Statistica Sinica **28** (2018), 2353-2373
doi:https://doi.org/10.5705/ss.202017.0101

# ESTIMATION OF ERRORS-IN-VARIABLES PARTIALLY LINEAR ADDITIVE MODELS

Eun Ryung Lee, Kyunghee Han and Byeong U. Park

*Sungkyunkwan University, University of California, Davis and Seoul National University*

*Abstract:* In this paper we consider partially linear additive models where the predictors in the parametric and in the nonparametric parts are contaminated by measurement errors. We propose an estimator of the parametric part and show that it achieves $\sqrt{n}$-consistency in a certain range of the smoothness of the measurement errors in the nonparametric part. We also derive the convergence rate of the parametric estimator in case the smoothness of the measurement errors is off the range. Furthermore, we suggest an estimator of the additive function in the nonparametric part that achieves the optimal one-dimensional convergence rate in nonparametric deconvolution problems. We conducted a simulation study that confirms our theoretical findings.

*Key words and phrases:* Attenuation, deconvolution, errors in variables, kernel smoothing, rate of convergence, smooth backfitting.

## 1. Introduction

In this paper, dedicated to the memory of Peter G. Hall, we consider an errors-in-variables regression model. A typical type of errors-in-variables problem is to estimate the density of a variable $X$ or the regression function $E(Y|X=\cdot)$ for a response $Y$ and a predictor $X$ when $X$ is not observed but $X^* = X + U$ is observed with measurement error $U$ that is independent of $X$. This topic is one of Peter Hall's areas in which he made fundamental contributions. Carroll and Hall (1988) provides the minimax rate of convergence for nonparametric density estimation. Delaigle, Hall and Meister (2008) studies the problem when the density of $U$ is unknown but is estimated from repeated contaminated measurements. Recently, his last paper on the topic, Delaigle and Hall (2016), demonstrates that one can estimate the density of $X$ using its phase function. Carroll, Delaigle and Hall (2009) tackles a prediction problem when the measurement error $U_F$ on $X$ for future observations is not identically distributed as $U$ so that the main task is to estimate $E(Y|X+U_F=\cdot)$ given a random sample of $(X^*, Y)$. Peter also made groundbreaking contributions to the topic when considering Berkson measurement error, where $U$ is independent of $X^*$, but not of $X$. Some of the main achievements in this area are Delaigle, Hall and Qiu (2006), Carroll, Delaigle and Hall (2007) and Delaigle, Hall and Müller (2007), among others. For other contributions of Peter Hall to the topic and for an excellent account of his achievements, the reader is referred to Delaigle (2016).

The present paper complements Peter Hall's work in nonparametric errors-in-variables problems. Specifically, we study the estimation of partially linear additive models when the predictors in the nonparametric part as well as those in the parametric part are contaminated by measurement errors. There have been some earlier works on partially linear models with errors-in-variables. Two works that are most closely related to the problem we study in this paper are Liang, Härdle and Carroll (1999) and Zhu and Cui (2003) which consider partially linear models where the nonparametric component is univariate. Liang, Härdle and Carroll (1999) treated the case where only the predictors in the parametric part are contaminated; Zhu and Cui (2003) extended the work to the case where both predictors in the parametric and in the nonparametric parts are observed with measurement errors. One can extend the latter work in a straightforward manner to the case where the predictor in the nonparametric part is multi-dimensional, but the procedure could then lead to the curse of dimensionality.

In this paper we study the estimation of partially linear models where the multivariate nonparametric part has an additive structure. In multivariate nonparametric regression, additive models are known to avoid the curse of dimensionality, see Mammen, Linton and Nielsen (1999), Yu, Park and Mammen (2008) and Lee, Mammen and Park (2010, 2012), among others. Specifically, we consider the case where we observe a response $Y$ and predictors $\mathbf{X} = (X_1, \dots, X_p)^\top$ and $\mathbf{Z} = (Z_1, \dots, Z_d)^\top$ such that

$$Y = \boldsymbol{\theta}^\top \mathbf{X} + m_0 + m_1(Z_1) + \cdots + m_d(Z_d) + \varepsilon, \tag{1.1}$$

where $\mathrm{E}(\varepsilon|\mathbf{X}, \mathbf{Z}) = 0$. We discuss how to estimate $\boldsymbol{\theta}$ and the univariate nonparametric component functions $m_j$ when we observe the contaminated predictors

$$\mathbf{X}^* = \mathbf{X} + \mathbf{U}, \quad \mathbf{Z}^* = \mathbf{Z} + \mathbf{V}$$

instead of $\mathbf{X}$ and $\mathbf{Z}$, where $\mathbf{U} = (U_1, \dots, U_p)^\top$ and $\mathbf{V} = (V_1, \dots, V_d)^\top$ are vectors of measurement errors.

In (1.1), we assume that the $m_j$ are square integrable and that the predictors $Z_j$ are supported on compact sets, say [0,1]. For identifiability of the additive component functions $m_j$, we put on the constraints $\mathrm{E}m_j(Z_j) = 0$, $1 \le j \le d$, introducing a constant $m_0$ in the model. The response error $\varepsilon$ is independent of $\mathbf{X}, \mathbf{Z}, \mathbf{U}, \mathbf{V}$, for simplicity of presentation. The measurement error vectors $\mathbf{U}$ and $\mathbf{V}$ are independent of each other, $\mathbf{U}$ has mean zero and known variance matrix $\mathbf{\Sigma_U}$, and $\mathbf{V}$ has a symmetric density $p_{\mathbf{V}}$. We also assume that the components $V_j$ and $V_k$ of $\mathbf{V}$ are independent for $j \neq k$, and that $(\mathbf{U}, \mathbf{V})$ is independent of $(\mathbf{X}, \mathbf{Z})$.

The parametric component $\boldsymbol{\theta}$ is identifiable in the model (1.1) if

$$\mathbf{D}_0 \equiv \mathrm{E}\left\{\mathbf{X} - \mathrm{E}(\mathbf{X}|\mathbf{Z})\right\}\left\{\mathbf{X} - \mathrm{E}(\mathbf{X}|\mathbf{Z})\right\}^\top$$

is invertible. This is true even in a wider model where the nonparametric part may not be an additive function, but is allowed to be a $d$-dimensional multivariate function. This follows simply from the identity

$$\mathrm{E}\left\{\mathbf{X} - \mathrm{E}(\mathbf{X}|\mathbf{Z})\right\}\left\{Y - \mathrm{E}(Y|\mathbf{Z})\right\} = \mathbf{D}_0\boldsymbol{\theta}. \tag{1.2}$$

Thus, $\boldsymbol{\theta}$ is identifiable in the smaller model (1.1). In fact, with the additive structure of the nonparametric function in (1.1), it is identifiable under the weaker condition that $\mathrm{E}\left\{\mathbf{X} - \boldsymbol{\eta}(\mathbf{Z})\right\}\left\{\mathbf{X} - \boldsymbol{\eta}(\mathbf{Z})\right\}^\top$ is invertible, where $\boldsymbol{\eta} = \Pi\left(\mathrm{E}(\mathbf{X}|\mathbf{Z} = \cdot)|\mathcal{H}\right)$, the projection of the multivariate function $\mathrm{E}(\mathbf{X}|\mathbf{Z} = \cdot)$ onto the space of additive functions, denoted by $\mathcal{H}$. Under (1.1) we have

$$\mathrm{E}\left\{\mathbf{X} - \boldsymbol{\eta}(\mathbf{Z})\right\}\left\{Y - \xi(\mathbf{Z})\right\} = \mathrm{E}\left\{\mathbf{X} - \boldsymbol{\eta}(\mathbf{Z})\right\}\left\{\mathbf{X} - \boldsymbol{\eta}(\mathbf{Z})\right\}^\top \boldsymbol{\theta}, \tag{1.3}$$

where $\xi = \Pi\left(\mathrm{E}(Y|\mathbf{Z} = \cdot)|\mathcal{H}\right)$. Here $\mathrm{E}\left\{\mathbf{X} - \boldsymbol{\eta}(\mathbf{Z})\right\}\left\{\mathbf{X} - \boldsymbol{\eta}(\mathbf{Z})\right\}^\top - \mathbf{D}_0$ is nonnegative definite.

We propose an estimator of $\boldsymbol{\theta}$ that basically solves an empirical version of (1.3). In Section 2 we take a different perspective to motivate our estimator. To get an empirical version of (1.3) we estimate $\boldsymbol{\eta}$ and $\xi$ using a kernel smoothing technique. In particular, we use the smooth backfitting technique of Mammen, Linton and Nielsen (1999) and the smoothed normalized deconvolution kernel of Han and Park (2018). To the best of our knowledge, this work is the first to study kernel estimation as (1.1) based on the observation of the contaminated predictors $\mathbf{X}^*$ and $\mathbf{Z}^*$. The problem is harder than when only $\mathbf{X}$ is contaminated or the nonparametric part is univariate, the focus of most earlier works.

