A Semiparametric Approach to Dimension Reduction: Supplementary Materials

Yanyuan MA and Liping ZHU

1 Statement and Proof of Lemma 1

<u>Lemma</u> **1.** *Assume* $\mathbf{\Lambda}$ *is a* $p \times p$ *symmetric matrix of rank* d*. If and only if* $\boldsymbol{\beta}$ *satisfies*

$$\mathbf{\Lambda} - \mathbf{P}\mathbf{\Lambda}\mathbf{P} = \mathbf{0},$$

then the span of the columns in $\boldsymbol{\beta}$ *is the eigen-space of* $\mathbf{\Lambda}$ *corresponding to the* d *nonzero eigenvalues.*

Proof: First we assume $\mathbf{\Lambda} - \mathbf{P}\mathbf{\Lambda}\mathbf{P} = \mathbf{0}$. By the definition of $\mathbf{P}$, we have

$$\mathbf{\Lambda} = \boldsymbol{\beta}_0(\boldsymbol{\beta}_0^{\mathrm{T}}\boldsymbol{\beta}_0)^{-1}\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{\Lambda}\boldsymbol{\beta}_0(\boldsymbol{\beta}_0^{\mathrm{T}}\boldsymbol{\beta}_0)^{-1}\boldsymbol{\beta}_0^{\mathrm{T}}.$$

Thus, $\mathbf{\Lambda}\boldsymbol{\beta}_0 = \boldsymbol{\beta}_0(\boldsymbol{\beta}_0^{\mathrm{T}}\boldsymbol{\beta}_0)^{-1}\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{\Lambda}\boldsymbol{\beta}_0$. Denote $\mathbf{A} = (\boldsymbol{\beta}_0^{\mathrm{T}}\boldsymbol{\beta}_0)^{-1}\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{\Lambda}\boldsymbol{\beta}_0$. $\mathbf{A}$ is a $d \times d$ full rank matrix. Thus, we can write the eigen-decomposition of $\mathbf{A}$ as $\mathbf{AU} = \mathbf{UD}$, where $\mathbf{D}$ is a diagonal matrix. Thus we have $\mathbf{\Lambda}\boldsymbol{\beta}_0\mathbf{U} = \boldsymbol{\beta}_0\mathbf{AU} = \boldsymbol{\beta}_0\mathbf{UD}$, indicating that $\boldsymbol{\beta}_0\mathbf{U}$ is a matrix of d eigenvectors of $\mathbf{\Lambda}$ corresponding to the d non-zero eigenvalues. Thus, the column space of $\boldsymbol{\beta}_0$ is indeed the d dimensional eigen-space corresponding to the d nonzero eigenvalues of $\mathbf{\Lambda}$.

On the other hand, assume now the column space of $\boldsymbol{\beta}_0$ is the d dimensional eigen-space corresponding to the d nonzero eigenvalues of $\mathbf{\Lambda}$. Denote $\mathbf{b}$ the $p \times d$ matrix formed by the d eigenvectors of $\mathbf{\Lambda}$ corresponding to the d non-zero eigenvalues of $\mathbf{\Lambda}$. Denote $\mathbf{c}$ the $p \times (p-d)$ matrix formed by the $p-d$ eigenvectors of $\mathbf{\Lambda}$ corresponding to the remaining zero eigenvalues. Because $\mathbf{\Lambda}$ is symmetric, we can assume $(\mathbf{b}, \mathbf{c})$ to be an orthonormal matrix. This means $\mathbf{b}^{\mathrm{T}}\mathbf{b} = \mathbf{I}_d$ and $\mathbf{b}^{\mathrm{T}}\mathbf{c} = \mathbf{0}$. We now have $\mathbf{\Lambda}\mathbf{b} = \mathbf{bD}$ and $\mathbf{\Lambda}\mathbf{c} = \mathbf{0}$, where $\mathbf{D}$ is the diagonal matrix with the d nonzero eigenvalues on the diagonal. In addition, we can also find a full rank $d \times d$ matrix $\mathbf{U}$ so that $\boldsymbol{\beta}_0 = \mathbf{bU}$. We have

$$\begin{aligned}
\mathbf{P\Lambda P} - \mathbf{\Lambda} &= \boldsymbol{\beta}_0(\boldsymbol{\beta}_0^{\mathrm{T}}\boldsymbol{\beta}_0)^{-1}\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{\Lambda}\mathbf{b}\mathbf{U}(\boldsymbol{\beta}_0^{\mathrm{T}}\boldsymbol{\beta}_0)^{-1}\boldsymbol{\beta}_0^{\mathrm{T}} - \mathbf{\Lambda} \\
&= \boldsymbol{\beta}_0(\boldsymbol{\beta}_0^{\mathrm{T}}\boldsymbol{\beta}_0)^{-1}\boldsymbol{\beta}_0^{\mathrm{T}}\mathbf{b}\mathbf{D}\mathbf{U}(\boldsymbol{\beta}_0^{\mathrm{T}}\boldsymbol{\beta}_0)^{-1}\boldsymbol{\beta}_0^{\mathrm{T}} - \mathbf{\Lambda} \\
&= \boldsymbol{\beta}_0\mathbf{U}^{-1}\mathbf{D}\mathbf{U}(\boldsymbol{\beta}_0^{\mathrm{T}}\boldsymbol{\beta}_0)^{-1}\boldsymbol{\beta}_0^{\mathrm{T}} - \mathbf{\Lambda} \\
&= \mathbf{\Lambda}\mathbf{b}(\mathbf{b}^{\mathrm{T}}\mathbf{b})^{-1}\mathbf{b}^{\mathrm{T}} - \mathbf{\Lambda}.
\end{aligned}$$

Thus, we can obtain $(\mathbf{P\Lambda P} - \mathbf{\Lambda})\mathbf{b} = \mathbf{0}$ and $(\mathbf{P\Lambda P} - \mathbf{\Lambda})\mathbf{c} = \mathbf{0} - \mathbf{0} = \mathbf{0}$. Because $(\mathbf{b}, \mathbf{c})$ is orthonormal, hence $\mathbf{P\Lambda P} - \mathbf{\Lambda} = \mathbf{0}$. □

2 Statement and Proof of Lemma 2

Lemma 2. *Assume $\mathbf{\Lambda}$ is a $p \times p$ symmetric non-negative definite matrix of rank d. If and only if $\boldsymbol{\beta}$ satisfies*

$$\mathbf{Q\Lambda Q} = \mathbf{0},$$

then the span of the columns in $\boldsymbol{\beta}$ is the eigen-space of $\mathbf{\Lambda}$ corresponding to the d nonzero eigenvalues.

Proof: We first show that assume $\mathbf{\Lambda}$ is a $p \times p$ non-negative definite symmetric matrix of rank d, then $\mathbf{\Lambda} = \mathbf{P\Lambda P}$ if and only if $\mathbf{Q\Lambda Q} = \mathbf{0}$.

Assume $\mathbf{\Lambda} = \mathbf{P\Lambda P}$, then $\mathbf{\Lambda P} = \mathbf{P\Lambda P}^2 = \mathbf{P\Lambda P}$ and $\mathbf{P\Lambda} = \mathbf{P}^2\mathbf{\Lambda P} = \mathbf{P\Lambda P}$. So we have $\mathbf{Q\Lambda Q} = \mathbf{\Lambda} - \mathbf{P\Lambda} - \mathbf{\Lambda P} + \mathbf{P\Lambda P} = \mathbf{\Lambda} - \mathbf{P\Lambda P} = \mathbf{0}$.

On the other hand, assume $\mathbf{Q\Lambda Q} = \mathbf{0}$. Because $\mathbf{\Lambda}$ is non-negative definite, hence $\mathbf{\Lambda Q} = \mathbf{0}$ and $\mathbf{Q\Lambda} = \mathbf{0}$. Thus, $\mathbf{P\Lambda P} = (\mathbf{I}_p - \mathbf{Q})\mathbf{\Lambda}(\mathbf{I}_p - \mathbf{Q}) = \mathbf{\Lambda} - \mathbf{Q\Lambda} - \mathbf{\Lambda Q} + \mathbf{Q\Lambda Q} = \mathbf{\Lambda}$.

Lemma 2 is a direct consequence of Lemma 1 and the above result.

