Supplement to " Semiparametric Estimation and Inference of Variance Function with Large Dimensional Covariates"

Yanyuan Ma and[1] and Liping Zhu[2]

[1]*Pennsylvania State University,* [2]*Renmin University of China*

Abstract

In this supplementary document, we provide some technical details.

1. The derivation of the efficient score $\mathbf{S}_{\rm eff}$ in model (1.2)

The derivation is split into three steps. We first derive the nuisance tangent space Λ, we then derive its orthogonal complement $\Lambda^{\perp}$, and finally, we calculate the score function and project it onto $\Lambda^{\perp}$ to obtain the efficient score.

The nuisance tangent spaces with respect to η_1 and η_2 are respectively

$$
\begin{aligned}
\Lambda_1 &= \{\mathbf{g}(\mathbf{x}) : E(\mathbf{g}) = \mathbf{0}, \mathbf{g} \in \mathbb{R}^{d_t}\}, \\
\Lambda_2 &= \{\mathbf{f}(\epsilon, \mathbf{x}) : E(\mathbf{f} \mid \mathbf{x}) = \mathbf{0}, E(\epsilon \mathbf{f} \mid \mathbf{x}) = \mathbf{0}, E(\epsilon^2 \mathbf{f} \mid \mathbf{x}) = \mathbf{0}, \mathbf{f} \in \mathbb{R}^{d_t}\}
\end{aligned}
$$

$$= \{\mathbf{f}(\epsilon, \mathbf{x}) : E(\mathbf{f} \mid \mathbf{x}) = \mathbf{0}, E(\epsilon\mathbf{f} \mid \mathbf{x}) = \mathbf{0}, E(u\mathbf{f} \mid \mathbf{x}) = \mathbf{0}, \mathbf{f} \in \mathbb{R}^{d_t}\}.$$

where u is defined in (2.1). Note that $E(u \mid \mathbf{x}) = 0$, $E(\epsilon u \mid \mathbf{x}) = 0$, $E(u^2 \mid \mathbf{x}) = 1$. Further calculating the nuisance tangent space with respect to m and σ, we have

$$\begin{aligned}\Lambda_m &= \{\eta'_{2\epsilon}/\eta_2 \mathbf{a}(\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{x})/\sigma(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}) : \mathbf{a} \in \mathbb{R}^{d_t}\} \\ \Lambda_\sigma &= \{(\epsilon\eta'_{2\epsilon}/\eta_2 + 1)\mathbf{b}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}) : \mathbf{b} \in \mathbb{R}^{d_t}\}.\end{aligned}$$

Combining these four spaces, we obtain the nuisance tangent space as $\Lambda = \Lambda_1 \oplus (\Lambda_2 + \Lambda_m + \Lambda_\sigma)$. Here $\Lambda_1 \perp \Lambda_2 + \Lambda_m + \Lambda_\sigma$ and the notation $\oplus$ is used to emphasize the orthogonality. Note that $E(\eta'_{2\epsilon}/\eta_2 \mid \mathbf{x}) = 0$, $E(\eta'_{2\epsilon}/\eta_2\epsilon \mid \mathbf{x}) = -1$, $E(\eta'_{2\epsilon}/\eta_2\epsilon^2 \mid \mathbf{x}) = 0$, $E(\eta'_{2\epsilon}/\eta_2\epsilon^3 \mid \mathbf{x}) = -3$, so $E(\eta'_{2\epsilon}/\eta_2 u \mid \mathbf{x}) = c(\mathbf{x})E(\epsilon^3 \mid \mathbf{x})$, $E(\eta'_{2\epsilon}/\eta_2 u\epsilon \mid \mathbf{x}) = -2c(\mathbf{x})$. Here $c(\mathbf{x})$ is defined in (2.1). Calculating the residual of projecting any function in $\Lambda_m, \Lambda_\sigma$ to Λ_2, we obtain

$$\begin{aligned}&\frac{\eta'_{2\epsilon}}{\eta_2}\frac{\mathbf{a}(\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{x})}{\sigma(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x})} = \left\{\frac{\eta'_{2\epsilon}}{\eta_2} + \epsilon - c(\mathbf{x})E(\epsilon^3 \mid \mathbf{x})u\right\}\frac{\mathbf{a}(\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{x})}{\sigma(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x})} - \{\epsilon - uc(\mathbf{x})E(\epsilon^3 \mid \mathbf{x})\}\frac{\mathbf{a}(\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{x})}{\sigma(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x})}, \\ &\left(\epsilon\frac{\eta'_{2\epsilon}}{\eta_2} + 1\right)\mathbf{b}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}) = \left\{\epsilon\frac{\eta'_{2\epsilon}}{\eta_2} + 1 + 2c(\mathbf{x})u\right\}\mathbf{b}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}) - 2uc(\mathbf{x})\mathbf{b}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}),\end{aligned}$$