The difficulty of deconvoluting measurement errors in nonparametric smoothing depends on the smoothness of the measurement error distributions, or the tail behavior of their characteristic functions, as well as the smoothness of the object function being estimated. In this paper we consider the so-called 'ordinary smooth' case where $\phi_{V_j}(t)$, the characteristic functions of $V_j$, decay at the tails at a rate $|t|^{-\beta}$ as $|t| \to \infty$ for some $\beta > 0$. We show that our estimator of $\boldsymbol{\theta}$ achieves $\sqrt{n}$-consistency regardless of the dimention $d$ when $\beta < 1/2$. In case $\beta \geq 1/2$ we find that the estimator has the rate $O_p(n^{-1/(1+2\beta)})$ up to a logarithmic factor. An estimator of $\boldsymbol{\theta}$ based on (1.2) and multivariate smoothing for estimating $\mathrm{E}(Y|\mathbf{Z}=\cdot)$ and $\mathrm{E}(\mathbf{X}|\mathbf{Z}=\cdot)$ does not give these rates. Our main focus is on the estimation of the parametric part, but we also suggest an estimator of the additive function in the nonparametric part and show that it achieves the optimal one-dimensional convergence rate in nonparametric deconvolution problems regardless of the dimension $d$. We conducted a simulation study to demonstrate the finite sample performance of the proposed estimator and found that it supports our theoretical findings.

## 2. Least Favorable Submodel and Smooth Backfitting

Here, we motivate our estimator of $\boldsymbol{\theta}$ from the theory of semiparametric efficient estimation. For this, we briefly review an estimation procedure when there are no measurement errors in the predictors. The latter, studied by Yu, Mammen and Park (2011), finds the ‘least favorable’ regular parametric submodel of (1.1) and estimates the true value of $\boldsymbol{\theta}$ in the submodel where the estimation is hardest in the sense of efficiency. By the standard theory of semiparametric efficient estimation, this procedure leads to a semiparametric efficient estimator of the parametric component. For the standard theory of semiparametric efficient estimation, the reader is referred to Bickel et al. (1993).

We write $m(\mathbf{z}) = m_0 + m_1(z_1) + \cdots + m_d(z_d)$, where $\mathrm{E}\, m_j(Z_j) = 0$, $1 \leq j \leq d$. Let $\mathcal{H}$ denote the space of all additive square integrable functions $g$ such that $g(\mathbf{z}) = g_1(z_1) + \cdots + g_d(z_d)$. Let $(\boldsymbol{\theta}^0, m^0)$ denote a fixed value of the parameter $(\boldsymbol{\theta}, m)$. Then, a regular parametric submodel of (1.1) at $(\boldsymbol{\theta}^0, m^0)$ may be written as $\mathcal{P}_0 = \{(\boldsymbol{\theta}, m(\,\cdot\,, \boldsymbol{\theta})) : \boldsymbol{\theta} \in \mathbb{R}^p, m(\,\cdot\,, \boldsymbol{\theta}^0) = m^0\}$ for a Fréchet differentiable map $\boldsymbol{\theta} \mapsto m(\,\cdot\,, \boldsymbol{\theta}) \in \mathcal{H}$. The least favorable regular parametric submodel is the one that has the smallest Fisher information. For a map $\boldsymbol{\theta} \mapsto m(\,\cdot\,, \boldsymbol{\theta})$ with the Fréchet derivative $\boldsymbol{\delta} = \partial m(\,\cdot\,, \boldsymbol{\theta})/\partial \boldsymbol{\theta}\big|_{\boldsymbol{\theta}=\boldsymbol{\theta}^0}$, the score function of $\ell(\boldsymbol{\theta}, m(\,\cdot\,, \boldsymbol{\theta})) \equiv \log p_\varepsilon(Y - \boldsymbol{\theta}^\top \mathbf{X} - m(\mathbf{Z}, \boldsymbol{\theta}))$ at $\boldsymbol{\theta} = \boldsymbol{\theta}^0$ is given by

$$\frac{d}{d\boldsymbol{\theta}} \ell(\boldsymbol{\theta}, m(\,\cdot\,, \boldsymbol{\theta}))\big|_{\boldsymbol{\theta}=\boldsymbol{\theta}^0} = -\frac{p_\varepsilon'}{p_\varepsilon}\{Y - \boldsymbol{\theta}^{0\top}\mathbf{X} - m^0(\mathbf{Z})\} \cdot \{\mathbf{X} + \boldsymbol{\delta}(\mathbf{Z})\},$$

where $p_\varepsilon$ denotes the density of the response error $\varepsilon$. This gives the Fisher information at $\boldsymbol{\theta}^0$ in the submodel with direction $\boldsymbol{\delta}$ as

$$I(\boldsymbol{\delta}) = I_0 \cdot \mathrm{E}\{\mathbf{X} + \boldsymbol{\delta}(\mathbf{Z})\}\{\mathbf{X} + \boldsymbol{\delta}(\mathbf{Z})\}^\top,$$

where $I_0 = \int (p_\varepsilon')^2/p_\varepsilon$. Thus, the direction $\boldsymbol{\delta}^*$ that minimizes $I(\boldsymbol{\delta})$ among all

$\boldsymbol{\delta} = (\delta_1, \ldots, \delta_p)^\top$, with each $\delta_j \in \mathcal{H}$, is given by $\boldsymbol{\delta}^* = -\boldsymbol{\eta}$, where

$$\boldsymbol{\eta} = \Pi\left(\mathrm{E}(\mathbf{X}|\mathbf{Z} = \cdot)|\mathcal{H}\right),$$

and $\Pi(\,\cdot\,|\mathcal{H})$ denotes the projection operator onto the space $\mathcal{H}$. The projection $\boldsymbol{\eta} = (\eta_1, \ldots, \eta_p)^\top$ has each $\eta_j$ belonging to the space of additive functions $\mathcal{H}$.

Now, let $(\boldsymbol{\theta}^0, m^0)$ be the true parameter value that generates i.i.d. copies $(Y^i, \mathbf{X}^i, \mathbf{Z}^i)$ of $(Y, \mathbf{X}, \mathbf{Z})$. The most difficult submodel $m(\,\cdot\,, \boldsymbol{\theta})$ of the nonparametric part of (1.1) for estimating $\boldsymbol{\theta}^0$ is given by

$$\begin{aligned} m^*(\,\cdot\,, \boldsymbol{\theta}) &= m^0 - (\boldsymbol{\theta} - \boldsymbol{\theta}^0)^\top \Pi\left(\mathrm{E}(\mathbf{X}|\mathbf{Z} = \cdot)|\mathcal{H}\right) \\ &= \Pi\left(\mathrm{E}(Y|\mathbf{Z} = \cdot)|\mathcal{H}\right) - \boldsymbol{\theta}^\top \Pi\left(\mathrm{E}(\mathbf{X}|\mathbf{Z} = \cdot)|\mathcal{H}\right). \end{aligned} \tag{2.1}$$

The second idenity in (2.1) follows from $\mathrm{E}(Y|\mathbf{Z}) = \mathrm{E}(\mathbf{X}|\mathbf{Z})^\top \boldsymbol{\theta}^0 + m^0(\mathbf{Z})$ and the fact that the projection operator is linear. One can then estimate the true parameter $\boldsymbol{\theta}^0$ in the least favorable submodel where the nonparametric additive function $m$ in (1.1) is replaced by $m^*(\,\cdot\,, \boldsymbol{\theta})$ in (2.1). Let $\hat{m}_Y^{\mathrm{add}}$ and $\hat{m}_{X_j}^{\mathrm{add}}$ denote estimators of $\Pi\left(\mathrm{E}(Y|\mathbf{Z} = \cdot)|\mathcal{H}\right)$ and $\eta_j = \Pi\left(\mathrm{E}(X_j|\mathbf{Z} = \cdot)|\mathcal{H}\right)$, respectively. Then, $\hat{m}_{\mathbf{X}}^{\mathrm{add}} \equiv (\hat{m}_{X_1}^{\mathrm{add}}, \ldots, \hat{m}_{X_p}^{\mathrm{add}})^\top$ is an estimator of $\boldsymbol{\eta} = \Pi\left(\mathrm{E}(\mathbf{X}|\mathbf{Z} = \cdot)|\mathcal{H}\right)$. Plugging in $\hat{m}_Y^{\mathrm{add}} - \boldsymbol{\theta}^\top \hat{m}_{\mathbf{X}}^{\mathrm{add}}$ as an estimator of the least favorable curve $m^*(\,\cdot\,, \boldsymbol{\theta})$ in the least squares criterion, one can estimate $\boldsymbol{\theta}^0$ by