3 SIR for Non-Elliptical Predictors

The first attempt to remove the linearity condition within the inverse regression family is Li and Dong (2009). With a slight modification of the classic SIR, they proposed to recover $\mathcal{S}_{Y|\mathbf{x}}$ through minimizing

$$E\left(\left\|E(\mathbf{x} \mid Y) - E\left\{E(\mathbf{x} \mid \mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}) \mid Y\right\}\right\|^2\right),$$

where $\|\cdot\|$ denotes the Frobenius norm of a matrix. The minimizer spans $\mathcal{S}_{Y|\mathbf{x}}$ even when the linearity condition is violated. The above minimization is equivalent to

$$\begin{aligned}
& E\left(\left[E(\mathbf{x}\mid Y)-E\left\{E(\mathbf{x}\mid \mathbf{x}^{\mathrm{T}}\boldsymbol{\beta})\mid Y\right\}\right]^{\mathrm{T}}\frac{\partial E\left\{E(\mathbf{x}\mid \mathbf{x}^{\mathrm{T}}\boldsymbol{\beta})\mid Y\right\}}{\partial\left\{\mathrm{vec}(\boldsymbol{\beta})\right\}^{\mathrm{T}}}\right)\\
= \ & E\left(\left[E(\mathbf{x}\mid Y)-E\left\{E(\mathbf{x}\mid \mathbf{x}^{\mathrm{T}}\boldsymbol{\beta})\mid Y\right\}\right]^{\mathrm{T}} E\left\{\frac{\partial E(\mathbf{x}\mid \mathbf{x}^{\mathrm{T}}\boldsymbol{\beta})}{\partial\left\{\mathrm{vec}(\boldsymbol{\beta})\right\}^{\mathrm{T}}}\mid Y\right\}\right)\\
= \ & \mathbf{0}.
\end{aligned}$$

In obtaining the second equality, we used the relation

$$\frac{\partial E\left\{E(\mathbf{x}\mid \mathbf{x}^{\mathrm{T}}\boldsymbol{\beta})\mid Y\right\}}{\partial\left\{\mathrm{vec}(\boldsymbol{\beta})\right\}^{\mathrm{T}}}=E\left[\frac{\partial E(\mathbf{x}\mid \mathbf{x}^{\mathrm{T}}\boldsymbol{\beta})}{\partial\left\{\mathrm{vec}(\boldsymbol{\beta})\right\}^{\mathrm{T}}}\;\Big|\; Y\right],$$

which holds under mild regularity conditions. With the law of iterated expectations, the above equality can be equivalently written as

$$E\left(\left\{\mathbf{x}-E(\mathbf{x}\mid \mathbf{x}^{\mathrm{T}}\boldsymbol{\beta})\right\}^{\mathrm{T}} E\left[\frac{\partial E(\mathbf{x}\mid \mathbf{x}^{\mathrm{T}}\boldsymbol{\beta})}{\partial\left\{\mathrm{vec}(\boldsymbol{\beta})\right\}^{\mathrm{T}}}\mid Y\right]\right)=\mathbf{0}. \tag{1}$$

To see the above method as a special case of the semiparametric approach, we choose $\boldsymbol{\alpha}(\mathbf{x})=\mathbf{x}-E(\mathbf{x}\mid \mathbf{x}^{\mathrm{T}}\boldsymbol{\beta})$ and $\mathbf{g}(Y,\mathbf{x}^{\mathrm{T}}\boldsymbol{\beta})=E\left[\partial E(\mathbf{x}\mid \mathbf{x}^{\mathrm{T}}\boldsymbol{\beta})/\partial\left\{\mathrm{vec}(\boldsymbol{\beta})\right\}^{\mathrm{T}}\mid Y\right]$ in the estimating equation (3) in the main article. Obviously, $E(\boldsymbol{\alpha}\mid \mathbf{x}^{\mathrm{T}}\boldsymbol{\beta})=\mathbf{0}$. Making use of the double robustness, we mis-specify $E(\mathbf{g}\mid \mathbf{x}^{\mathrm{T}}\boldsymbol{\beta})=0$, which directly yields (1).

4 SAVE for Non-Elliptical Predictors

Dong and Li (2010) further extended the idea of Li and Dong (2009) to second-order methods. In this section we use SAVE as an example to illustrate their rationale.

Assuming the constant variance condition, Dong and Li (2010) used the equality

$$\mathbf{I}_p-\mathrm{cov}(\mathbf{x}\mid Y)=\mathrm{cov}\left\{E(\mathbf{x}\mid \mathbf{x}^{\mathrm{T}}\boldsymbol{\beta})\right\}-\mathrm{cov}\left\{E(\mathbf{x}\mid \mathbf{x}^{\mathrm{T}}\boldsymbol{\beta})\mid Y\right\}$$

and proposed to obtain $\mathcal{S}_{Y|\mathbf{x}}$ based on

$$\min_{\boldsymbol{\beta}} E\left(\left\|\mathbf{I}_p-\mathrm{cov}(\mathbf{x}\mid Y)-\mathrm{cov}\left\{E(\mathbf{x}\mid \mathbf{x}^{\mathrm{T}}\boldsymbol{\beta})\right\}+\mathrm{cov}\left\{E(\mathbf{x}\mid \mathbf{x}^{\mathrm{T}}\boldsymbol{\beta})\mid Y\right\}\right\|^2\right).$$

To simplify notation, we denote $\mathbf{u} = E(\mathbf{x} \mid \mathbf{x}^{\mathrm{T}}\boldsymbol{\beta})$. Similar as in Section 3, an equivalent form of the above is

$$
\begin{aligned}
& E\left[\operatorname{vec}\left\{\mathbf{I}_p - \operatorname{cov}(\mathbf{x} \mid Y) - \operatorname{cov}(\mathbf{u}) + \operatorname{cov}(\mathbf{u} \mid Y)\right\}^{\mathrm{T}} \frac{\partial \operatorname{vec}\left\{-\operatorname{cov}(\mathbf{u}) + \operatorname{cov}(\mathbf{u} \mid Y)\right\}}{\partial \operatorname{vec}(\boldsymbol{\beta})^{\mathrm{T}}}\right] \\
= \; & E\left[\operatorname{vec}\left\{\mathbf{I}_p - \operatorname{cov}(\mathbf{x} \mid Y) - \operatorname{cov}(\mathbf{u}) + \operatorname{cov}(\mathbf{u} \mid Y)\right\}^{\mathrm{T}} \frac{\partial \operatorname{vec}\left\{\operatorname{cov}(\mathbf{u} \mid Y)\right\}}{\partial \operatorname{vec}(\boldsymbol{\beta})^{\mathrm{T}}}\right] \\
= \; & E\Big(\operatorname{vec}\left\{\mathbf{I}_p - \operatorname{cov}(\mathbf{x} \mid Y) - \operatorname{cov}(\mathbf{u}) + \operatorname{cov}(\mathbf{u} \mid Y)\right\}^{\mathrm{T}} \\
& \cdot \left[E\left\{\frac{\partial \operatorname{vec}(\mathbf{u}\mathbf{u}^{\mathrm{T}})}{\partial \operatorname{vec}(\boldsymbol{\beta})^{\mathrm{T}}} \mid Y\right\} - E\left\{\mathbf{u} \otimes \frac{\partial \mathbf{u}}{\partial \operatorname{vec}(\boldsymbol{\beta})^{\mathrm{T}}} - \frac{\partial \mathbf{u}}{\partial \operatorname{vec}(\boldsymbol{\beta})^{\mathrm{T}}} \otimes \mathbf{u} \mid Y\right\}\right]\Big) \\
= \; & E\Big[\operatorname{vec}\left\{\mathbf{I}_p - \operatorname{cov}(\mathbf{x} \mid Y) - \operatorname{cov}(\mathbf{u}) + \operatorname{cov}(\mathbf{u} \mid Y)\right\}^{\mathrm{T}} \\
& \cdot \left\{\frac{\partial \operatorname{vec}(\mathbf{u}\mathbf{u}^{\mathrm{T}})}{\partial \operatorname{vec}(\boldsymbol{\beta})^{\mathrm{T}}} - \mathbf{u} \otimes \frac{\partial \mathbf{u}}{\partial \operatorname{vec}(\boldsymbol{\beta})^{\mathrm{T}}} - \frac{\partial \mathbf{u}}{\partial \operatorname{vec}(\boldsymbol{\beta})^{\mathrm{T}}} \otimes \mathbf{u}\right\}\Big] \\
= \; & \mathbf{0}.
\end{aligned}
$$