where the first summand on the right side of each display is an element in

Λ_2, while the second summand is orthogonal to Λ_2. Hence

$$\begin{aligned}
\Lambda'_\sigma &= \Pi(\Lambda_\sigma \mid \Lambda_2^\perp) = \left\{uc(\mathbf{x})\mathbf{b}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}) : \mathbf{b} \in \mathbb{R}^{d_t}\right\},\\
\Lambda'_m &= \Pi(\Lambda_m \mid \Lambda_2^\perp) = \left[\left\{\epsilon - uc(\mathbf{x})E(\epsilon^3 \mid \mathbf{x})\right\}\frac{\mathbf{a}(\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{x})}{\sigma(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x})} : \mathbf{a} \in \mathbb{R}^{d_t}\right],\\
\Lambda'_m + \Lambda'_\sigma &= \left[uc(\mathbf{x})\mathbf{b}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}) + \left\{\epsilon - uc(\mathbf{x})E(\epsilon^3 \mid \mathbf{x})\right\}\frac{\mathbf{a}(\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{x})}{\sigma(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x})} : \mathbf{a}, \mathbf{b} \in \mathbb{R}^{d_t}\right],
\end{aligned}$$

and subsequently

$$\begin{aligned}
\Lambda &= \Lambda_1 \oplus \Lambda_2 \oplus (\Lambda'_\sigma + \Lambda'_m)\\
&= \{\mathbf{g}(\mathbf{x}) : E(\mathbf{g}) = \mathbf{0}, \mathbf{g} \in \mathbb{R}^{d_t}\} \oplus \{\mathbf{f}(\epsilon, \mathbf{x}) : E(\mathbf{f} \mid \mathbf{x}) = \mathbf{0}, E(\epsilon\mathbf{f} \mid \mathbf{x}) = \mathbf{0}, E(u\mathbf{f} \mid \mathbf{x}) = \mathbf{0}, \mathbf{f} \in \mathbb{R}^{d_t}\}\\
&\quad \oplus \left[uc(\mathbf{x})\mathbf{b}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}) + \left\{\epsilon - uc(\mathbf{x})E(\epsilon^3 \mid \mathbf{x})\right\}\frac{\mathbf{a}(\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{x})}{\sigma(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x})} : \mathbf{a}, \mathbf{b} \in \mathbb{R}^{d_t}\right].
\end{aligned}$$

We can now calculate the orthogonal complement of Λ by sequentially considering the orthogonal complement of Λ_1, $\Lambda_1 \oplus \Lambda_2$ and $\Lambda_1 \oplus \Lambda_2 \oplus (\Lambda'_\sigma + \Lambda'_m)$, and obtain

$$\Lambda^\perp = [\epsilon\sigma(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x})\{\mathbf{a}(\mathbf{x}) - E(\mathbf{a} \mid \boldsymbol{\alpha}^{\mathrm{T}}\mathbf{x})\} + (\epsilon^2 - 1)\left\{\mathbf{b}(\mathbf{x}) - E(\mathbf{b} \mid \boldsymbol{\beta}^{\mathrm{T}}\mathbf{x})\right\} : \mathbf{a}, \mathbf{b} \in \mathbb{R}^{d_t}].$$

To further calculate the efficient score, we first need the score function, which can be easily verified to be

$$\mathbf{S}_\theta = \begin{bmatrix} -\text{vecm}\left\{\frac{\eta'_{2\epsilon}}{\eta_2}\mathbf{x} \otimes \frac{m'(\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{x})^{\mathrm{T}}}{\sigma(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x})}\right\} \\ -\text{vecl}\left\{\left(\frac{\eta'_{2\epsilon}}{\eta_2}\epsilon + 1\right)\mathbf{x} \otimes \frac{\sigma'(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x})^{\mathrm{T}}}{\sigma(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x})}\right\} \end{bmatrix},$$

where $\eta'_{2\epsilon} = \partial\eta_2/\partial\epsilon$. $\mathbf{S}_\theta$ can be further decomposed as

$$\begin{bmatrix} \left\{-\frac{\eta'_{2\epsilon}}{\eta_2} - \epsilon + c(\mathbf{x})E(\epsilon^3 \mid \mathbf{x})u\right\}\text{vecm}\left\{\mathbf{x} \otimes \frac{m'(\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{x})^{\mathrm{T}}}{\sigma(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x})}\right\} \\ -\left\{\epsilon\frac{\eta'_{2\epsilon}}{\eta_2} + 1 + 2c(\mathbf{x})u\right\}\text{vecl}\left\{\mathbf{x} \otimes \frac{\sigma'(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x})^{\mathrm{T}}}{\sigma(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x})}\right\} \end{bmatrix} + \mathbf{S}_1 \quad (A.1)$$

where

$$\mathbf{S}_1 = \begin{bmatrix} \{\epsilon - uc(\mathbf{x})E(\epsilon^3 \mid \mathbf{x})\}\text{vecm}\left\{\mathbf{x} \otimes \frac{m'(\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{x})^{\mathrm{T}}}{\sigma(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x})}\right\} \\ 2uc(\mathbf{x})\text{vecl}\left\{\mathbf{x} \otimes \frac{\sigma'(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x})^{\mathrm{T}}}{\sigma(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x})}\right\} \end{bmatrix}.$$

It is easy to verify that the first summand in (A.1) is an element of the nuisance tangent space Λ, hence to obtain the efficient score, we only need to further study the second summand $\mathbf{S}_1$. The essential work is to decompose $\mathbf{S}_1$ into an element in Λ and an element in $\Lambda^\perp$, taking advantage of the known form of these two spaces. We skip the tedious derivation procedure, and point out that it is easy to verify $uc\mathbf{b} + (\epsilon - uc\mu_3)\sigma^{-1}\mathbf{a}$ is an element of Λ and $\mathbf{S}_1 - uc\mathbf{b} - (\epsilon - uc\mu_3)\sigma^{-1}\mathbf{a}$ is an element of $\Lambda^\perp$, hence the efficient score is $\mathbf{S}_{\text{eff}} = \mathbf{S}_1 - uc\mathbf{b} - (\epsilon - uc\mu_3)\sigma^{-1}\mathbf{a}$, which has the desired form.