$$\begin{aligned} \hat{\boldsymbol{\theta}} &= \arg\min_{\boldsymbol{\theta}} \sum_{i=1}^n \left\{ Y^i - \hat{m}_Y^{\mathrm{add}}(\mathbf{Z}^i) - \boldsymbol{\theta}^\top \left( \mathbf{X}^i - \hat{m}_{\mathbf{X}}^{\mathrm{add}}(\mathbf{Z}^i) \right) \right\}^2 \\ &= \left( n^{-1} \sum_{i=1}^n \tilde{\mathbf{X}}^i \tilde{\mathbf{X}}^{i\top} \right)^{-1} n^{-1} \sum_{i=1}^n \tilde{\mathbf{X}}^i \tilde{Y}^i, \end{aligned} \tag{2.2}$$

where $\tilde{\mathbf{X}}^i = \mathbf{X}^i - \hat{m}_{\mathbf{X}}^{\mathrm{add}}(\mathbf{Z}^i)$ and $\tilde{Y}^i = Y^i - \hat{m}_Y^{\mathrm{add}}(\mathbf{Z}^i)$.

Yu, Mammen and Park (2011) studied the estimator $\hat{\boldsymbol{\theta}}$ when $\hat{m}_Y^{\mathrm{add}}$ and $\hat{m}_{\mathbf{X}}^{\mathrm{add}}$ are obtained by the smooth backfitting technique. This method was proposed by Mammen, Linton and Nielsen (1999) for estimating additive models and found to avoid the curse of dimensionality under weaker conditions than the ordinary backfitting (Opsomer and Ruppert (1997)) and marginal integration (Linton and Nielsen (1995)). The idea of smooth backfitting was successfully implemented for fitting various other structured nonparametric models, see Yu, Park and Mammen (2008) and Lee, Mammen and Park (2010, 2012), among others. For a response variable $W$, and in case $\mathrm{E}(W|\mathbf{Z})$ is not an additive function as in our cases with $W = Y$ and $W = X_j$, the method estimates $\Pi\left(\mathrm{E}(W|\mathbf{Z} = \cdot)\middle|\mathcal{H}\right)$. It gives $\hat{m}_W^{\mathrm{add}}$ as an estimator of $\Pi\left(\mathrm{E}(W|\mathbf{Z} = \cdot)\middle|\mathcal{H}\right)$, where $\hat{m}_W^{\mathrm{add}}(\mathbf{z}) = \hat{m}_{W,0} + \hat{m}_{W,1}(z_1) + \cdots + \hat{m}_{W,d}(z_d)$ and the $d$-tuple $(\hat{m}_{W,j} : 1 \le j \le d)$

solves the system of integral equations

$$\hat{m}_{W,j}(z_j) = \tilde{m}_{W,j}(z_j) - \hat{m}_{W,0} - \sum_{k\neq j}\int_0^1 \hat{m}_{W,k}(z_k)\frac{\hat{p}_{jk}(z_j,z_k)}{\hat{p}_j(z_j)}\,dz_k, \quad 1\le j\le d, \tag{2.3}$$

subject to the constraints $\int_0^1 \hat{m}_{W,j}(z_j)\hat{p}_j(z_j)\,dz_j = 0$, $1\le j\le d$. In these equations, $\hat{m}_{W,0} = n^{-1}\sum_{i=1}^n W^i$, $\tilde{m}_{W,j}$ is a marginal regression estimator of $\mathrm{E}(W|Z_j=\cdot)$, with $\hat{p}_j$ and $\hat{p}_{jk}$ estimators of the marginal density $p_j$ of $Z_j$ and of the joint density $p_{jk}$ of $(Z_j,Z_k)$.

Specifically,

$$\begin{aligned}
\tilde{m}_{W,j}(z_j) &= \hat{p}_j(z_j)^{-1} n^{-1}\sum_{i=1}^n K_h(z_j,Z_j^i)W^i,\\
\hat{p}_j(z_j) &= n^{-1}\sum_{i=1}^n K_{h_j}(z_j,Z_j^i),\\
\hat{p}_{jk}(z_j,z_k) &= n^{-1}\sum_{i=1}^n K_{h_j}(z_j,Z_j^i)K_{h_k}(z_k,Z_k^i).
\end{aligned} \tag{2.4}$$

Here $K_h(z,u)$ is the so-called normalized kernel defined by

$$K_h(z,u) = \frac{K_h(z-u)}{\int_0^1 K_h(t-u)\,dt}, \quad z,u\in[0,1], \tag{2.5}$$

where $K_h(u) = h^{-1}K(u/h)$, $K$ is a baseline kernel function and $h>0$ is the bandwidth. The normalized kernel $K_h(\,\cdot\,,\cdot)$ satisfies $\int_0^1 K_h(z,u)\,dz = 1$ for all $u\in[0,1]$ and it equals the conventional kernel $K_h(z-u)$ for $z\in[2h,1-2h]$. For more details, see Mammen, Linton and Nielsen (1999) and Yu, Park and Mammen (2008). Yu, Mammen and Park (2011) proved that the estimator $\hat{\boldsymbol{\theta}}$ at (2.2) achieves $\sqrt{n}$-consistency if $p_\varepsilon$ has finite second moment, and is semiparametric efficient in case $p_\varepsilon$ is Gaussian.

## 3. Estimation of the Model

In case only the $X_j^i$ are contaminated and we observe $X_j^{*i} = X_j^i + U_j^i$ and $Z_k^i$, one can simply correct for the 'attenuation effect' due to the measurement errors $U_j^i$, in (2.2). Specifically, one can estimate $\boldsymbol{\theta}^0$ by

$$\tilde{\boldsymbol{\theta}} = \left(n^{-1}\sum_{i=1}^n \tilde{\mathbf{X}}^{*i}\tilde{\mathbf{X}}^{*i\top} - \boldsymbol{\Sigma}_{\mathbf{U}}\right)^{-1} n^{-1}\sum_{i=1}^n \tilde{\mathbf{X}}^{*i}\tilde{Y}^i, \tag{3.1}$$

where $\tilde{\mathbf{X}}^{*i} = \mathbf{X}^{*i} - \hat{m}^{\mathrm{add}}_{\mathbf{X}^*}(\mathbf{Z}^i)$ and $\boldsymbol{\Sigma}_{\mathbf{U}}$ is the covariance matrix of $\mathbf{U} = (U_1,\ldots,U_p)^\top$.

Liang, Härdle and Carroll (1999) studied this type of estimator for the case $d = 1$. When $d = 1$, there is no need for backfitting such as at (2.3). In this case, one simply puts $\tilde{\mathbf{X}}^{*i} = \mathbf{X}^{*i} - \tilde{m}_{\mathbf{X}^*}(Z^i)$ and $\tilde{Y}^i = Y^i - \tilde{m}_Y(Z^i)$ in (3.1), where $\tilde{m}_{\mathbf{X}^*} = (\tilde{m}_{X_1^*}, \ldots, \tilde{m}_{X_p^*})^\top$ and $\tilde{m}_W$ with $W = Y$ or $W = X_j^*$ is defined as in (2.4). When $d > 1$, and with the smooth backfitting estimation at (2.3) being applied to $W = X_j^*$ for each $j$, one can prove that the estimator $\tilde{\boldsymbol{\theta}}$ at (3.1) satisfies

$$\sqrt{n}(\tilde{\boldsymbol{\theta}} - \boldsymbol{\theta}^0) \xrightarrow{d} N\left(\mathbf{0}, \mathrm{var}(\varepsilon - \mathbf{U}^\top \boldsymbol{\theta}^0) \cdot \left[\mathrm{E}\{\mathbf{X} - \boldsymbol{\eta}(\mathbf{Z})\}\{\mathbf{X} - \boldsymbol{\eta}(\mathbf{Z})\}^\top\right]^{-1}\right).$$

These results may be obtained by adapting the theory developed in Yu, Mammen and Park (2011) to the correction for attenuation, and using the fact that $\mathrm{E}(\mathbf{X}^*|\mathbf{Z}) = \mathrm{E}(\mathbf{X}|\mathbf{Z})$ so that $\hat{m}_{\mathbf{X}^*}^{\mathrm{add}}$ estimates $\boldsymbol{\eta}$ consistently and has similar asymptotic properties as $\hat{m}_{\mathbf{X}}^{\mathrm{add}}$.