Here, the first equality holds because $\partial \operatorname{vec}\{\operatorname{cov}(\mathbf{u})\}/\partial \operatorname{vec}(\boldsymbol{\beta})^{\mathrm{T}}$ is a constant, and $\otimes$ denotes Kronecker product. If we choose

$$
\begin{aligned}
\mathbf{g}(Y, \mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}) &= \operatorname{vec}\left\{\mathbf{I}_p - \operatorname{cov}(\mathbf{x} \mid Y) - \operatorname{cov}(\mathbf{u}) + \operatorname{cov}(\mathbf{u} \mid Y)\right\}^{\mathrm{T}}, \\
\boldsymbol{\alpha}(\mathbf{x}) &= \frac{\partial \operatorname{vec}(\mathbf{u}\mathbf{u}^{\mathrm{T}})}{\partial \operatorname{vec}(\boldsymbol{\beta})^{\mathrm{T}}} - \mathbf{u} \otimes \frac{\partial \mathbf{u}}{\partial \operatorname{vec}(\boldsymbol{\beta})^{\mathrm{T}}} - \frac{\partial \mathbf{u}}{\partial \operatorname{vec}(\boldsymbol{\beta})^{\mathrm{T}}} \otimes \mathbf{u},
\end{aligned}
$$

and mis-specify $E(\boldsymbol{\alpha} \mid \mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}) = \mathbf{0}$ while making use of $E(\mathbf{g} \mid \mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}) = \mathbf{0}$, a direct result from $\mathbf{g} = \mathbf{0}$, we can easily see that this improved version of SAVE is also a special case of the semiparametric approach.

5 Simulations: Example 3

We further illustrate the performance of the semiparametric methods at different mean-to-variance ratios, which we define as $\operatorname{var}\{E(Y \mid \mathbf{x})\}/E\{\operatorname{var}(Y \mid \mathbf{x})\}$. Note that in the problem of estimating $\mathcal{S}_{E(Y|\mathbf{x})}$, this mean-to-variance ratio can be viewed as signal-to-noise ratio. However, in the problem of estimating $\mathcal{S}_{Y|\mathbf{x}}$, the concept of signal or noise no longer applies because all aspects of the model contains useful information. For this reason, we avoid using the notion of signal-to-noise ratio.

We generate the response variable from the following two models:

$$
\begin{aligned}
\text{model (V)}: &\quad Y = \mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}_1 + \sigma_0 \exp\left(\mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}_2\right)\varepsilon, \\
\text{model (VI)}: &\quad Y = 0.5\left(\mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}_1\right)^2 + 4\left(\mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}_2\right)^2 + \sigma_0\varepsilon
\end{aligned}
$$

where $\boldsymbol{\beta}_1$, $\boldsymbol{\beta}_2$, $\mathbf{x}$ and ε are the same as in Example 1. We use different σ_0 values to control the levels of mean-to-variance ratio. Specifically, we set $\sigma_0 = 0.5$, 1 and 2, which roughly corresponds to the respective mean-to-variance ratio of 4, 1 and 0.25 in model (V). The corresponding mean-to-variance ratio in model (VI) is slightly different, with the smallest ratio about 2. We use models (V) and (VI) to compare respectively SIR and PHD with their semiparametric counterparts. Because the simulation results for $p = 6$ are quite similar to those for $p = 12$, we only report the results for $p = 12$ below in Figure 1.

The performance of the various estimators in model (V) does not follow a monotone pattern in general. This reflects the fact that in estimating $\mathcal{S}_{Y|\mathbf{x}}$, larger variability does not necessarily lead to worse performance. How an estimator performs when the mean-to-variance ratio increases depends on how the mean and variance information is utilized in the particular estimator. However, in all the different mean-to-variance ratio situations, the semiparametric estimator out-performs the others. On the contrary, all the estimators show different degrees of improvement when the mean-to-variance ratio increases in model (VI). This is within our expectation because in identifying $\mathcal{S}_{E(Y|\mathbf{x})}$, the only useful information is in the mean function and increasing the error variance truly reflects increased noise. Again, among all the estimators and in all the different noise situations, the semiparametric estimator performs the best.

6 Statement and Proof of Theorem 1

Theorem 1. *Under conditions (C1)-(C4) given in Appendix 4, the estimator $\widehat{\boldsymbol{\beta}}$ obtained from the estimating equation*

$$\sum_{i=1}^{n} \left[\mathbf{g}(Y_i, \mathbf{x}_i^{\mathrm{T}}\widehat{\boldsymbol{\beta}}) - \widehat{E}\{\mathbf{g}(Y_i, \mathbf{x}_i^{\mathrm{T}}\widehat{\boldsymbol{\beta}}) \mid \mathbf{x}_i^{\mathrm{T}}\widehat{\boldsymbol{\beta}}\}\right] \left[\boldsymbol{\alpha}(\mathbf{x}_i) - \widehat{E}\{\boldsymbol{\alpha}(\mathbf{x}_i) \mid \mathbf{x}_i^{\mathrm{T}}\widehat{\boldsymbol{\beta}}\}\right] = \mathbf{0}$$

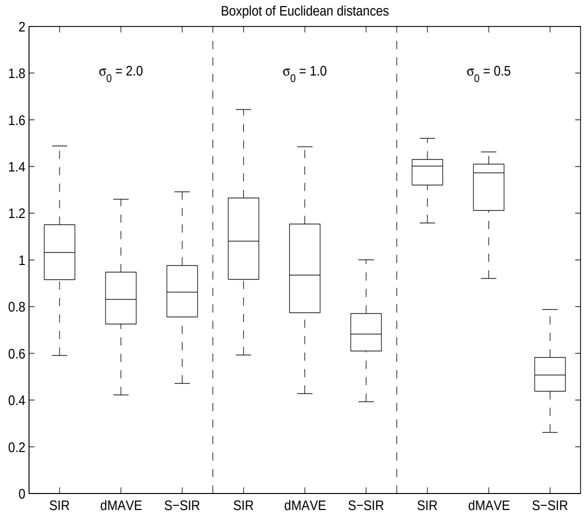
satisfies

$$\sqrt{n}\mathbf{A}\mathrm{vec}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \to \mathcal{N}(\mathbf{0}, \mathbf{B})$$

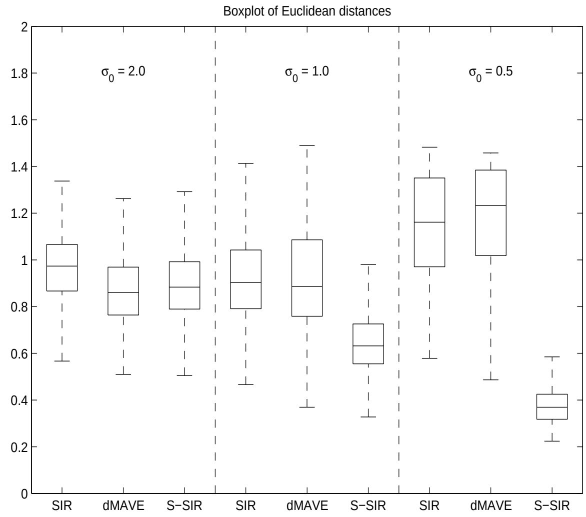
in distribution, where

$$\begin{aligned} \mathbf{A} &= E\left[\partial \mathrm{vec}\left(\left[\mathbf{g}(Y, \mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}) - E\{\mathbf{g}(Y, \mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}) \mid \mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}\}\right]\left[\boldsymbol{\alpha}(\mathbf{x}) - E\{\boldsymbol{\alpha}(\mathbf{x}) \mid \mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}\}\right]\right) / \partial\left\{\mathrm{vec}(\boldsymbol{\beta})\right\}^{\mathrm{T}}\right], \\ \mathbf{B} &= \mathrm{cov}\left\{\mathrm{vec}\left(\left[\mathbf{g}(Y, \mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}) - E\{\mathbf{g}(Y, \mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}) \mid \mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}\}\right]\left[\boldsymbol{\alpha}(\mathbf{x}) - E\{\boldsymbol{\alpha}(\mathbf{x}) \mid \mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}\}\right]\right)\right\}. \end{aligned}$$

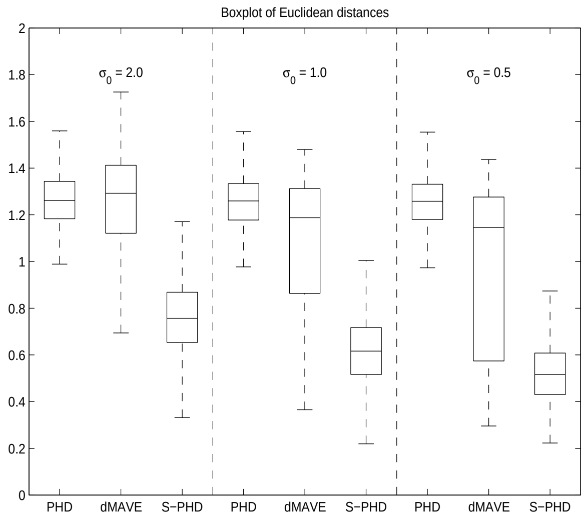
Here $\mathrm{vec}(\mathbf{M})$ *denotes the vector formed by concatenating the columns of* $\mathbf{M}$.