2. The derivation of the efficient score $\mathrm{S}_{\mathrm{eff}}$ in model (4.1)

The joint density function of $\mathbf{x}, Y$ is

$$f_{\mathbf{x},Y}(\mathbf{x}, Y) = \eta_1(\mathbf{x})\eta_2(\epsilon, \mathbf{x})/\sigma(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x})$$

where $\epsilon = \{Y - m(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x})\}/\sigma(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x})$, $\int \eta_1(\mathbf{x})d\mu(\mathbf{x}) = 1$, $\int \eta_2(\epsilon, \mathbf{x})d\mu(\epsilon) = 1$, $\int \epsilon\eta_2(\epsilon, \mathbf{x})d\mu(\epsilon) = 0$, $\int \epsilon^2\eta_2(\epsilon, \mathbf{x})d\mu(\epsilon) = 1$. Like before, the derivation is split into three steps. We first derive the nuisance tangent space Λ, we then derive its orthogonal complement $\Lambda^{\perp}$, and finally, we calculate the score function and project it onto $\Lambda^{\perp}$ to obtain the efficient score.

Calculating the nuisance tangent space with respect to η_1, η_2, we have

$$\begin{aligned}
\Lambda_1 &= \{\mathbf{g}(\mathbf{x}) : E(\mathbf{g}) = \mathbf{0}, \mathbf{g} \in \mathbb{R}^{d_t}\}, \\
\Lambda_2 &= \{\mathbf{f}(\epsilon, \mathbf{x}) : E(\mathbf{f} \mid \mathbf{x}) = \mathbf{0}, E(\epsilon\mathbf{f} \mid \mathbf{x}) = \mathbf{0}, E(\epsilon^2\mathbf{f} \mid \mathbf{x}) = \mathbf{0}, \mathbf{f} \in \mathbb{R}^{d_t}\} \\
&= \{\mathbf{f}(\epsilon, \mathbf{x}) : E(\mathbf{f} \mid \mathbf{x}) = \mathbf{0}, E(\epsilon\mathbf{f} \mid \mathbf{x}) = \mathbf{0}, E(u\mathbf{f} \mid \mathbf{x}) = \mathbf{0}, \mathbf{f} \in \mathbb{R}^{d_t}\}.
\end{aligned}$$

where c, u are defined in (2.1). Note that we still have $E(u \mid \mathbf{x}) = 0$, $E(\epsilon u \mid \mathbf{x}) = 0$, $E(u^2 \mid \mathbf{x}) = 1$. Calculating the nuisance tangent space with respect to m and σ, we have

$$\Lambda_m = \{\eta_{2\epsilon}'/\eta_2\mathbf{a}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}) : \mathbf{a} \in \mathbb{R}^{d_t}\}$$

$$\Lambda_\sigma \;=\; \{(\epsilon\eta_{2\epsilon}'/\eta_2+1)\mathbf{b}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}):\mathbf{b}\in\mathbb{R}^{d_t}\}.$$

Thus, we again have $\Lambda = \Lambda_1 \oplus (\Lambda_2 + \Lambda_m + \Lambda_\sigma)$. As in Appendix 1, we still have $E(\eta_{2\epsilon}'/\eta_2 \mid \mathbf{x}) = 0$, $E(\eta_{2\epsilon}'/\eta_2\epsilon \mid \mathbf{x}) = -1$, $E(\eta_{2\epsilon}'/\eta_2\epsilon^2 \mid \mathbf{x}) = 0$, $E(\eta_{2\epsilon}'/\eta_2\epsilon^3 \mid \mathbf{x}) = -3$, so $E(\eta_{2\epsilon}'/\eta_2 u \mid \mathbf{x}) = c(\mathbf{x})E(\epsilon^3 \mid \mathbf{x})$, $E(\eta_{2\epsilon}'/\eta_2 u\epsilon \mid \mathbf{x}) = -2c(\mathbf{x})$. In order to calculating the residual of projecting any function in $\Lambda_m, \Lambda_\sigma$ to Λ_2, we obtain the decomposition

$$\frac{\eta_{2\epsilon}'}{\eta_2}\mathbf{a}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}) = \left\{\frac{\eta_{2\epsilon}'}{\eta_2} + \epsilon - c(\mathbf{x})E(\epsilon^3 \mid \mathbf{x})u\right\}\mathbf{a}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}) - \left\{\epsilon - uc(\mathbf{x})E(\epsilon^3 \mid \mathbf{x})\right\}\mathbf{a}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}),$$
$$\left(\epsilon\frac{\eta_{2\epsilon}'}{\eta_2} + 1\right)\mathbf{b}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}) = \left\{\epsilon\frac{\eta_{2\epsilon}'}{\eta_2} + 1 + 2c(\mathbf{x})u\right\}\mathbf{b}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}) - 2uc(\mathbf{x})\mathbf{b}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}),$$

where the first summand on the right side of each display is an element in Λ_2, while the second summand is orthogonal to Λ_2. Hence