When both $X_j$ and $Z_k$ are contaminated by measurement errors $U_j$ and $V_k$, respectively, the problem is much more complicated. The difficulty arises since $\tilde{m}_{W,j}$, $\hat{p}_j$ and $\hat{p}_{jk}$ at (2.4) with $Z_j^i$ and $Z_k^i$ replaced by the corresponding contaminated $Z_j^{*i}$ and $Z_k^{*i}$ have nonnegligible biases as estimators of $E(W|Z_j = \cdot)$, $p_j$ and $p_{jk}$, respectively. For the contaminated $Z_j^{*i}$ and $Z_k^{*i}$ that are close to points of interest, say $z_j$ and $z_k$, respectively, the corresponding unobserved true predictor values $Z_j^i$ and $Z_k^i$ may be far away from the points $z_j$ and $z_k$ due to measurement errors. Thus, $Z_j^{*i}$ and $Z_k^{*i}$ may not have relevant information about the target functions at $z_j$ and $z_k$, respectively.

When $d = 1$, this difficulty can be resolved by using the deconvolution kernel suggested and studied by Stefanski and Carroll (1990) and Fan and Truong (1993), among others. The salient feature of the deconvolution kernel, denoted by $K^D$, is the 'unbiased scoring' property that

$$\mathrm{E}\{K_h^D(z - Z^*)|Z\} = K_h(z - Z). \tag{3.2}$$

The property (3.2) entails that the bias properties of the kernel estimators with $K_h^D$ based on contaminated predictor values $Z^{*i}$ are the same as those of the estimators with the conventional kernel weight $K_h$ based on the true predictor values $Z^i$. Indeed, Zhu and Cui (2003) proved that the use of a deconvolution kernel, in conjunction with the correction for attenuation as is done in (3.1), gives a $\sqrt{n}$-consistent estimator of $\boldsymbol{\theta}^0$ under suitable conditions.

Han and Park (2018) introduced a special kernel scheme that has both the properties of normalization and unbiased scoring, and we adopt it here. Let $\phi_f$ for a function $f$ denote the Fourier transform of $f$, and $\phi_V$ for a random variable $V$ the characteristic function of $V$. Let

$$\phi_K(t;z) = \int_0^1 e^{it(z-u)/h} K_h(z,u)\,du,$$

where $K_h(z,u)$ is the normalized kernel defined at (2.5). For $z \in [2h, 1-2h]$ one can show that $\phi_K(\,\cdot\,;z) = \phi_K$, the Fourier transform of the baseline kernel $K$. The kernel function of Han and Park (2018) is given by

$$\begin{aligned} K_h^\star(z,z^*) &= \frac{1}{2\pi h}\int_{-\infty}^{\infty} e^{-it(z-z^*)/h} \frac{\phi_K(t;z)\phi_K(t)}{\phi_V(t/h)}\,dt \\ &= \frac{1}{2\pi}\int_{-\infty}^{\infty} e^{-itz^*} \frac{\phi_{K_h(z,\cdot)*K_h}(t)}{\phi_V(-t)}\,dt, \end{aligned} \tag{3.3}$$

where $K_h(z,\cdot) * K_h(u) = \int_{-\infty}^{\infty} K_h(u-t)K_h(z,t)\,dt$. Here $K_h(z,\cdot) * K_h(u) = K_h * K_h(z-u)$ when $z \in [2h, 1-2h]$.

Han and Park (2018) showed that, under (A1) and (A2) in the next section, $K_h^\star(z,z^*)$ at (3.3) is well-defined for all $z \in [0,1]$ and $z^* \in \mathbb{R}$, and satisfies

$$\begin{aligned} &\int_0^1 K_h^\star(z,z^*)\,dx = 1 \quad \text{for all } z^* \in \mathbb{R}, \\ &\mathrm{E}\left\{K_h^\star(z,Z^*)\,\middle|\,Z=u\right\} = K_h(z,\cdot) * K_h(u) \quad \text{for all } z,u \in [0,1], \end{aligned} \tag{3.4}$$

where $Z^* = Z + V$ with $V$ independent of $Z$. The first identity of (3.4) is the normalization property that is essential for the success of the smooth backfitting method, and the second corresponds to the unbiased scoring property (3.2). Thus, the bias properties of the smooth backfitting estimator of $\Pi\left(\mathrm{E}(W|\mathbf{Z}=\cdot)\middle|\mathcal{H}\right)$ based on $\mathbf{Z}^{*i}$ and the kernel scheme $K_h^\star(z,u)$ is the same as those based on the true but unobservable $\mathbf{Z}^i$ and the kernel $K_h(z,\cdot) * K_h(u)$. Recall that $K_h(z,\cdot) * K_h(u) = (K * K)_h(z-u)$ for $z$ in the interior region $[2h, 1-2h]$.

Now we set out our estimator of $\boldsymbol{\theta}^0$. Let

$$\hat{\eta}_j(\mathbf{z}) = \hat{m}_{X_j^*}^{\mathrm{add}}(\mathbf{z}) = \hat{m}_{X_j^*,0} + \hat{m}_{X_j^*,1}(z_1) + \cdots + \hat{m}_{X_j^*,d}(z_d), \tag{3.5}$$

where $\hat{m}_{X_j^*,0} = n^{-1}\sum_{i=1}^n X_j^{*i}$ and $(\hat{m}_{X_j^*,k} : 1 \le k \le d)$ solves the system (2.3) with $W^i = X_j^{*i}$ and $K_{h_j}(z_j, Z_j^i)$ replaced by $K_{h_j}^\star(z_j, Z_j^{*i})$ at (2.4). Put $\hat{\boldsymbol{\eta}} = (\hat{\eta}_1,\ldots,\hat{\eta}_p)^\top$. Likewise, let

$$\hat{\xi}(\mathbf{z}) = \hat{m}_Y^{\mathrm{add}}(\mathbf{z}) = \hat{m}_{Y,0} + \hat{m}_{Y,1}(z_1) + \cdots + \hat{m}_{Y,d}(z_d), \tag{3.6}$$

with $Y^i$ taking the role of $X_j^{*i}$ in the definition of $\hat{\eta}_j(\mathbf{z})$. Here $\hat{\xi}$ is an estimator of $\xi \equiv \Pi(\mathrm{E}(Y|\mathbf{Z}=\cdot)|\mathcal{H})$ in the presence of measurement errors. We want to replace $\tilde{\mathbf{X}}^{*i}$ and $\tilde{Y}^i$ by $\mathbf{X}^{*i} - \hat{\boldsymbol{\eta}}(\mathbf{Z}^i)$ and $Y^i - \hat{\xi}(\mathbf{Z}^i)$, respectively, in (3.1), but this is infeasible since the $\mathbf{Z}^i$ are not observed. Replacing them by $\mathbf{X}^{*i} - \hat{\boldsymbol{\eta}}(\mathbf{Z}^{*i})$ and $Y^i - \hat{\xi}(\mathbf{Z}^{*i})$ would lead to an inconsistent estimator due to the mesaurement errors in $\mathbf{Z}^{*i}$.

In the case of no measurement errors, $n^{-1}\sum_{i=1}^{n}\tilde{\mathbf{X}}^i\tilde{\mathbf{X}}^{i\top}$ in (2.2) targets $\mathbf{D}\equiv E\{\mathbf{X}-\boldsymbol{\eta}(\mathbf{Z})\}\{\mathbf{X}-\boldsymbol{\eta}(\mathbf{Z})\}^\top$. We have

$$\begin{aligned}\mathbf{D}&=\mathrm{E}\{\mathbf{X}^*-\boldsymbol{\eta}(\mathbf{Z})\}\{\mathbf{X}^*-\boldsymbol{\eta}(\mathbf{Z})\}^\top-\boldsymbol{\Sigma}_{\mathbf{U}}\\&=\int_{[0,1]^d}\mathrm{E}\left[\{\mathbf{X}^*-\boldsymbol{\eta}(\mathbf{z})\}\{\mathbf{X}^*-\boldsymbol{\eta}(\mathbf{z})\}^\top\Big|\mathbf{Z}=\mathbf{z}\right]p_{\mathbf{Z}}(\mathbf{z})\,d\mathbf{z}-\boldsymbol{\Sigma}_{\mathbf{U}}.\end{aligned}\tag{3.7}$$