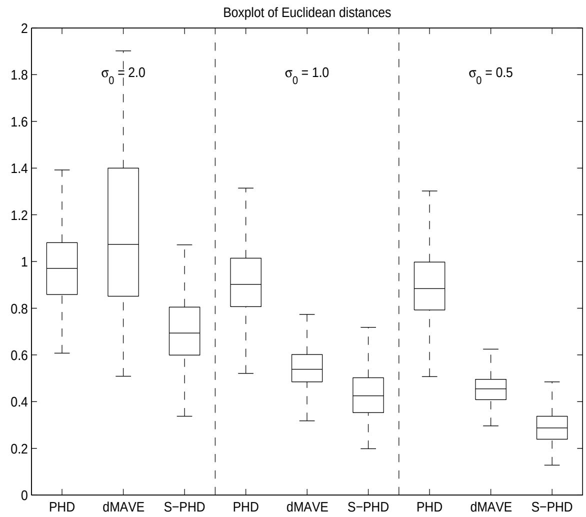
(A): case 1 in model (V)



(B): case 2 in model (V)



(C): case 1 in model (VI)



(D): case 2 in model (VI)

Figure 1: Boxplots of Euclidean distances for models (V) and (VI) with $p = 12$.

Proof: We first prove two lemmas.

<u>**Lemma**</u> **3.** *Assume Conditions (C1)-(C4) hold. Let* $\Omega_\beta = \{(\mathbf{x}, Y, \widehat{\boldsymbol{\beta}}) : \mathbf{x} \in \mathbb{R}^p, Y \in \mathbb{R}, \|\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}\| \leq Cn^{-1/2}\}$*, where* $\|\cdot\|$ *is the Euclidean norm and* C *is a generic constant. Then there*

exists a basis $\boldsymbol\beta$ *of* $\mathcal{S}_{Y|\mathbf{x}}$ *such that*

$$\begin{aligned}
& \sup_{\Omega_\beta}\left|\widehat{E}\{\boldsymbol\alpha(\mathbf{x}) \mid \mathbf{x}^{\mathrm T}\widehat{\boldsymbol\beta}\} - \widehat{E}\{\boldsymbol\alpha(\mathbf{x}) \mid \mathbf{x}^{\mathrm T}\boldsymbol\beta\} - E\{\boldsymbol\alpha(\mathbf{x}) \mid \mathbf{x}^{\mathrm T}\widehat{\boldsymbol\beta}\} + E\{\boldsymbol\alpha(\mathbf{x}) \mid \mathbf{x}^{\mathrm T}\boldsymbol\beta\}\right| \\
= \; & O_p\left\{h^m n^{-1/2} + n^{-1}h^{-(d+1)}\mathrm{log}n\right\}, \qquad (2) \\
\textit{and} \quad & \sup_{\Omega_\beta}\left|\widehat{E}\{\mathbf{g}(Y,\mathbf{x}^{\mathrm T}\widehat{\boldsymbol\beta}) \mid \mathbf{x}^{\mathrm T}\widehat{\boldsymbol\beta}\} - \widehat{E}\{\mathbf{g}(Y,\mathbf{x}^{\mathrm T}\boldsymbol\beta) \mid \mathbf{x}^{\mathrm T}\boldsymbol\beta\} - E\{\mathbf{g}(Y,\mathbf{x}^{\mathrm T}\widehat{\boldsymbol\beta}) \mid \mathbf{x}^{\mathrm T}\widehat{\boldsymbol\beta}\}\right. \\
& \left. + E\{\mathbf{g}(Y,\mathbf{x}^{\mathrm T}\boldsymbol\beta) \mid \mathbf{x}^{\mathrm T}\boldsymbol\beta\}\right| \\
= \; & O_p\left\{h^m n^{-1/2} + n^{-1}h^{-(d+1)}\mathrm{log}n\right\}. \qquad (3)
\end{aligned}$$

Proof for Lemma 3. We only prove (2) since the proof of (3) is almost identical. Recall that we estimate the nonparametric functions with kernel regressions to obtain

$$\widehat{E}\{\boldsymbol\alpha(\mathbf{x}) \mid \mathbf{x}^{\mathrm T}\widehat{\boldsymbol\beta}\} = \frac{1}{n}\sum_{i=1}^n K_h(\mathbf{x}_i^{\mathrm T}\widehat{\boldsymbol\beta} - \mathbf{x}^{\mathrm T}\widehat{\boldsymbol\beta})\boldsymbol\alpha(\mathbf{x}_i)\Big/\frac{1}{n}\sum_{i=1}^n K_h(\mathbf{x}_i^{\mathrm T}\widehat{\boldsymbol\beta} - \mathbf{x}^{\mathrm T}\widehat{\boldsymbol\beta}).$$

In the sequel we inspect the numerator and denominator of $\widehat{E}\{\boldsymbol\alpha(\mathbf{x}) \mid \mathbf{x}^{\mathrm T}\boldsymbol\beta\}$ respectively. Using the law of iterated expectations, we obtain that

$$\begin{aligned}
& E\left[\{K_h(\mathbf{x}_i^{\mathrm T}\widehat{\boldsymbol\beta} - \mathbf{x}^{\mathrm T}\widehat{\boldsymbol\beta}) - K_h(\mathbf{x}_i^{\mathrm T}\boldsymbol\beta - \mathbf{x}^{\mathrm T}\boldsymbol\beta)\}^2\boldsymbol\alpha(\mathbf{x}_i)\boldsymbol\alpha^{\mathrm T}(\mathbf{x}_i)\right] \\
= \; & E\left[\{K_h(\mathbf{x}_i^{\mathrm T}\widehat{\boldsymbol\beta} - \mathbf{x}^{\mathrm T}\widehat{\boldsymbol\beta}) - K_h(\mathbf{x}_i^{\mathrm T}\boldsymbol\beta - \mathbf{x}^{\mathrm T}\boldsymbol\beta)\}^2 E\{\boldsymbol\alpha(\mathbf{x}_i)\boldsymbol\alpha^{\mathrm T}(\mathbf{x}_i) \mid \mathbf{x}_i^{\mathrm T}\boldsymbol\beta, \mathbf{x}_i^{\mathrm T}\widehat{\boldsymbol\beta}\}\right] \\
= \; & E\left[\left\{K_h(\mathbf{x}_i^{\mathrm T}\widehat{\boldsymbol\beta} - \mathbf{x}^{\mathrm T}\widehat{\boldsymbol\beta}) - K_h(\mathbf{x}_i^{\mathrm T}\boldsymbol\beta - \mathbf{x}^{\mathrm T}\boldsymbol\beta)\right\}^2 E\left\{\boldsymbol\alpha(\mathbf{x}_i)\boldsymbol\alpha^{\mathrm T}(\mathbf{x}_i) \mid \mathbf{x}_i^{\mathrm T}\boldsymbol\beta\right\}\right] \\
& + E\left\{K_h(\mathbf{x}_i^{\mathrm T}\widehat{\boldsymbol\beta} - \mathbf{x}^{\mathrm T}\widehat{\boldsymbol\beta}) - K_h(\mathbf{x}_i^{\mathrm T}\boldsymbol\beta - \mathbf{x}^{\mathrm T}\boldsymbol\beta)\right\}^2 \\
& \cdot\left[E\{\boldsymbol\alpha(\mathbf{x}_i)\boldsymbol\alpha^{\mathrm T}(\mathbf{x}_i) \mid \mathbf{x}_i^{\mathrm T}\boldsymbol\beta, \mathbf{x}_i^{\mathrm T}\widehat{\boldsymbol\beta}\} - E\{\boldsymbol\alpha(\mathbf{x}_i)\boldsymbol\alpha^{\mathrm T}(\mathbf{x}_i) \mid \mathbf{x}_i^{\mathrm T}\boldsymbol\beta\}\right]. \qquad (4)
\end{aligned}$$