$$\begin{aligned}
\Lambda_\sigma' &= \Pi(\Lambda_\sigma \mid \Lambda_2^\perp) = \left\{uc(\mathbf{x})\mathbf{b}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}) : \mathbf{b}\in\mathbb{R}^{d_t}\right\},\\
\Lambda_m' &= \Pi(\Lambda_m \mid \Lambda_2^\perp) = \left[\left\{\epsilon - uc(\mathbf{x})E(\epsilon^3 \mid \mathbf{x})\right\}\mathbf{a}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}) : \mathbf{a}\in\mathbb{R}^{d_t}\right],\\
\Lambda_m' + \Lambda_\sigma' &= \left[uc(\mathbf{x})\mathbf{b}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}) + \left\{\epsilon - uc(\mathbf{x})E(\epsilon^3 \mid \mathbf{x})\right\}\mathbf{a}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}) : \mathbf{a},\mathbf{b}\in\mathbb{R}^{d_t}\right].
\end{aligned}$$

We therefore have obtained

$$\Lambda \;=\; \Lambda_1 \oplus \Lambda_2 \oplus (\Lambda_\sigma' + \Lambda_m')$$

$$= \{\mathbf{g}(\mathbf{x}) : E(\mathbf{g}) = \mathbf{0}, \mathbf{g} \in \mathbb{R}^{d_t}\} \oplus \{\mathbf{f}(\epsilon, \mathbf{x}) : E(\mathbf{f} \mid \mathbf{x}) = \mathbf{0}, E(\epsilon\mathbf{f} \mid \mathbf{x}) = \mathbf{0}, E(u\mathbf{f} \mid \mathbf{x}) = \mathbf{0}, \mathbf{f} \in \mathbb{R}^{d_t}\}$$
$$\oplus \left[uc(\mathbf{x})\mathbf{b}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}) + \left\{\epsilon - uc(\mathbf{x})E(\epsilon^3 \mid \mathbf{x})\right\}\mathbf{a}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}) : \mathbf{a}, \mathbf{b} \in \mathbb{R}^{d_t}\right].$$

We can now easily obtain the orthogonal complement of Λ by sequentially considering the orthogonal complement of Λ_1, $\Lambda_1 \oplus \Lambda_2$ and Λ, and obtain

$$\Lambda^{\perp} = [\epsilon\{\mathbf{a}(\mathbf{x}) - E(\mathbf{a} \mid \boldsymbol{\beta}^{\mathrm{T}}\mathbf{x})\} + (\epsilon^2 - 1)\left\{\mathbf{b}(\mathbf{x}) - E(\mathbf{b} \mid \boldsymbol{\beta}^{\mathrm{T}}\mathbf{x})\right\} : \mathbf{a}, \mathbf{b} \in \mathbb{R}^{d_t}].$$

To further derive the efficient score, we first calculate the score function

$$\begin{aligned}\mathbf{S}_{\beta} &= -\frac{\eta_{2\epsilon}'}{\eta_2}\mathrm{vecl}\left\{\mathbf{x} \otimes \frac{m'(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x})^{\mathrm{T}}}{\sigma(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x})}\right\} - \left(\frac{\eta_{2\epsilon}'}{\eta_2}\epsilon + 1\right)\mathrm{vecl}\left\{\mathbf{x} \otimes \frac{\sigma'(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x})^{\mathrm{T}}}{\sigma(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x})}\right\} \\ &= \left\{-\frac{\eta_{2\epsilon}'}{\eta_2} - \epsilon + c(\mathbf{x})E(\epsilon^3 \mid \mathbf{x})u\right\}\mathrm{vecl}\left\{\mathbf{x} \otimes \frac{m'(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x})^{\mathrm{T}}}{\sigma(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x})}\right\} \\ &\quad - \left\{\epsilon\frac{\eta_{2\epsilon}'}{\eta_2} + 1 + 2c(\mathbf{x})u\right\}\mathrm{vecl}\left\{\mathbf{x} \otimes \frac{\sigma'(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x})^{\mathrm{T}}}{\sigma(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x})}\right\} + \mathbf{S}_1,\end{aligned}$$

where

$$\mathbf{S}_1 = \left\{\epsilon - uc(\mathbf{x})E(\epsilon^3 \mid \mathbf{x})\right\}\mathrm{vecl}\left\{\mathbf{x} \otimes \frac{m'(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x})^{\mathrm{T}}}{\sigma(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x})}\right\} + 2uc(\mathbf{x})\mathrm{vecl}\left\{\mathbf{x} \otimes \frac{\sigma'(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x})^{\mathrm{T}}}{\sigma(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x})}\right\}.$$

Using the form of Λ, we can easily verify that $\mathbf{S}_{\beta} - \mathbf{S}_1 \in \Lambda$, hence to obtain $\mathbf{S}_{\mathrm{eff}}$, we only need to further study $\mathbf{S}_1$. Again, using the form of $\Lambda, \Lambda^{\perp}$, we can verify that

$$uc\left\{\frac{2\mathrm{vecl}(\mathbf{x}\otimes\sigma'^{\mathrm{T}})}{\sigma}-\mathbf{a}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x})\right\}+(\epsilon-uc\mu_3)\left\{\frac{\mathrm{vecl}(\mathbf{x}\otimes m'^{\mathrm{T}})}{\sigma}-\mathbf{b}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x})\right\}$$

is an element of $\Lambda^{\perp}$, while its difference from $\mathbf{S}_1$ is an element in Λ. Hence

$$\mathbf{S}_{\mathrm{eff}}=uc\left\{\frac{2\mathrm{vecl}(\mathbf{x}\otimes\sigma'^{\mathrm{T}})}{\sigma}-\mathbf{a}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x})\right\}+(\epsilon-uc\mu_3)\left\{\frac{\mathrm{vecl}(\mathbf{x}\otimes m'^{\mathrm{T}})}{\sigma}-\mathbf{b}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x})\right\}.$$

It is easy to check that this yields the desired form of $\mathbf{S}_{\mathrm{eff}}$ in Section 4.