We can estimate the joint density $p_{\mathbf{Z}}(\mathbf{z})$ in (3.7) by $\hat{p}_{\mathbf{Z}}(\mathbf{z})=n^{-1}\sum_{i=1}^{n}K^\star_{\mathbf{g}}(\mathbf{z},\mathbf{Z}^{*i})$, where $K^\star_{\mathbf{g}}(\mathbf{z},\mathbf{Z}^{*i})=K^\star_{g_1}(z_1,Z_1^{*i})\times\cdots\times K^\star_{g_d}(z_d,Z_d^{*i})$ allowing $g_j$ to be different from the bandwidth $h_j$ in the smooth backfitting. Also, we can estimate the conditional expectation inside the integral on the right side of the second equation of (3.7) by the Nadaraya-Watson type estimator

$$\hat{p}_{\mathbf{Z}}(\mathbf{z})^{-1}\cdot n^{-1}\sum_{i=1}^{n}\{\mathbf{X}^{*i}-\hat{\boldsymbol{\eta}}(\mathbf{z})\}\{\mathbf{X}^{*i}-\hat{\boldsymbol{\eta}}(\mathbf{z})\}^\top K^\star_{\mathbf{g}}(\mathbf{z},\mathbf{Z}^{*i}).\tag{3.8}$$

Putting these together with (3.7) we estimate $\mathbf{D}$ by

$$\hat{\mathbf{D}}=n^{-1}\sum_{i=1}^{n}\int_{[0,1]^d}\{\mathbf{X}^{*i}-\hat{\boldsymbol{\eta}}(\mathbf{z})\}\{\mathbf{X}^{*i}-\hat{\boldsymbol{\eta}}(\mathbf{z})\}^\top K^\star_{\mathbf{g}}(\mathbf{z},\mathbf{Z}^{*i})\,d\mathbf{z}-\boldsymbol{\Sigma}_{\mathbf{U}}.\tag{3.9}$$

Similarly, we estimate $E\{\mathbf{X}-\boldsymbol{\eta}(\mathbf{Z})\}\{Y-\xi(\mathbf{Z})\}$, the target of $n^{-1}\sum_{i=1}^{n}\tilde{\mathbf{X}}^i\tilde{Y}^i$ in (2.2), by

$$n^{-1}\sum_{i=1}^{n}\int_{[0,1]^d}\{\mathbf{X}^{*i}-\hat{\boldsymbol{\eta}}(\mathbf{z})\}\{Y^i-\hat{\xi}(\mathbf{z})\}K^\star_{\mathbf{g}}(\mathbf{z},\mathbf{Z}^{*i})\,d\mathbf{z}.\tag{3.10}$$

This gives our proposed estimator of $\boldsymbol{\theta}^0$ as

$$\hat{\boldsymbol{\theta}}=\hat{\mathbf{D}}^{-1}\,n^{-1}\sum_{i=1}^{n}\int_{[0,1]^d}\{\mathbf{X}^{*i}-\hat{\boldsymbol{\eta}}(\mathbf{z})\}\{Y^i-\hat{\xi}(\mathbf{z})\}K^\star_{\mathbf{g}}(\mathbf{z},\mathbf{Z}^{*i})\,d\mathbf{z}.\tag{3.11}$$

In the case where only the $Z_j^i$ are contaminated so that we observe the true predictor values $X_j^i$, we can simply replace $\mathbf{X}^{*i}$ by $\mathbf{X}^i$ in the definitions of $\hat{\boldsymbol{\eta}}$, $\hat{\mathbf{D}}$ and $\hat{\boldsymbol{\theta}}$ at (3.5), (3.9) and (3.11), respectively, and put $\boldsymbol{\Sigma}_{\mathbf{U}}=\mathbf{0}$ in (3.9). Here $\hat{\mathbf{D}}$ and $\hat{\boldsymbol{\theta}}$ involve only two-dimensional integration as $\int_0^1 K^\star_g(z_j,u)\,dz_j=1$ for all $u\in\mathbb{R}$, and both $\hat{\boldsymbol{\eta}}$ and $\hat{\xi}$ are sums of univariate functions.

Once we estimate $\boldsymbol{\theta}^0$ by $\hat{\boldsymbol{\theta}}$, we can estimate the true nonparametric additive function $m^0=m_0^0+m_1^0+\cdots+m_d^0$ by applying the smooth backfitting method of Han and Park (2018) to $Y-\hat{\boldsymbol{\theta}}^\top\mathbf{X}^*$ as the response variable and $\mathbf{Z}^*$ as the contaminated predictor vector. Since the rate of convergence of the parametric estimator $\hat{\boldsymbol{\theta}}$ is faster than the nonparametric rate, the resulting estimators of $m^0$ and its components $m_j^0$ have the same first-order asymptotic properties as the corresponding oracle smooth backfitting estimators obtained by taking $Y-$

$\boldsymbol{\theta}^{0\top}\mathbf{X}^*$ as the response variable and $\mathbf{Z}^*$ as the contaminated predictor vector. The asymptotic properties of the oracle estimators can be easily obtained by adapting the theory developed in Han and Park (2018).

## 4. Theoretical Properties

For simplicity, we assume $h_j \asymp h$ and $g_j \asymp g$. We need some assumptionst.

(A1) There exist some positive constants $\beta$, $c_1$ and $c_2$ such that $c_1(1+|t|)^{-\beta} \le |\phi_{V_j}(t)| \le c_2(1+|t|)^{-\beta}$, and $|t^{\beta+1}\phi'_{V_j}(t)| = O(1)$ as $|t| \to \infty$.

(A2) The baseline kernel function $K$ is supported on $[-1,1]$ and is $\lfloor \beta+1 \rfloor$-times continuously differentiable. $K^{(\ell)}(-1) = K^{(\ell)}(1) = 0$ for $0 \le \ell \le \lfloor \beta \rfloor$, where $\lfloor \beta \rfloor$ denotes the largest integer that is less than or equal to $\beta$. $K^{(\ell)}$ the $\ell$-th derivative of $K$. Also, $\int_0^1 |t^\beta \phi_K(t)|\, dt < \infty$.

(A3) The joint density $p$ of $\mathbf{Z}$ is bounded away from zero and infinity on $[0,1]^d$ and partially continuously differentiable, and the one- and two-dimensional marginal densities $p_j$ and $p_{jk}$ are also (partially) continuously differentiable.

(A4) The $\mathrm{E}(X_j^2 \mid \mathbf{Z} = \cdot)$ are bounded on $[0,1]^d$.

(A5) The $\eta_{j,\ell}$ for $1 \le j \le p$ and $1 \le \ell \le d$ are twice continuously differentiable on $[0,1]$.

(A6) $\mathrm{E}\{\mathbf{X} - \boldsymbol{\eta}(\mathbf{Z})\}\{\mathbf{X} - \boldsymbol{\eta}(\mathbf{Z})\}^\top$ is positive definite.

(A7) There exist constants $C > 0$ such that $\mathrm{E}e^{uW} \le \exp(Cu^2/2)$ for all $u$, for $W = U_j$, $X_j$ and $\varepsilon$.

The conditions (A3)–(A5) are typically assumed in kernel smoothing theory , see Mammen, Linton and Nielsen (1999), Yu, Park and Mammen (2008) and Lee, Mammen and Park (2012), among others. The condition (A6) is assumed for identifiability of $\theta$ in the model (1.1), see also Yu, Mammen and Park (2011) , and (A7) is used to get exponential bounds in our applications of empirical process theory to concentration inequalities. The conditions (A1) and (A2) are usually assumed in deconvolution problems, see Delaigle, Fan and Carroll (2009). They enable us to obtain an inequality enveloping $K_h^\star$ that we use to get bounds for terms involving $K_h^\star$, see Lemma 5.1 in Han and Park (2018).

Put

$$\tau(h;\beta) = \begin{cases} 1 & \beta < 1/2, \\ \sqrt{\log h^{-1}} & \beta = 1/2, \\ h^{1/2-\beta} & \beta > 1/2, \end{cases}$$

$$r_n(g,h;\beta) = n^{-1/2}\tau(g;\beta)^2 + n^{-1/2}hg^{-2\beta}\sqrt{\log n} + n^{-1}h^{-1-2\beta}\tau(h;\beta)^2 \log n.$$

**Theorem 1.** *If* (A1)–(A7) *hold, and if* $nh^{3+2\beta}\tau(h;\beta)^{-2}(\log n)^{-1}$ *is bounded away from zero, then,*

$$\hat{\boldsymbol{\theta}} = \boldsymbol{\theta}^0 + O(g^2) + O(h^3) + O_p(r_n(g,h;\beta)).$$

When only the $Z_j^i$ are contaminated, Theorem 1 remains valid for the modified version of $\hat{\boldsymbol{\theta}}$ that we described after (3.11). We can derive the rates of convergence of $\hat{\boldsymbol{\theta}}$ from Theorem 1, depending on the smoothness $\beta$ of the distributions of the measurement errors $V_j$.

