient score function $\mathcal{S}^*$ is given in equation (A.5). The Fisher information lower bound is $\mathrm{E}\{\mathcal{S}^*\mathcal{S}^{*T}\}$, which is equal to $\mathbf{A}^{-1}$; see Bickel et al. (1993). The theorem follows.

References

Bickel, P., Klaasen, C., Ritov, Y., and Wellner, J. (1993). *Efficient and Adaptive Inference for Semiparametric Models.* Baltimore: Johns Hopkins University Press.

Cai, Z., Fan, J. and Li, R.Z. (2000). Efficient estimation and inferences for varying-coefficient models. *Journal of American Statistical Association*, 95, 888-902.

Carroll, R. J., Fan, J., Gijbels, I. and Wand, M. P. (1997). Generalized partially linear single-index models. *Journal of American Statistical Association*, 92, 477–489.

Härdle, W., Hall, P. and Ichimura, H. (1993) Optimal Smoothing of Single Index models. Annals of Statistics, 21, 157-178.

Weisberg, S. and Welsh, A. H. (1994). Adapting for the missing link. *The Annals of Statistics.* 22, 1674-1700.