3. Regularity conditions

(C1) The density functions of $(\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{x})$ and $(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x})$, denoted by $f_{\boldsymbol{\alpha}}(\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{x})$ and $f_{\boldsymbol{\beta}}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x})$, are continuous and bounded away from zero and infinity for all $\mathbf{x}\in\mathbb{R}^p$, and have locally Lipschitz second derivatives.

(C2) The functions $m(\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{x})$, $m(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x})$, $\sigma^2(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x})$, $E(\mathbf{x}\mid\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{x})f_{\boldsymbol{\alpha}}(\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{x})$, $E(\mathbf{x}\mid\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x})f_{\boldsymbol{\beta}}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x})$, $m(\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{x})f_{\boldsymbol{\alpha}}(\boldsymbol{\alpha}^{\mathrm{T}}\mathbf{x})$, $m(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x})f_{\boldsymbol{\beta}}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x})$, and $\sigma^2(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x})f_{\boldsymbol{\beta}}(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x})$ are continuous and bounded for all $\mathbf{x}\in\mathbb{R}^p$, and their third derivatives are locally Lipschitz.

(C3) The mth order kernel function $K(\cdot)$ is twice continuously differentiable with compact support, and is Lipschitz continuous. For $d_\alpha>1$ or $d_\beta>1$, we use multivariate kernel function which is the product of d_α

or d_β univariate kernel functions.

(C4) The bandwidths h_0, h_1, h_2, and h_3 satisfy $nh_i^{2d_\alpha}/\log^2(n) \to \infty$, $nh_i^{2d_\beta}/\log^2(n) \to \infty$ and $nh_i^{4m} \to 0$, for $i = 0, 1, 2$ and 3. In addition, $h_i^4 \log^2 n/h_j \to 0$, $\log^4 n/(nh_ih_j) \to 0$ for $i \neq j$.

(C5) $E(Y^4) < \infty$ and $\max E(X_k^4) < \infty$, for $k = 1, \ldots, p$. In addition, the variance function $\text{var}(Y \mid \mathbf{x})$ is bounded away from 0 and infinity.

4. Outline proof of Theorems 1 and 2

The proof of Theorem 1 follows the same line as that of Theorem 2 but is more tedious, so we only outline the proof of Theorem 2.

We will repetitively use Lemmas 3 and 4 in the supplement of Ma and Zhu (2012). Observe that there are two summands in the estimating equations (4.2). We treat them separately. We first decompose

$$\sum_{i=1}^n \frac{\{\widehat{\varepsilon}_i^2 - \widehat{\sigma}^2(\widehat{\boldsymbol\beta}^{\mathrm T}\mathbf{x}_i)\}}{\{\widehat{\sigma}^2(\widehat{\boldsymbol\beta}^{\mathrm T}\mathbf{x}_i)\}^2}\text{vecl}\left[\left\{\mathbf{x}_i - \widehat{E}(\mathbf{x}_i \mid \widehat{\boldsymbol\beta}^{\mathrm T}\mathbf{x}_i)\right\} \otimes \{\widehat{\sigma}^2(\widehat{\boldsymbol\beta}^{\mathrm T}\mathbf{x}_i)\}'^{\mathrm T}\right]$$

into a summation of the following five summands, denoted $J_1, \ldots, J_5$ respectively.

$$J_1 \stackrel{\text{def}}{=} \sum_{i=1}^n \frac{\{\varepsilon_i^2 - \sigma^2(\boldsymbol\beta^{\mathrm T}\mathbf{x}_i)\}}{\{\sigma^2(\boldsymbol\beta^{\mathrm T}\mathbf{x}_i)\}^2}\text{vecl}\left[\{\mathbf{x}_i - E(\mathbf{x}_i \mid \boldsymbol\beta^{\mathrm T}\mathbf{x}_i)\} \otimes \{\sigma^2(\boldsymbol\beta^{\mathrm T}\mathbf{x}_i)\}'^{\mathrm T}\right]$$

is a summation of independent and identically distributed random vectors.

It is clearly of order $O_p(n^{1/2})$.

$$J_2 \stackrel{\text{def}}{=} \sum_{i=1}^{n} \frac{\{\varepsilon_i^2 - \sigma^2(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}_i)\}}{\{\sigma^2(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}_i)\}^2}\mathrm{vecl}\left[\left\{E(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}_i) - \widehat{E}(\mathbf{x}_i \mid \widehat{\boldsymbol{\beta}}^{\mathrm{T}}\mathbf{x}_i)\right\} \otimes \{\sigma^2(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}_i)\}'^{\mathrm{T}}\right].$$

Lemmas 3 and 4 in the supplement of Ma and Zhu (2012) yields $J_2 = o_p(n^{1/2})$.