In case $\beta < 1/2$,

$$\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^0 = O(g^2) + O(h^3) + O_p\left(n^{-1/2} + n^{-1/2}hg^{-2\beta}\sqrt{\log n} + n^{-1}h^{-1-2\beta}\log n\right).$$

Let $h \asymp n^{-a}$ and $g \asymp n^{-b}$ for $a, b > 0$. If we choose $a$ and $b$ so that $1/4 \le b < a/(2\beta)$ and $\max\{1/6, \beta/2\} < a < 1/(3+2\beta)$, then $\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^0 = O_p(n^{-1/2})$. If $\beta = 1/2$, we get $\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^0 = O_p(n^{-1/2}\log n)$ by choosing $h \asymp g \asymp n^{-1/4}\sqrt{\log n}$.

The case where $\beta > 1/2$ is more involved. We get from Theorem 1 that

$$\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^0 = O(g^2) + O(h^3) + O_p\left(n^{-1/2}g^{1-2\beta} + n^{-1/2}hg^{-2\beta}\sqrt{\log n} + n^{-1}h^{-4\beta}\log n\right).$$

The best rate here is $O_p(n^{-1/(1+2\beta)}\sqrt{\log n})$, achieved by choosing $h \asymp g \asymp n^{-1/(2+4\beta)}(\log n)^{1/4}$. This size of $h$ satisfies the condition in Theorem 1. To see that it is the best rate, we let $h \asymp n^{-a}$ and $g \asymp n^{-b}$ up to a factor of size $\log n$ or its power. First consider that $b \le a$. By trading off $g^2$ and $n^{-1/2}g^{1-2\beta}$, we get the optimal order of $g$, $n^{-1/(2+4\beta)}$. This gives $g^2 + n^{-1/2}g^{1-2\beta} = n^{-1/(1+2\beta)}$. The term of order $n^{-1}h^{-4\beta}$ can achieve this rate only when $a \le 1/(2+4\beta)$. This implies that the choice $a = b = 1/(2+4\beta)$ gives the best rate $n^{-1/(1+2\beta)}$ up to a logarithmic factor among all $b \le a$. If $b > a$, we can trade off $n^{-1}h^{-4\beta}$ and $h^3$ to get the best rate for the sum of the two terms, which gives $a = 1/(3+4\beta)$ and the rate $n^{-3/(3+4\beta)}$. For $\hat{\boldsymbol{\theta}}$ to achieve the latter rate, $n^{-1/2}hg^{-2\beta}$ must be smaller than or equal to $n^{-3/(3+4\beta)}$, but this is impossible for any choice of $b > 1/(3+4\beta)$. One can find that trading off other combinations of the four terms $g^2$, $h^3$, $n^{-1/2}hg^{-2\beta}$ and $n^{-1}h^{-4\beta}$ do not lead to a rate for $\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^0$ faster than $n^{-1/(1+2\beta)}$.

**Theorem 2.** *If the conditions in Theorem* 1 *hold, and* $\beta < 1/2$, *then* $\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^0 =$

$O_p(n^{-1/2})$ *if* $h \asymp n^{-a}$ *and* $g \asymp n^{-b}$ *with* $\max\{1/6, \beta/2\} < a < 1/(3+2\beta)$ *and* $1/4 \le b < a/(2\beta)$. *If* $\beta = 1/2$, $\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^0 = O_p(n^{-1/2}\log n)$ *if* $h \asymp g \asymp n^{-1/4}\sqrt{\log n}$. *If* $\beta > 1/2$, $\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}^0 = O_p(n^{-1/(1+2\beta)}\sqrt{\log n})$ *if* $h \asymp g \asymp n^{-1/(2+4\beta)}(\log n)^{1/4}$.

Next, we discuss the rates of convergence of the estimators of the nonparametric component functions $m_j^0$ described at the end of Section 3. Let $h^0$ denote the bandwidth in the smooth backfitting with $Y - \hat{\boldsymbol{\theta}}^{*\top}\mathbf{X}^*$ as the response variable and $\mathbf{Z}^*$ as the contaminated predictor vector. Then, by choosing the bandwidth size $h^0 \asymp n^{-1/(5+2\beta)}$ we get that, for $1 \le j \le d$,

$$\sup_{2h^0 \le z_j \le 1-2h^0} \left|\hat{m}_j(z_j) - m_j^0(z_j)\right| = O_p\left(n^{-2/(5+2\beta)}\sqrt{\log n}\right),$$

$$\sup_{0 \le z_j \le 1} \left|\hat{m}_j(z_j) - m_j^0(z_j)\right| = O_p\left(n^{-1/(5+2\beta)}\right)$$

when $\beta < 1/2$. For $\beta > 1/2$, it holds that by choosing $h^0 \asymp n^{-1/(4+4\beta)}$,

$$\sup_{2h^0 \le z_j \le 1-2h^0} \left|\hat{m}_j(z_j) - m_j^0(z_j)\right| = O_p\left(n^{-1/(2+2\beta)}\sqrt{\log n}\right),$$

$$\sup_{0 \le z_j \le 1} \left|\hat{m}_j(z_j) - m_j^0(z_j)\right| = O_p\left(n^{-1/(4+4\beta)}\right).$$

If $\beta = 1/2$, we get the rates $n^{-1/3}\log n$ in the interior and $n^{-1/6}\sqrt{\log n}$ on the boundary with $h^0 \asymp n^{-1/6}$. These results follow basically from the fact that the estimation error of $\hat{\boldsymbol{\theta}}$ in Theorem 2 is of smaller order than the nonparametric rate.

*Proof of Theorem* 1. Write $\hat{\eta}_j(\mathbf{z}) = \hat{\eta}_{j,0} + \hat{\eta}_{j,1}(z_1) + \cdots + \hat{\eta}_{j,d}(z_d)$, and recall the constraints $\int_0^1 \hat{\eta}_{j,\ell}(z_\ell)\hat{p}_\ell(z_\ell)\,dz_\ell = 0$ on $\hat{\eta}_{j,\ell}$, $1 \le \ell \le d$. Likewise, write $\eta_j(\mathbf{z}) = \eta_{j,0} + \eta_{j,1}(z_1) + \cdots + \eta_{j,d}(z_d)$ with the constraints $\int_0^1 \eta_{j,\ell}(z_\ell)p_\ell(z_\ell)\,dz_\ell = 0$, $1 \le \ell \le d$. We also write $\xi(\mathbf{z}) = \xi_0 + \xi_1(z_1) + \cdots + \xi_d(z_d)$ and $\hat{\xi}(\mathbf{z}) = \hat{\xi}_0 + \hat{\xi}_1(z_1) + \cdots + \hat{\xi}_d(z_d)$ with the corresponding constraints on $\xi_\ell$ and $\hat{\xi}_\ell$ for $1 \le \ell \le d$. In proceeding, we take $\eta_{j,0} = \xi_0 = 0$ as well as $m_0^0 = 0$ and ignore $\hat{\eta}_{j,0}$ and $\hat{\xi}_0$, for simplicity, since their estimation errors are of smaller order than those of the nonparametric estimators $\hat{\eta}_{j,\ell}$ and $\hat{\xi}_\ell$.

**Lemma 1.** *Under the conditions of Theorem* 1,

$$\begin{aligned} &n^{-1}\sum_{i=1}^{n}\int_{[0,1]^d}\{X_j^i - \eta_j(\mathbf{z})\}\{\hat{\eta}_{k,\ell}(z_\ell) - \eta_{k,\ell}(z_\ell)\}K_g^\star(\mathbf{z}, \mathbf{Z}^{*i})\,d\mathbf{z} \\ &= O_p\left(g^2 h + n^{-1/2}hg^{-2\beta}\sqrt{\log n}\right) \end{aligned}$$

*for all* $1 \le j, k \le p$ *and* $1 \le \ell \le d$.

**Lemma 2.** *Under the conditions of Theorem* 1,

$$\mathrm{E}\left|\int_{[0,1]^d}\{X_j-\eta_j(\mathbf{z})\}K_g^\star(\mathbf{z},\mathbf{Z}^*)\,d\mathbf{z}\right|^2=O\left(\tau(g;\beta)^2\right)$$

*for all* $1\le j\le p$.

**Lemma 3.** *Under the conditions of Theorem* 1,

$$\mathrm{var}\left(\int_{[0,1]^d}\{X_j-\eta_j(\mathbf{z})\}\{X_k-\eta_k(\mathbf{z})\}K_g^\star(\mathbf{z},\mathbf{Z}^*)\,d\mathbf{z}\right)=O\left(\tau(g;\beta)^4\right)$$

*for all* $1\le j,k\le d$.