We bound the first quantity in the last equality of (4) first. Under the condition (C3) that $E\{\boldsymbol\alpha(\mathbf{x}_i)\boldsymbol\alpha^{\mathrm T}(\mathbf{x}_i) \mid \mathbf{x}_i^{\mathrm T}\boldsymbol\beta\} \le C_1$ and $f_{\mathbf{x}}(\mathbf{x}_i) \le C_2$, the first quantity is less than

$$\begin{aligned}
& C_1C_2\int\{K_h(\mathbf{x}_i^{\mathrm T}\widehat{\boldsymbol\beta} - \mathbf{x}^{\mathrm T}\widehat{\boldsymbol\beta}) - K_h(\mathbf{x}_i^{\mathrm T}\boldsymbol\beta - \mathbf{x}^{\mathrm T}\boldsymbol\beta)\}^2 d\mathbf{x}_i \\
= \; & C_1C_2h^{-2d}\int\left\{K\left(\frac{\mathbf{x}_i^{\mathrm T}\widehat{\boldsymbol\beta} - \mathbf{x}^{\mathrm T}\widehat{\boldsymbol\beta}}{h}\right) - K\left(\frac{\mathbf{x}_i^{\mathrm T}\boldsymbol\beta - \mathbf{x}^{\mathrm T}\boldsymbol\beta}{h}\right)\right\}^2 d\mathbf{x}_i.
\end{aligned}$$

By letting $\mathbf{z}_i = (\mathbf{x}_i - \mathbf{x})/h$, the above display is equal to

$$\begin{aligned}
& C_1C_2\int\left\{K_h(\mathbf{x}_i^{\mathrm T}\widehat{\boldsymbol\beta} - \mathbf{x}^{\mathrm T}\widehat{\boldsymbol\beta}) - K_h(\mathbf{x}_i^{\mathrm T}\boldsymbol\beta - \mathbf{x}^{\mathrm T}\boldsymbol\beta)\right\}^2 d\mathbf{x}_i \\
= \; & C_1C_2h^{p-2d}\int\left\{K(\mathbf{z}_i^{\mathrm T}\widehat{\boldsymbol\beta}) - K(\mathbf{z}_i^{\mathrm T}\boldsymbol\beta)\right\}^2 d\mathbf{z}_i. \qquad (5)
\end{aligned}$$

Because condition (C1) assumes that $K(\cdot)$ is Lipschitz continuous and has compact support,

$$\begin{aligned}
& C_1 C_2 h^{-2d} h^p \int \left\{ K(\mathbf{z}_i^{\mathrm{T}} \widehat{\boldsymbol{\beta}}) - K(\mathbf{z}_i^{\mathrm{T}} \boldsymbol{\beta}) \right\}^2 d\mathbf{z}_i \\
\leq \; & C_1 C_2 C_3 h^{-2d} h^p \int \left\| \mathbf{z}_i^{\mathrm{T}} (\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \right\|^2 d\mathbf{z}_i \\
= \; & C_1 C_2 C_3 h^{-(p+2d+2)} h^p \int \left\| (h\mathbf{z}_i)^{\mathrm{T}} (\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \right\|^2 d(h\mathbf{z}_i) \\
= \; & O\left\{ n^{-1} h^{-(2d+2)} \right\}.
\end{aligned}$$

Next we deal with the second quantity in the last equality of (4). Condition (C3) assumes that $E\{\boldsymbol{\alpha}(\mathbf{x}_i)\boldsymbol{\alpha}^{\mathrm{T}}(\mathbf{x}_i) \mid \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}\}$ is locally Lipschitz-continuous. Thus

$$\left| E\{\boldsymbol{\alpha}(\mathbf{x}_i)\boldsymbol{\alpha}^{\mathrm{T}}(\mathbf{x}_i) \mid \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}, \mathbf{x}_i^{\mathrm{T}}\widehat{\boldsymbol{\beta}}\} - E\{\boldsymbol{\alpha}(\mathbf{x}_i)\boldsymbol{\alpha}^{\mathrm{T}}(\mathbf{x}_i) \mid \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}\} \right| \leq C_4 \left| \mathbf{x}_i^{\mathrm{T}} (\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \right|.$$

Accordingly, the second quantity in the last equality of (4) is bounded by

$$C_2 C_4 \int \left\{ K_h(\mathbf{x}_i^{\mathrm{T}}\widehat{\boldsymbol{\beta}} - \mathbf{x}^{\mathrm{T}}\widehat{\boldsymbol{\beta}}) - K_h(\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta} - \mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}) \right\}^2 \left| \mathbf{x}_i^{\mathrm{T}} (\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}) \right| d\mathbf{x}_i. \tag{6}$$

Following similar arguments for dealing with (5), we can obtain that (6) is clearly of order $o\left\{n^{-1}h^{-(2d+2)}\right\}$. These two results imply that (4) is clearly bounded by $O\left\{n^{-1}h^{-(2d+2)}\right\}$. The above result, together with Chebyshev's inequality, imply that

$$\begin{aligned}
& \Bigg| n^{-1} \sum_{i=1}^n \{K_h(\mathbf{x}_i^{\mathrm{T}}\widehat{\boldsymbol{\beta}} - \mathbf{x}^{\mathrm{T}}\widehat{\boldsymbol{\beta}}) - K_h(\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta} - \mathbf{x}^{\mathrm{T}}\boldsymbol{\beta})\}\boldsymbol{\alpha}(\mathbf{x}_i) \\
& -E\left[\{K_h(\mathbf{x}_i^{\mathrm{T}}\widehat{\boldsymbol{\beta}} - \mathbf{x}^{\mathrm{T}}\widehat{\boldsymbol{\beta}}) - K_h(\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta} - \mathbf{x}^{\mathrm{T}}\boldsymbol{\beta})\}\boldsymbol{\alpha}(\mathbf{x}_i) \right] \Bigg| \\
= \; & O_p\left\{ n^{-1} h^{-(d+1)} \right\}.
\end{aligned}$$

This result enables us to use Theorem 37 in Pollard (1984, page 34) to prove that

$$\begin{aligned}
& \sup_{\Omega_\beta} \Bigg| n^{-1} \sum_{i=1}^n \{K_h(\mathbf{x}_i^{\mathrm{T}}\widehat{\boldsymbol{\beta}} - \mathbf{x}^{\mathrm{T}}\widehat{\boldsymbol{\beta}}) - K_h(\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta} - \mathbf{x}^{\mathrm{T}}\boldsymbol{\beta})\}\boldsymbol{\alpha}(\mathbf{x}_i) \\
& -E\left[\{K_h(\mathbf{x}_i^{\mathrm{T}}\widehat{\boldsymbol{\beta}} - \mathbf{x}^{\mathrm{T}}\widehat{\boldsymbol{\beta}}) - K_h(\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta} - \mathbf{x}^{\mathrm{T}}\boldsymbol{\beta})\}\boldsymbol{\alpha}(\mathbf{x}_i) \right] \Bigg| \\
= \; & O_p\left\{ n^{-1} h^{-(d+1)} \mathrm{log} n \right\}.
\end{aligned} \tag{7}$$

One can also refer to Appendix A.1 in Wang, Xue, Zhu and Chong (2010) where they illustrate in detail how to use Theorem 37 in Pollard (1984, page 34) to derive (7). Thus the variance term is bounded. Next we bound the bias term. In the sequel we prove that

$$\sup_{\Omega_\beta}\left|E\left[\{K_h(\mathbf{x}_i^{\mathrm{T}}\widehat{\boldsymbol\beta}-\mathbf{x}^{\mathrm{T}}\widehat{\boldsymbol\beta})-K_h(\mathbf{x}_i^{\mathrm{T}}\boldsymbol\beta-\mathbf{x}^{\mathrm{T}}\boldsymbol\beta)\}\boldsymbol\alpha(\mathbf{x}_i)\right]-\mathbf{r}_1(\mathbf{x}_i^{\mathrm{T}}\widehat{\boldsymbol\beta})+\mathbf{r}_1(\mathbf{x}_i^{\mathrm{T}}\boldsymbol\beta)\right|=O\left(\frac{h^m}{\sqrt{n}}\right), \quad (8)$$

where $\mathbf{r}_1(\mathbf{x}_i^{\mathrm{T}}\boldsymbol\beta) = E\{\boldsymbol\alpha(\mathbf{x}_i) \mid \mathbf{x}_i^{\mathrm{T}}\boldsymbol\beta\}f(\mathbf{x}_i^{\mathrm{T}}\boldsymbol\beta)$, and the result holds elementwise. By using Taylor's expansion with Lagrange remainder, we have