$$J_3 \stackrel{\text{def}}{=} \sum_{i=1}^{n} \frac{(\widehat{\varepsilon}_i^2 - \varepsilon_i^2)}{\{\sigma^2(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}_i)\}^2}\mathrm{vecl}\left[\left\{\mathbf{x}_i - \widehat{E}(\mathbf{x}_i \mid \widehat{\boldsymbol{\beta}}^{\mathrm{T}}\mathbf{x}_i)\right\} \otimes \{\sigma^2(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}_i)\}'^{\mathrm{T}}\right]$$

With the identity $\widehat{\varepsilon}_i^2 - \varepsilon_i^2 = (\widehat{\varepsilon}_i - \varepsilon_i)^2 + 2\varepsilon_i(\widehat{\varepsilon}_i - \varepsilon_i)$, we have

$$\begin{aligned} J_3 &= \sum_{i=1}^{n} \frac{(\widehat{\varepsilon}_i - \varepsilon_i)^2}{\{\sigma^2(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}_i)\}^2}\mathrm{vecl}\left[\left\{\mathbf{x}_i - \widehat{E}(\mathbf{x}_i \mid \widehat{\boldsymbol{\beta}}^{\mathrm{T}}\mathbf{x}_i)\right\} \otimes \{\sigma^2(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}_i)\}'^{\mathrm{T}}\right] \\ &+ \sum_{i=1}^{n} \frac{2\varepsilon_i(\widehat{\varepsilon}_i - \varepsilon_i)}{\{\sigma^2(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}_i)\}^2}\mathrm{vecl}\left[\left\{\mathbf{x}_i - \widehat{E}(\mathbf{x}_i \mid \widehat{\boldsymbol{\beta}}^{\mathrm{T}}\mathbf{x}_i)\right\} \otimes \{\sigma^2(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}_i)\}'^{\mathrm{T}}\right]. \end{aligned}$$

The first summand of J_3 is of order $O_p\left\{(nh_3^{d_\alpha})^{-1}\log^2(n) + h_3^{2m}\right\}O_p(n^{1/2})$, and the second is of order $O_p\left\{(nh_3^{d_\alpha})^{-1/2}\log(n) + h_3^{m}\right\}O_p(n^{1/2})$. Both are of order $o_p(n^{1/2})$.

$$J_4 \stackrel{\text{def}}{=} \sum_{i=1}^{n} \frac{\{\sigma^2(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}_i) - \widehat{\sigma}^2(\widehat{\boldsymbol{\beta}}^{\mathrm{T}}\mathbf{x}_i)\}}{\{\sigma^2(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}_i)\}^2}\mathrm{vecl}\left[\left\{\mathbf{x}_i - \widehat{E}(\mathbf{x}_i \mid \widehat{\boldsymbol{\beta}}^{\mathrm{T}}\mathbf{x}_i)\right\} \otimes \{\sigma^2(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}_i)\}'^{\mathrm{T}}\right].$$

Applying Taylor expansion to $\sigma^2(\widehat{\boldsymbol\beta}^{\rm T}\mathbf{x}_i)$ at around $\boldsymbol\beta$, we obtain

$$
\begin{aligned}
J_4 &= \sum_{i=1}^n \frac{\{\sigma^2(\boldsymbol\beta^{\rm T}\mathbf{x}_i)\}'(\widehat{\boldsymbol\beta}-\boldsymbol\beta)^{\rm T}\mathbf{x}_i}{\{\sigma^2(\boldsymbol\beta^{\rm T}\mathbf{x}_i)\}^2}\mathrm{vecl}\left[\left\{\mathbf{x}_i-\widehat{E}(\mathbf{x}_i \mid \widehat{\boldsymbol\beta}^{\rm T}\mathbf{x}_i)\right\}\otimes\{\sigma^2(\boldsymbol\beta^{\rm T}\mathbf{x}_i)\}'^{\rm T}\right] \\
&= nE\left\{\left(\mathrm{vecl}\left[\left\{\mathbf{x}-E(\mathbf{x}\mid\boldsymbol\beta^{\rm T}\mathbf{x})\right\}\otimes\frac{\{\sigma^2(\boldsymbol\beta^{\rm T}\mathbf{x})\}'^{\rm T}}{\sigma^2(\boldsymbol\beta^{\rm T}\mathbf{x})}\right]\right)^{\otimes 2}\right\}\{\mathrm{vecl}(\widehat{\boldsymbol\beta}-\boldsymbol\beta)\}+o_p(n^{1/2}).
\end{aligned}
$$

$$
J_5 \stackrel{\rm def}{=} \sum_{i=1}^n \frac{\{\sigma^2(\boldsymbol\beta^{\rm T}\mathbf{x}_i)-\widehat\sigma^2(\widehat{\boldsymbol\beta}^{\rm T}\mathbf{x}_i)\}}{\{\sigma^2(\boldsymbol\beta^{\rm T}\mathbf{x}_i)\}^2}\mathrm{vecl}\left[\left\{E(\boldsymbol\beta^{\rm T}\mathbf{x}_i)-\widehat{E}(\mathbf{x}_i\mid\widehat{\boldsymbol\beta}^{\rm T}\mathbf{x}_i)\right\}\otimes\{\sigma^2(\boldsymbol\beta^{\rm T}\mathbf{x}_i)\}'^{\rm T}\right]
$$