**Lemma 4.** *Under the conditions of Theorem* 1,

$$n^{-1}\sum_{i=1}^n U_j^i\int_0^1\{\hat\eta_{k,\ell}(z_\ell)-\eta_{k,\ell}(z_\ell)\}K_g^\star(z_\ell,Z_\ell^{*i})\,dz_\ell=O_p\left(n^{-1/2}hg^{-\beta}\sqrt{\log n}\right)$$

*for all* $1\le j,k\le p$ *and* $1\le\ell\le d$.

We turn to the proof of Theorem 1. Let $\hat m^{\mathrm{ora}}(\mathbf{z})=\hat\xi(\mathbf{z})-\hat\eta(\mathbf{z})^\top\boldsymbol{\theta}^0$. This is an additive function and is an oracle estimator of the true additive function $m^0(\mathbf{z})=m_1^0(z_1)+\cdots+m_d^0(z_d)$. To see this, write $\xi(\mathbf{z})-\eta(\mathbf{z})^\top\boldsymbol{\theta}^0=\Pi\left(\mathrm{E}(Y-\mathbf{X}^\top\boldsymbol{\theta}^0|\mathbf{Z}=\cdot)\middle|\mathcal{H}\right)(\mathbf{z})=m^0(\mathbf{z})$, and take

$$\hat{\boldsymbol{\delta}}=n^{-1}\sum_{i=1}^n\int_{[0,1]^d}\{\mathbf{X}^{*i}-\hat{\boldsymbol{\eta}}(\mathbf{z})\}\{Y^i-\boldsymbol{\theta}^{0\top}\mathbf{X}^{*i}-\hat m^{\mathrm{ora}}(\mathbf{z})\}K_{\mathbf{g}}^\star(\mathbf{z},\mathbf{Z}^{*i})\,d\mathbf{z}+\boldsymbol{\Sigma}_{\mathbf{U}}\boldsymbol{\theta}^0.$$

Then $\hat{\boldsymbol{\theta}}=\boldsymbol{\theta}^0+\hat{\mathbf{D}}^{-1}\cdot\hat{\boldsymbol{\delta}}$. We show that $\hat{\mathbf{D}}=\mathbf{D}+o_p(1)$, and analyze the size of $\hat{\boldsymbol{\delta}}$.

We first approximate $\hat{\mathbf{D}}$. Decompose $\hat{\mathbf{D}}$ as $\hat{\mathbf{D}}=\hat{\mathbf{D}}_1+\hat{\mathbf{D}}_2+\hat{\mathbf{D}}_3+\hat{\mathbf{D}}_4$, where

$$\hat{\mathbf{D}}_1=n^{-1}\sum_{i=1}^n\int_{[0,1]^d}\{\mathbf{X}^i-\hat{\boldsymbol{\eta}}(\mathbf{z})\}\{\mathbf{X}^i-\hat{\boldsymbol{\eta}}(\mathbf{z})\}^\top K_g^\star(\mathbf{z},\mathbf{Z}^{*i})\,d\mathbf{z},$$

$$\hat{\mathbf{D}}_2=n^{-1}\sum_{i=1}^n\mathbf{U}^i\cdot\int_{[0,1]^d}\{\mathbf{X}^i-\hat{\boldsymbol{\eta}}(\mathbf{z})\}^\top K_g^\star(\mathbf{z},\mathbf{Z}^{*i})\,d\mathbf{z},$$

$$\hat{\mathbf{D}}_3=n^{-1}\sum_{i=1}^n\int_{[0,1]^d}\{\mathbf{X}^i-\hat{\boldsymbol{\eta}}(\mathbf{z})\}K_g^\star(\mathbf{z},\mathbf{Z}^{*i})\,d\mathbf{z}\cdot\mathbf{U}^{i\top},$$

$$\hat{\mathbf{D}}_4=n^{-1}\sum_{i=1}^n\mathbf{U}^i\mathbf{U}^{i\top}-\boldsymbol{\Sigma}_{\mathbf{U}}.$$

It is clear that $\hat{\mathbf{D}}_4=O_p(n^{-1/2})$. Using Lemmas 2 and 4, we can show that both $\hat{\mathbf{D}}_2$ and $\hat{\mathbf{D}}_3$ are of order $O_p\left(n^{-1/2}\tau(g;\beta)+n^{-1/2}hg^{-\beta}\sqrt{\log n}\right)$. We claim

$$\hat{\mathbf{D}}_1=\mathbf{D}+O(g^2)+O(h^3)+O_p(r_n(g,h;\beta)).\tag{4.1}$$

To prove (4.1), we further decompose $\hat{\mathbf{D}}_1$ into four terms as $\hat{\mathbf{D}}_1 = \hat{\mathbf{D}}_{11} + \hat{\mathbf{D}}_{12} + \hat{\mathbf{D}}_{13} + \hat{\mathbf{D}}_{14}$, where

$$\hat{\mathbf{D}}_{11} = n^{-1} \sum_{i=1}^{n} \int_{[0,1]^d} \{\mathbf{X}^i - \boldsymbol{\eta}(\mathbf{z})\}\{\mathbf{X}^i - \boldsymbol{\eta}(\mathbf{z})\}^\top K_g^\star(\mathbf{z}, \mathbf{Z}^{*i})\, d\mathbf{z},$$
$$\hat{\mathbf{D}}_{12} = n^{-1} \sum_{i=1}^{n} \int_{[0,1]^d} \{\mathbf{X}^i - \boldsymbol{\eta}(\mathbf{z})\}\{\boldsymbol{\eta}(\mathbf{z}) - \hat{\boldsymbol{\eta}}(\mathbf{z})\}^\top K_g^\star(\mathbf{z}, \mathbf{Z}^{*i})\, d\mathbf{z},$$
$$\hat{\mathbf{D}}_{13} = n^{-1} \sum_{i=1}^{n} \int_{[0,1]^d} \{\boldsymbol{\eta}(\mathbf{z}) - \hat{\boldsymbol{\eta}}(\mathbf{z})\}\{\mathbf{X}^i - \boldsymbol{\eta}(\mathbf{z})\}^\top K_g^\star(\mathbf{z}, \mathbf{Z}^{*i})\, d\mathbf{z},$$
$$\hat{\mathbf{D}}_{14} = \int_{[0,1]^d} \{\hat{\boldsymbol{\eta}}(\mathbf{z}) - \boldsymbol{\eta}(\mathbf{z})\}\{\hat{\boldsymbol{\eta}}(\mathbf{z}) - \boldsymbol{\eta}(\mathbf{z})\}^\top \hat{p}_{\mathbf{Z}}(\mathbf{z})\, d\mathbf{z}.$$

By Lemma 1, both $\hat{\mathbf{D}}_{12}$ and $\hat{\mathbf{D}}_{13}$ are of order $O_p\left(g^2 h + n^{-1/2} h g^{-2\beta} \sqrt{\log n}\right)$.

For $\hat{\mathbf{D}}_{11}$, we get the magnitude of its variance from Lemma 3. We compute $\mathrm{E}(\hat{\mathbf{D}}_{11})$. Put $\tilde{K}_h(z,u) = K_h(z,\cdot) * K_h(u)$ and $\tilde{K}_h(\mathbf{z},\mathbf{u}) = \tilde{K}_h(z_1,u_1) \times \cdots \times \tilde{K}_h(z_d,u_d)$, with a slight abuse of notation. We observe that

$$\mathrm{E}(\hat{\mathbf{D}}_{11}) = \mathbf{D} + \mathrm{E} \int_{[0,1]^d} \{\boldsymbol{\eta}(\mathbf{Z}) - \boldsymbol{\eta}(\mathbf{z})\}\{\boldsymbol{\eta}(\mathbf{Z}) - \boldsymbol{\eta}(\mathbf{z})\}^\top \tilde{K}_g(\mathbf{z}, \mathbf{Z})\, d\mathbf{z}. \tag{4.2}$$

The identity (4.2) follows from the unbiased scoring property of $K_g^\star$ and

$$\mathrm{E}\{\mathbf{X} - \boldsymbol{\eta}(\mathbf{Z})\} \int_{[0,1]^d} \{\boldsymbol{\eta}(\mathbf{Z}) - \boldsymbol{\eta}(\mathbf{z})\}^\top \tilde{K}_g(\mathbf{z}, \mathbf{Z})\, d\mathbf{z} = \mathbf{0}.$$