$$\begin{aligned}
& E\left\{K_h(\mathbf{x}_i^{\mathrm{T}}\boldsymbol\beta-\mathbf{x}^{\mathrm{T}}\boldsymbol\beta)\boldsymbol\alpha(\mathbf{x}_i)\right\}-\mathbf{r}_1(\mathbf{x}^{\mathrm{T}}\boldsymbol\beta)\\
=\ & E\left[K_h(\mathbf{x}_i^{\mathrm{T}}\boldsymbol\beta-\mathbf{x}^{\mathrm{T}}\boldsymbol\beta)E\{\boldsymbol\alpha(\mathbf{x}_i)\mid \mathbf{x}_i^{\mathrm{T}}\boldsymbol\beta\}\right]-\mathbf{r}_1(\mathbf{x}^{\mathrm{T}}\boldsymbol\beta)\\
=\ & \int K_h(\mathbf{x}_i^{\mathrm{T}}\boldsymbol\beta-\mathbf{x}^{\mathrm{T}}\boldsymbol\beta)\mathbf{r}_1(\mathbf{x}_i^{\mathrm{T}}\boldsymbol\beta)d(\mathbf{x}_i^{\mathrm{T}}\boldsymbol\beta)-\mathbf{r}_1(\mathbf{x}^{\mathrm{T}}\boldsymbol\beta)\\
=\ & \int K(\mathbf{z}_i)\mathbf{r}_1(\mathbf{x}^{\mathrm{T}}\boldsymbol\beta+h\mathbf{z}_i)d\mathbf{z}_i-\mathbf{r}_1(\mathbf{x}^{\mathrm{T}}\boldsymbol\beta)\\
=\ & \int K(\mathbf{z}_i)(h\mathbf{z}_i)^m\{\mathbf{r}_1^{(m)}(\mathbf{x}^{\mathrm{T}}\boldsymbol\beta+h\mathbf{z}_i^{\star})\}/m!d\mathbf{z}_i,
\end{aligned}$$

where $\mathbf{z}_i^{\star}$ is between $\mathbf{x}^{\mathrm{T}}\boldsymbol\beta$ and $\mathbf{x}^{\mathrm{T}}\boldsymbol\beta+\mathbf{x}_i^{\mathrm{T}}\boldsymbol\beta$. Let $\mathbf{z}_i^{\star\star}$ be a proper value between $\mathbf{x}^{\mathrm{T}}\widehat{\boldsymbol\beta}$ and $\mathbf{x}^{\mathrm{T}}\widehat{\boldsymbol\beta}+\mathbf{x}_i^{\mathrm{T}}\widehat{\boldsymbol\beta}$. It follows from the local Lipschitz continuity of $\mathbf{r}_1^{(m)}$ in Condition (C2) that

$$\begin{aligned}
& \sup_{\Omega_\beta}\left|E\left[\{K_h(\mathbf{x}_i^{\mathrm{T}}\widehat{\boldsymbol\beta}-\mathbf{x}^{\mathrm{T}}\widehat{\boldsymbol\beta})-K_h(\mathbf{x}_i^{\mathrm{T}}\boldsymbol\beta-\mathbf{x}^{\mathrm{T}}\boldsymbol\beta)\}\boldsymbol\alpha(\mathbf{x}_i)\right]-\{\mathbf{r}_1(\mathbf{x}_i^{\mathrm{T}}\widehat{\boldsymbol\beta})-\mathbf{r}_1(\mathbf{x}_i^{\mathrm{T}}\boldsymbol\beta)\}\right|\\
=\ & \sup_{\Omega_\beta}\left|\int K(\mathbf{z}_i)(h\mathbf{z}_i)^m\{\mathbf{r}_1^{(m)}(\mathbf{x}^{\mathrm{T}}\boldsymbol\beta+h\mathbf{z}_i^{\star})-\mathbf{r}_1^{(m)}(\mathbf{x}^{\mathrm{T}}\widehat{\boldsymbol\beta}+h\mathbf{z}_i^{\star\star})\}/m!d\mathbf{z}_i\right|\\
=\ & \sup_{\Omega_\beta}\left|\int K(\mathbf{z}_i)(h\mathbf{z}_i)^m\frac{\mathbf{r}_1^{(m)}(\mathbf{x}^{\mathrm{T}}\boldsymbol\beta)\{1+O(h\mathbf{z}_i^{\star})\}-\mathbf{r}_1^{(m)}(\mathbf{x}^{\mathrm{T}}\widehat{\boldsymbol\beta})\{1+O(h\mathbf{z}_i^{\star\star})\}}{m!}d\mathbf{z}_i\right|\\
=\ & O\left(\frac{h^m}{\sqrt{n}}\right),
\end{aligned}$$

which proves (8). Combining the results of (7) and (8), we obtain that

$$\begin{aligned}
& \sup_{\Omega_\beta}\left|\frac{1}{n}\sum_{i=1}^{n}\{K_h(\mathbf{x}_i^{\mathrm{T}}\widehat{\boldsymbol\beta}-\mathbf{x}^{\mathrm{T}}\widehat{\boldsymbol\beta})-K_h(\mathbf{x}_i^{\mathrm{T}}\boldsymbol\beta-\mathbf{x}^{\mathrm{T}}\boldsymbol\beta)\}\boldsymbol\alpha(\mathbf{x}_i)-\{\mathbf{r}_1(\mathbf{x}_i^{\mathrm{T}}\widehat{\boldsymbol\beta})-\mathbf{r}_1(\mathbf{x}_i^{\mathrm{T}}\boldsymbol\beta)\}\right|\\
=\ & O_p\left\{h^m/\sqrt{n}+n^{-1}h^{-(d+1)}\log n\right\}. \qquad (9)
\end{aligned}$$

Following similar arguments for proving (9) by letting $\boldsymbol\alpha(\mathbf{x}_i)=1$, we can prove that

$$\begin{aligned}
& \sup_{\Omega_\beta}\left|\frac{1}{n}\sum_{i=1}^{n}\{K_h(\mathbf{x}_i^{\mathrm{T}}\widehat{\boldsymbol\beta}-\mathbf{x}^{\mathrm{T}}\widehat{\boldsymbol\beta})-K_h(\mathbf{x}_i^{\mathrm{T}}\boldsymbol\beta-\mathbf{x}^{\mathrm{T}}\boldsymbol\beta)\}-\{f(\mathbf{x}_i^{\mathrm{T}}\widehat{\boldsymbol\beta})-f(\mathbf{x}_i^{\mathrm{T}}\boldsymbol\beta)\}\right|\\
=\ & O_p\left\{h^m/\sqrt{n}+n^{-1}h^{-(d+1)}\log n\right\}. \qquad (10)
\end{aligned}$$

The result of (9) and (10) imply immediately (2).

We omit the details for proving (3) because it can be proven similarly. □

Lemma 4. *Assume Conditions (C1)-(C3) hold. Then*

$$\begin{aligned} & \frac{1}{n}\sum_{i=1}^{n}\left[\mathbf{g}(Y_i,\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}) - E\{\mathbf{g}(Y_i,\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}) \mid \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}\}\right]\left[\widehat{E}\{\boldsymbol{\alpha}(\mathbf{x}_i) \mid \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}\} - E\{\boldsymbol{\alpha}(\mathbf{x}_i) \mid \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}\}\right] \\ = \; & O_p\left\{1/\left(nh^{d/2}\right) + h^m/n^{1/2} + h^{2m} + \log^2 n/\left(nh^d\right)\right\}, \\ \textit{and} \quad & \frac{1}{n}\sum_{i=1}^{n}\left[\widehat{E}\{\mathbf{g}(Y_i,\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}) \mid \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}\} - E\{\mathbf{g}(Y_i,\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}) \mid \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}\}\right]\left[\boldsymbol{\alpha}(\mathbf{x}_i) - E\{\boldsymbol{\alpha}(\mathbf{x}_i) \mid \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}\}\right] \\ = \; & O_p\left\{1/\left(nh^{d/2}\right) + h^m/n^{1/2} + h^{2m} + \log^2 n/\left(nh^d\right)\right\}. \end{aligned}$$