This quantity is of order $O_p\left\{(nh_3^{d_\alpha})^{-1}\log^2(n)+h_3^{2m}\right\}O_p(n)$, which is of order $o_p(n^{1/2})$ as long as $nh_3^{2d_\alpha}/\log^2(n)\to\infty$ and $nh_3^{4m}\to 0$. Through summarizing the above derivations, we obtain that

$$
\begin{aligned}
&\sum_{i=1}^n \frac{\{\widehat\varepsilon_i^2-\widehat\sigma^2(\widehat{\boldsymbol\beta}^{\rm T}\mathbf{x}_i)\}}{\{\widehat\sigma^2(\widehat{\boldsymbol\beta}^{\rm T}\mathbf{x}_i)\}^2}\mathrm{vecl}\left[\left\{\mathbf{x}_i-\widehat{E}(\mathbf{x}_i\mid\widehat{\boldsymbol\beta}^{\rm T}\mathbf{x}_i)\right\}\otimes\{\widehat\sigma^2(\widehat{\boldsymbol\beta}^{\rm T}\mathbf{x}_i)\}'^{\rm T}\right] \\
=&\sum_{i=1}^n \frac{\{\varepsilon_i^2-\sigma^2(\boldsymbol\beta^{\rm T}\mathbf{x}_i)\}}{\{\sigma^2(\boldsymbol\beta^{\rm T}\mathbf{x}_i)\}^2}\mathrm{vecl}\left[\left\{\mathbf{x}_i-E(\mathbf{x}_i\mid\boldsymbol\beta^{\rm T}\mathbf{x}_i)\right\}\otimes\{\sigma^2(\boldsymbol\beta^{\rm T}\mathbf{x}_i)\}'^{\rm T}\right] \qquad \text{(A.1)}\\
+&\; nE\left\{\left(\mathrm{vecl}\left[\left\{\mathbf{x}-E(\mathbf{x}\mid\boldsymbol\beta^{\rm T}\mathbf{x})\right\}\otimes\frac{\{\sigma^2(\boldsymbol\beta^{\rm T}\mathbf{x})\}'^{\rm T}}{\sigma^2(\boldsymbol\beta^{\rm T}\mathbf{x})}\right]\right)^{\otimes 2}\right\}\{\mathrm{vecl}(\widehat{\boldsymbol\beta}-\boldsymbol\beta)\}+o_p(n^{1/2}).
\end{aligned}
$$

Similarly, we can show that

$$
\sum_{i=1}^n \frac{\widehat\varepsilon_i}{\widehat\sigma^2(\widehat{\boldsymbol\beta}^{\rm T}\mathbf{x}_i)}\mathrm{vecl}\left[\left\{\mathbf{x}_i-\widehat{E}(\mathbf{x}_i\mid\widehat{\boldsymbol\beta}^{\rm T}\mathbf{x}_i)\right\}\otimes\widehat{m}'(\widehat{\boldsymbol\beta}^{\rm T}\mathbf{x}_i)^{\rm T}\right]
$$

$$
\begin{aligned}
&= \sum_{i=1}^{n} \frac{\varepsilon_i}{\sigma^2(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}_i)} \mathrm{vecl}\left[\left\{\mathbf{x}_i - E(\mathbf{x}_i \mid \boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}_i)\right\} \otimes m'(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x}_i)^{\mathrm{T}}\right] \qquad \text{(A.2)} \\
&+ nE\left\{\left(\mathrm{vecl}\left[\left\{\mathbf{x} - E(\mathbf{x} \mid \boldsymbol{\beta}^{\mathrm{T}}\mathbf{x})\right\} \otimes \frac{m'(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x})}{\sigma(\boldsymbol{\beta}^{\mathrm{T}}\mathbf{x})}^{\mathrm{T}}\right]\right)^{\otimes 2}\right\}\{\mathrm{vecl}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})\} + o_p(n^{1/2}).
\end{aligned}
$$

The proof of Theorem 2 is completed by combining (A.1) and (A.2). □

5. Additional simulation results

Table S1: The bias ("bias") and the sample standard errors ("std") for our local and oracle efficient estimating equations estimators (EEE), and the inference results, respectively the average of the estimated standard deviation ("$\widehat{\text{std}}$") and the coverage of the estimated 95% confidence interval ("cp"), of our proposals. All numbers reported below are multiplied by 100.

		$\alpha_{1,3}$	$\alpha_{1,4}$	$\alpha_{1,5}$	$\alpha_{1,6}$	$\alpha_{2,3}$	$\alpha_{2,4}$	$\alpha_{2,5}$	$\alpha_{2,6}$
	true	-0.20	-0.20	0.20	0.20	-0.50	0.20	-0.20	0.20
EEE(local)	bias	0.10	0.04	-0.04	-0.10	0.14	-0.08	0.14	0.08
	std	2.24	2.16	1.16	1.02	2.39	2.27	0.94	0.93
	$\widehat{\text{std}}$	2.25	2.18	1.01	0.93	2.32	2.14	0.99	0.93
	cp	95.10	94.70	93.60	94.90	95.00	94.70	95.40	95.30
EEE(oracle)	bias	0.10	0.04	-0.04	-0.10	0.14	-0.08	0.14	0.08
	std	2.24	2.16	1.16	1.02	2.39	2.27	0.94	0.93
	$\widehat{\text{std}}$	2.25	2.18	1.01	0.93	2.32	2.14	0.99	0.93
	cp	95.10	94.70	93.60	94.90	95.00	94.70	95.40	95.30
		$\beta_{1,3}$	$\beta_{1,4}$	$\beta_{1,5}$	$\beta_{1,6}$	$\beta_{2,3}$	$\beta_{2,4}$	$\beta_{2,5}$	$\beta_{2,6}$
	true	-0.50	-0.20	-0.50	-0.20	-0.20	-0.50	-0.20	-0.50
EEE(local)	bias	-1.57	-0.71	-1.02	-0.57	-1.66	-1.80	-0.91	-1.50
	std	10.52	10.40	5.17	5.63	9.91	9.62	5.71	6.04
	$\widehat{\text{std}}$	13.51	13.46	6.39	6.57	13.34	13.35	6.47	6.50
	cp	96.90	96.90	94.90	95.10	97.30	97.80	94.20	94.60
EEE(oracle)	bias	-0.96	-0.95	-0.08	0.06	-1.73	-1.30	-0.43	-0.38
	std	10.59	11.15	5.67	6.22	11.00	10.46	5.90	6.33
	$\widehat{\text{std}}$	14.72	14.53	6.03	6.07	14.29	14.34	6.04	6.03
	cp	97.00	97.00	91.80	91.60	97.00	97.50	91.50	91.10

Table S2: The bias ("bias") and the sample standard errors ("std") for our local and oracle efficient estimating equations estimators (EEE), and the inference results, respectively the average of the estimated standard deviation ("$\widehat{\text{std}}$") and the coverage of the estimated 95% confidence interval ("cp"), of our proposals. All numbers reported below are multiplied by 100.

		$\beta_{1,3}$	$\beta_{1,4}$	$\beta_{1,5}$	$\beta_{1,6}$	$\beta_{2,3}$	$\beta_{2,4}$	$\beta_{2,5}$	$\beta_{2,6}$
	true	-0.20	-0.20	0.20	0.20	-0.50	0.20	-0.20	0.20
EEE(local)	bias	-0.23	-0.22	0.26	0.22	-0.33	0.16	-0.21	0.08
	std	1.82	1.75	1.73	1.80	2.07	1.72	1.85	1.83
	$\widehat{\text{std}}$	2.05	1.84	1.70	1.84	2.29	1.92	1.91	1.93
	cp	96.90	95.10	93.80	94.80	96.90	96.20	95.20	96.60
EEE(oracle)	bias	-0.10	-0.08	0.09	0.05	-0.12	0.07	-0.10	0.00
	std	1.07	0.99	0.98	1.02	1.11	1.00	1.03	1.03
	$\widehat{\text{std}}$	1.24	1.08	1.01	1.08	1.38	1.15	1.15	1.15
	cp	96.90	96.40	94.30	95.90	98.00	96.60	97.40	97.50

Table S3: The bias ("bias") and the sample standard errors ("std") for the local and oracle efficient estimators obtained from solving (6), and the inference results, respectively the average of the estimated standard deviation ("$\widehat{\text{std}}$") and the coverage of the estimated 95% confidence interval ("cp"), of our proposals. All numbers reported below are multiplied by 100.

		$\beta_{1,3}$	$\beta_{1,4}$	$\beta_{1,5}$	$\beta_{1,6}$	$\beta_{2,3}$	$\beta_{2,4}$	$\beta_{2,5}$	$\beta_{2,6}$
	true	-0.20	-0.20	0.20	0.20	-0.50	0.20	-0.20	0.20
EEE(local)	bias	-0.03	-0.06	0.01	0.07	-0.01	0.00	-0.12	-0.03
	std	1.69	1.44	1.61	1.55	2.04	1.69	1.76	1.77
	$\widehat{\text{std}}$	1.50	1.44	1.41	1.43	1.77	1.64	1.64	1.64
	cp	91.30	92.80	92.30	92.90	93.70	93.90	93.90	95.00
EEE(oracle)	bias	-0.03	-0.06	0.01	0.07	-0.01	0.00	-0.12	-0.03
	std	1.69	1.44	1.61	1.55	2.04	1.69	1.76	1.77
	$\widehat{\text{std}}$	1.50	1.44	1.41	1.43	1.77	1.64	1.64	1.64
	cp	91.30	92.80	92.30	92.90	93.70	93.90	93.90	95.00
		$\beta_{1,3}$	$\beta_{1,4}$	$\beta_{1,5}$	$\beta_{1,6}$	$\beta_{2,3}$	$\beta_{2,4}$	$\beta_{2,5}$	$\beta_{2,6}$
	true	-0.20	-0.20	0.20	0.20	-0.50	0.20	-0.20	0.20
EEE(local)	bias	-0.85	-0.45	0.50	0.79	-1.17	0.33	-0.39	0.77
	std	4.64	4.31	4.54	4.41	4.97	4.24	4.15	4.10
	$\widehat{\text{std}}$	6.12	5.44	5.47	5.37	6.06	5.44	5.42	5.46
	cp	95.90	96.70	96.50	95.90	96.60	96.90	95.80	96.10
EEE(oracle)	bias	-0.45	-0.14	0.19	0.40	-0.48	-0.19	-0.26	0.26
	std	5.21	4.84	5.18	5.13	5.29	4.58	4.62	4.74
	$\widehat{\text{std}}$	6.21	5.45	5.46	5.47	6.11	5.55	5.45	5.39
	cp	94.20	94.20	94.50	94.40	93.40	94.50	95.00	94.20