Proof for Lemma 4. Because these two equalities and their proofs are very similar, we only show the first one. For simplicity, we let $\varepsilon_i = \mathbf{g}(Y_i,\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}) - E\{\mathbf{g}(Y_i,\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}) \mid \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}\}$, $\widehat{f}(\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}) = (n-1)^{-1}\sum_{j\neq i}K_h(\mathbf{x}_j^{\mathrm{T}}\boldsymbol{\beta} - \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta})$ and $\widehat{\mathbf{r}}_1(\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}) = (n-1)^{-1}\sum_{j\neq i}K_h(\mathbf{x}_j^{\mathrm{T}}\boldsymbol{\beta} - \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta})\boldsymbol{\alpha}(\mathbf{x}_j)$. Notice that $E\{\boldsymbol{\alpha}(\mathbf{x}_i) \mid \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}\} = \mathbf{r}_1(\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta})/f(\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta})$. After some simple algebra, we have

$$\begin{aligned} & \frac{1}{n}\sum_{i=1}^{n}\varepsilon_i\left[\widehat{E}\{\boldsymbol{\alpha}(\mathbf{x}_i) \mid \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}\} - E\{\boldsymbol{\alpha}(\mathbf{x}_i) \mid \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}\}\right] \\ = \; & \frac{1}{n}\sum_{i=1}^{n}\varepsilon_i\left[\frac{\widehat{\mathbf{r}}_1(\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta})}{\widehat{f}(\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta})} - \frac{\mathbf{r}_1(\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta})}{f(\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta})}\right] \\ = \; & \frac{1}{n}\sum_{i=1}^{n}\varepsilon_i\left\{\frac{\widehat{\mathbf{r}}_1(\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}) - \mathbf{r}_1(\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta})}{f(\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta})}\right\} - \frac{1}{n}\sum_{i=1}^{n}\varepsilon_i\left[\frac{\mathbf{r}_1(\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta})\{\widehat{f}(\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}) - f(\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta})\}}{f^2(\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta})}\right] \\ & -\frac{1}{n}\sum_{i=1}^{n}\varepsilon_i\left[\frac{\{\widehat{\mathbf{r}}_1(\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}) - \mathbf{r}_1(\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta})\}\{\widehat{f}(\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}) - f(\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta})\}}{f(\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta})\widehat{f}(\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta})}\right] \\ & +\frac{1}{n}\sum_{i=1}^{n}\varepsilon_i\left[\frac{\mathbf{r}_1(\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta})\{\widehat{f}(\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}) - f(\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta})\}^2}{f^2(\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta})\widehat{f}(\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta})}\right]. \end{aligned} \tag{11}$$

We notice that the first two quantities in the right hand side of (11) have similar structure. Thus in the sequel we only deal with the first quantity. By the uniform convergence of nonparametric regression, the third and the fourth quantities are clearly of order $O_p\left\{h^{2m} + \log^2 n/\left(nh^d\right)\right\}$. Thus it suffices to study the convergence rate of the first quantity below. We write $n^{-1}\sum_{i=1}^{n}\widehat{\mathbf{r}}_1(\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta})\varepsilon_i$ as a second-order U-statistic:

$$\frac{1}{n}\sum_{i=1}^{n}\widehat{\mathbf{r}}_1(\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta})\varepsilon_i = \frac{1}{n(n-1)}\sum_{i\neq j}^{n}K_h(\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta} - \mathbf{x}_j^{\mathrm{T}}\boldsymbol{\beta})\left\{\varepsilon_i\boldsymbol{\alpha}(\mathbf{x}_j) + \varepsilon_j\boldsymbol{\alpha}(\mathbf{x}_i)\right\}.$$

By using Lemma 5.2.1.A of Serfling (1980, page 183), it follows that

$$\frac{1}{n}\sum_{i=1}^{n}\widehat{\mathbf{r}}_1(\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta})\varepsilon_i - \frac{1}{n}\sum_{i=1}^{n}\varepsilon_i E\left\{K_h(\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta} - \mathbf{x}_j^{\mathrm{T}}\boldsymbol{\beta})\mathbf{r}_1(\mathbf{x}_j^{\mathrm{T}}\boldsymbol{\beta}) \mid \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}\right\} = O_p\left\{1/\left(nh^{d/2}\right)\right\}. \tag{12}$$

The above equality follows because the difference on the left hand side is a degenerated U-statistic. Next we show that

$$\frac{1}{n}\sum_{i=1}^{n}\varepsilon_i\left[E\left\{K_h(\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}-\mathbf{x}_j^{\mathrm{T}}\boldsymbol{\beta})\mathbf{r}_1(\mathbf{x}_j^{\mathrm{T}}\boldsymbol{\beta})\mid \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}\right\}-\mathbf{r}_1\left(\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}\right)f\left(\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}\right)\right]=O_p(n^{-1/2}h^m). \quad (13)$$

By using the standard arguments to calculate the bias term in nonparametric regression, we can easily have $\sup_{\mathbf{x}_i}\left|E\left\{K_h(\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}-\mathbf{x}_j^{\mathrm{T}}\boldsymbol{\beta})\mathbf{r}_1(\mathbf{x}_j^{\mathrm{T}}\boldsymbol{\beta})\mid \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}\right\}-\mathbf{r}_1(\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta})f(\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta})\right|=O_p(h^m)$ by assuming that the m-th derivative of $\left\{\mathbf{r}_1(\mathbf{x}^{\mathrm{T}}\boldsymbol{\beta})f(\mathbf{x}^{\mathrm{T}}\boldsymbol{\beta})\right\}$ is locally Lipschitz-continuous. This proves (13). Combining (12) and (13), we obtain that

$$\frac{1}{n}\sum_{i=1}^{n}\varepsilon_i\left\{\widehat{\mathbf{r}}_1(\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta})-\mathbf{r}_1\left(\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}\right)f\left(\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}\right)\right\}=O_p\left\{1/\left(nh^{d/2}\right)+h^m/n^{1/2}\right\}.$$

This result together with (11) entails the desired result, which completes the proof. □

Proof for Theorem 1. We notice that the resulting estimator $\widehat{\boldsymbol{\beta}}$ satisfies

$$\sum_{i=1}^{n}\left[\mathbf{g}(Y_i,\mathbf{x}_i^{\mathrm{T}}\widehat{\boldsymbol{\beta}})-\widehat{E}\{\mathbf{g}(Y_i,\mathbf{x}_i^{\mathrm{T}}\widehat{\boldsymbol{\beta}})\mid \mathbf{x}_i^{\mathrm{T}}\widehat{\boldsymbol{\beta}}\}\right]\left[\boldsymbol{\alpha}(\mathbf{x}_i)-\widehat{E}\{\boldsymbol{\alpha}(\mathbf{x}_i)\mid \mathbf{x}_i^{\mathrm{T}}\widehat{\boldsymbol{\beta}}\}\right]=\mathbf{0}.$$

The left hand side of the above display can be decomposed into four terms.

$$\begin{aligned}
&\sum_{i=1}^{n}\left[\mathbf{g}(Y_i,\mathbf{x}_i^{\mathrm{T}}\widehat{\boldsymbol{\beta}})-\widehat{E}\{\mathbf{g}(Y_i,\mathbf{x}_i^{\mathrm{T}}\widehat{\boldsymbol{\beta}})\mid \mathbf{x}_i^{\mathrm{T}}\widehat{\boldsymbol{\beta}}\}\right]\left[\boldsymbol{\alpha}(\mathbf{x}_i)-\widehat{E}\{\boldsymbol{\alpha}(\mathbf{x}_i)\mid \mathbf{x}_i^{\mathrm{T}}\widehat{\boldsymbol{\beta}}\}\right]\\
=&\sum_{i=1}^{n}\left[\mathbf{g}(Y_i,\mathbf{x}_i^{\mathrm{T}}\widehat{\boldsymbol{\beta}})-E\{\mathbf{g}(Y_i,\mathbf{x}_i^{\mathrm{T}}\widehat{\boldsymbol{\beta}})\mid \mathbf{x}_i^{\mathrm{T}}\widehat{\boldsymbol{\beta}}\}\right]\left[\boldsymbol{\alpha}(\mathbf{x}_i)-E\{\boldsymbol{\alpha}(\mathbf{x}_i)\mid \mathbf{x}_i^{\mathrm{T}}\widehat{\boldsymbol{\beta}}\}\right]\\
+&\sum_{i=1}^{n}\left[\mathbf{g}(Y_i,\mathbf{x}_i^{\mathrm{T}}\widehat{\boldsymbol{\beta}})-E\{\mathbf{g}(Y_i,\mathbf{x}_i^{\mathrm{T}}\widehat{\boldsymbol{\beta}})\mid \mathbf{x}_i^{\mathrm{T}}\widehat{\boldsymbol{\beta}}\}\right]\left[E\{\boldsymbol{\alpha}(\mathbf{x}_i)\mid \mathbf{x}_i^{\mathrm{T}}\widehat{\boldsymbol{\beta}}\}-\widehat{E}\{\boldsymbol{\alpha}(\mathbf{x}_i)\mid \mathbf{x}_i^{\mathrm{T}}\widehat{\boldsymbol{\beta}}\}\right] \qquad (14)\\
+&\sum_{i=1}^{n}\left[E\{\mathbf{g}(Y_i,\mathbf{x}_i^{\mathrm{T}}\widehat{\boldsymbol{\beta}})\mid \mathbf{x}_i^{\mathrm{T}}\widehat{\boldsymbol{\beta}}\}-\widehat{E}\{\mathbf{g}(Y_i,\mathbf{x}_i^{\mathrm{T}}\widehat{\boldsymbol{\beta}})\mid \mathbf{x}_i^{\mathrm{T}}\widehat{\boldsymbol{\beta}}\}\right]\left[\boldsymbol{\alpha}(\mathbf{x}_i)-E\{\boldsymbol{\alpha}(\mathbf{x}_i)\mid \mathbf{x}_i^{\mathrm{T}}\widehat{\boldsymbol{\beta}}\}\right]\\
+&\sum_{i=1}^{n}\left[E\{\mathbf{g}(Y_i,\mathbf{x}_i^{\mathrm{T}}\widehat{\boldsymbol{\beta}})\mid \mathbf{x}_i^{\mathrm{T}}\widehat{\boldsymbol{\beta}}\}-\widehat{E}\{\mathbf{g}(Y_i,\mathbf{x}_i^{\mathrm{T}}\widehat{\boldsymbol{\beta}})\mid \mathbf{x}_i^{\mathrm{T}}\widehat{\boldsymbol{\beta}}\}\right]\left[E\{\boldsymbol{\alpha}(\mathbf{x}_i)\mid \mathbf{x}_i^{\mathrm{T}}\widehat{\boldsymbol{\beta}}\}-\widehat{E}\{\boldsymbol{\alpha}(\mathbf{x}_i)\mid \mathbf{x}_i^{\mathrm{T}}\widehat{\boldsymbol{\beta}}\}\right].
\end{aligned}$$

In the sequel we show that the first term in the right hand side of (14) is $O_p\left(n^{1/2}\right)$ while the other three terms are $o_p\left(n^{1/2}\right)$. We notice that the first term in (14) can be expanded as

$$\begin{aligned}
&\sum_{i=1}^{n}\left[\mathbf{g}(Y_i,\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta})-E\{\mathbf{g}(Y_i,\mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta})\mid \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}\}\right]\left[\boldsymbol{\alpha}(\mathbf{x}_i)-E\{\boldsymbol{\alpha}(\mathbf{x}_i)\mid \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}\}\right]\\
+\quad & n\mathrm{dvec}\ E\left\{\frac{\partial \mathrm{vec}\left(\left[\mathbf{g}(Y,\mathbf{x}^{\mathrm{T}}\boldsymbol{\beta})-E\{\mathbf{g}(Y,\mathbf{x}^{\mathrm{T}}\boldsymbol{\beta})\mid \mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}\}\right]\left[\boldsymbol{\alpha}(\mathbf{x})-E\{\boldsymbol{\alpha}(\mathbf{x})\mid \mathbf{x}^{\mathrm{T}}\boldsymbol{\beta}\}\right]\right)}{\partial\left\{\mathrm{vec}(\boldsymbol{\beta})\right\}^{\mathrm{T}}}\right\}\\
&\cdot\mathrm{vec}(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})+o_p(n^{1/2}). \qquad (15)
\end{aligned}$$

Here dvec$(\cdot)$ is such that dvec$\{\text{vec}(\mathbf{M})\} = \mathbf{M}$ for any matrix $\mathbf{M}$.

Next we deal with the second quantity in the right hand side of (14). By using Lemma 3, this quantity can be approximated by

$$\sum_{i=1}^{n} \left[\mathbf{g}(Y_i, \mathbf{x}_i^{\mathrm{T}}\widehat{\boldsymbol{\beta}}) - E\{\mathbf{g}(Y_i, \mathbf{x}_i^{\mathrm{T}}\widehat{\boldsymbol{\beta}}) \mid \mathbf{x}_i^{\mathrm{T}}\widehat{\boldsymbol{\beta}}\}\right] \left[E\{\boldsymbol{\alpha}(\mathbf{x}_i) \mid \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}\} - \widehat{E}\{\boldsymbol{\alpha}(\mathbf{x}_i) \mid \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}\}\right] \{1 + o_p(1)\},$$

which, again by Taylor's expansion, is asymptotically equivalent to

$$\sum_{i=1}^{n} \left[\mathbf{g}(Y_i, \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}) - E\{\mathbf{g}(Y_i, \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}) \mid \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}\}\right] \left[E\{\boldsymbol{\alpha}(\mathbf{x}_i) \mid \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}\} - \widehat{E}\{\boldsymbol{\alpha}(\mathbf{x}_i) \mid \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}\}\right] \{1 + o_p(1)\}.$$

Lemma 4 implies that the above display is of order $o_p(n^{1/2})$ when $nh^{4m} \to 0$ and $nh^{2d} \to \infty$.

We turn to the third quantity in the right hand side of (14). By using Lemma 3 again, this quantity can be approximated with

$$\begin{aligned}
&\sum_{i=1}^{n} \left[E\{\mathbf{g}(Y_i, \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}) \mid \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}\} - \widehat{E}\{\mathbf{g}(Y_i, \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}) \mid \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}\}\right] \left[\boldsymbol{\alpha}(\mathbf{x}_i) - E\{\boldsymbol{\alpha}(\mathbf{x}_i) \mid \mathbf{x}_i^{\mathrm{T}}\widehat{\boldsymbol{\beta}}\}\right] \{1 + o_p(1)\} \\
= &\sum_{i=1}^{n} \left[E\{\mathbf{g}(Y_i, \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}) \mid \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}\} - \widehat{E}\{\mathbf{g}(Y_i, \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}) \mid \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}\}\right] \left[\boldsymbol{\alpha}(\mathbf{x}_i) - E\{\boldsymbol{\alpha}(\mathbf{x}_i) \mid \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}\}\right] \{1 + o_p(1)\} \\
&+ \sum_{i=1}^{n} \left[E\{\mathbf{g}(Y_i, \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}) \mid \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}\} - \widehat{E}\{\mathbf{g}(Y_i, \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}) \mid \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}\}\right] \left[E\{\boldsymbol{\alpha}(\mathbf{x}_i) \mid \mathbf{x}_i^{\mathrm{T}}\widehat{\boldsymbol{\beta}}\} - E\{\boldsymbol{\alpha}(\mathbf{x}_i) \mid \mathbf{x}_i^{\mathrm{T}}\boldsymbol{\beta}\}\right].
\end{aligned}$$

Lemma 4 implies that the first term in the above display is of order $o_p\left(n^{1/2}\right)$. The second term is clearly of order $o_p(n^{1/2})$ by using the uniform convergence of nonparametric regression (Mack and Silverman, 1982) and Taylor's expansion.

The last quantity in the right hand side of (14) is of order $O_p\left\{n\left(h^{2m} + \log^2 n/(nh^d)\right)\right\}$ (Mack and Silverman, 1982), which is of order $o_p(n^{1/2})$ when $nh^{4m} \to 0$ and $nh^{2d} \to \infty$.

Combining the above results for all four terms in (14), we obtain the root-n rate and the asymptotic normality directly from (15), which completes the proof for the asymptotic normality of $\widehat{\boldsymbol{\beta}}$. □

References

Dong, Y. and Li, B. (2010) Dimension reduction for non-elliptically distributed predictors: second-order moments. *Biometrika*, **97**, 279–294.

Li, B. and Dong, Y. (2009) Dimension reduction for non-elliptically distributed predictors. *Annals of Statistics*, **37**, 1272–1298.

Mack, Y. P. and Silverman, B. W. (1982) Weak and strong uniform consistency of kernel regression estimates. *Zeitschrift für Wahrscheinlichkeitstheorie und verwandte Gebiete*, **61**, 405–415.

Pollard, D. (1984). *Convergence of Stochastic Processes*. Springer, New York.

Serfling, R. J. (1980) *Approximation Theorems of Mathematical Statistics*, New York: John Wiley & Sons Inc.

Wang, J. L., Xue, L. G., Zhu, L. X. and Chong, Y. S. (2010) Estimation for a partial-linear single-index model. *Annals of Statistics*, **38**, 246–274