The latter holds since $\mathrm{E}(X_j|\mathbf{Z} = \cdot) - \eta_j(\,\cdot\,)$, the projection of $\mathrm{E}(X_j|\mathbf{Z} = \cdot)$ onto $\mathcal{H}^\perp$ in the space of square integrable functions, is orthogonal to

$$\int \{\eta_k(\,\cdot\,) - \eta_k(\mathbf{z})\} \tilde{K}_g(\mathbf{z}, \cdot)\, d\mathbf{z} = \sum_{\ell=1}^{d} \int (\eta_{k,\ell}(\,\cdot\,) - \eta_{k,\ell}(z_\ell)) \tilde{K}_g(z_\ell, \cdot)\, dz_\ell \in \mathcal{H}.$$

From the standard theory of kernel smoothing, the second term in (4.2) is of magnitude $g^2$. This shows

$$\hat{\mathbf{D}}_{11} = \mathbf{D} + O(g^2) + O_p(n^{-1/2} \tau(g;\beta)^2). \tag{4.3}$$

Now, consider $\hat{\mathbf{D}}_{14}$. From Theorems 2 and 3 in Han and Park (2018), we get that, for $1 \le \ell \le d$,

$$\begin{aligned} \sup_{u \in [2h, 1-2h]} |\hat{\eta}_{j,\ell}(u) - \eta_{j,\ell}(u)| &= O_p\left(h^2 + n^{-1/2} h^{-1/2-\beta} \tau(h,\beta) \sqrt{\log n}\right), \\ \sup_{u \in [0,1]} |\hat{\eta}_{j,\ell}(u) - \eta_{j,\ell}(u)| &= O_p\left(h + n^{-1/2} h^{-1/2-\beta} \tau(h,\beta) \sqrt{\log n}\right). \end{aligned} \tag{4.4}$$

Here $\hat{\mathbf{D}}_{14}$ involves only one- and two-dimensional integrals because of the additivity of $\hat{\eta}_j(\mathbf{z})$ and $\eta_j(\mathbf{z})$. From (4.4) we get that the one-dimensional integrals are of order $O_p\left(h^3 + n^{-1}h^{-1-2\beta}\tau(h,\beta)^2\log n\right)$ since the length of the boundary region equals $4h$. The two-dimensional integrals have the magnitudes $O_p\left(h^4 + n^{-1}h^{-1-2\beta}\tau(h,\beta)^2\log n\right)$. This gives

$$\hat{\mathbf{D}}_{14} = O_p\left(h^3 + n^{-1}h^{-1-2\beta}\tau(h,\beta)^2\log n\right). \tag{4.5}$$

This completes the proof of (4.1) and establishes that

$$\hat{\mathbf{D}} = \mathbf{D} + O(g^2) + O(h^3) + O_p(r_n(g,h;\beta)). \tag{4.6}$$

To analyze the size of $\hat{\boldsymbol{\delta}}$, we decompose it into four terms, $\hat{\boldsymbol{\delta}} = \hat{\boldsymbol{\delta}}_1 + \hat{\boldsymbol{\delta}}_2 + \hat{\boldsymbol{\delta}}_3 + \hat{\boldsymbol{\delta}}_4$, where

$$\begin{aligned}
\hat{\boldsymbol{\delta}}_1 &= n^{-1}\sum_{i=1}^n(\varepsilon^i - \mathbf{U}^{i\top}\boldsymbol{\theta}^0)\cdot\int_{[0,1]^d}\{\mathbf{X}^i - \hat{\boldsymbol{\eta}}(\mathbf{z})\}K_g^\star(\mathbf{z},\mathbf{Z}^{*i})\,d\mathbf{z},\\
\hat{\boldsymbol{\delta}}_2 &= n^{-1}\sum_{i=1}^n\int_{[0,1]^d}\{\mathbf{X}^i - \hat{\boldsymbol{\eta}}(\mathbf{z})\}\{m^0(\mathbf{Z}^i) - \hat{m}^{\mathrm{ora}}(\mathbf{z})\}K_g^\star(\mathbf{z},\mathbf{Z}^{*i})\,d\mathbf{z},\\
\hat{\boldsymbol{\delta}}_3 &= n^{-1}\sum_{i=1}^n\mathbf{U}^i\varepsilon^i - n^{-1}\sum_{i=1}^n(\mathbf{U}^i\mathbf{U}^{i\top} - \boldsymbol{\Sigma}_{\mathbf{U}})\boldsymbol{\theta}^0,\\
\hat{\boldsymbol{\delta}}_4 &= n^{-1}\sum_{i=1}^n\mathbf{U}^i\cdot\int_{[0,1]^d}\{m^0(\mathbf{Z}^i) - \hat{m}^{\mathrm{ora}}(\mathbf{z})\}K_g^\star(\mathbf{z},\mathbf{Z}^{*i})\,d\mathbf{z}.
\end{aligned}$$

For the first term $\hat{\boldsymbol{\delta}}_1$, $n^{-1}\sum_{i=1}^n(\varepsilon^i - \mathbf{U}^{i\top}\boldsymbol{\theta}^0)\{X_j^i - \eta_j(\mathbf{Z}^i)\} = O_p(n^{-1/2})$, and

$$\begin{aligned}
&n^{-1}\sum_{i=1}^n(\varepsilon^i - \mathbf{U}^{i\top}\boldsymbol{\theta}^0)\int_0^1\{\eta_{j,\ell}(Z_\ell^i) - \eta_{j,\ell}(z_\ell)\}K_g^\star(z_\ell,Z_\ell^{*i})\,dz_\ell\\
&= O_p(n^{-1/2}\tau(g;\beta)),\\
&n^{-1}\sum_{i=1}^n(\varepsilon^i - \mathbf{U}^{i\top}\boldsymbol{\theta}^0)\int_0^1\{\hat{\eta}_{j,\ell}(z_\ell) - \eta_{j,\ell}(z_\ell)\}K_g^\star(z_\ell,Z_\ell^{*i})\,dz_\ell\\
&= O_p(n^{-1/2}hg^{-\beta}\sqrt{\log n}).
\end{aligned} \tag{4.7}$$

The first result of (4.7) follows from the fact that the second moment of the integral $\int_0^1\{\eta_{j,\ell}(Z_\ell^i) - \eta_{j,\ell}(z_\ell)\}K_g^\star(z_\ell,Z_\ell^{*i})\,dz_\ell$ is of size $O(\tau(g,\beta)^2)$, which can be proved as in the proof of Theorem 3.2 in Han and Park (2018). The second result is the direct consequence of an application of Lemma 4. Clearly, $\hat{\boldsymbol{\delta}}_3 = O_p(n^{-1/2})$. For the fourth term $\hat{\boldsymbol{\delta}}_4$, decomposing $m_j^0(Z_j^i) - \hat{m}_j^{\mathrm{ora}}(z_j)$ as $m_j^0(Z_j^i) - m_j^0(z_j)$ and $\hat{m}_j^{\mathrm{ora}}(z_j) - m_j^0(z_j)$, and using the arguments for deriving (4.7), gives $\hat{\boldsymbol{\delta}}_4 = O_p\left(n^{-1/2}\tau(g;\beta) + n^{-1/2}hg^{-\beta}\sqrt{\log n}\right)$. Thus, we have

$$\hat{\boldsymbol{\delta}}_1 + \hat{\boldsymbol{\delta}}_3 + \hat{\boldsymbol{\delta}}_4 = O_p\left(n^{-1/2}\tau(g;\beta) + n^{-1/2}hg^{-\beta}\sqrt{\log n}\right). \tag{4.8}$$

The analysis of $\hat{\boldsymbol{\delta}}_2$ is similar to that of $\hat{\mathbf{D}}_1$. We claim

$$\hat{\boldsymbol{\delta}}_2 = O(g^2) + O(h^3) + O_p(r_n(g,h;\beta)). \tag{4.9}$$

This and (4.8) establishes

$$\hat{\boldsymbol{\delta}} = O(g^2) + O(h^3) + O_p(r_n(g,h;\beta)). \tag{4.10}$$

To prove (4.9) we decompose $\hat{\boldsymbol{\delta}}_2$ further into four terms, $\hat{\boldsymbol{\delta}}_2 = \hat{\boldsymbol{\delta}}_{21} + \hat{\boldsymbol{\delta}}_{22} + \hat{\boldsymbol{\delta}}_{23} + \hat{\boldsymbol{\delta}}_{24}$, where

$$\begin{aligned}
\hat{\boldsymbol{\delta}}_{21} &= n^{-1}\sum_{i=1}^{n}\int_{[0,1]^d}\{\mathbf{X}^i - \boldsymbol{\eta}(\mathbf{z})\}\{m^0(\mathbf{Z}^i) - m^0(\mathbf{z})\}K_g^\star(\mathbf{z},\mathbf{Z}^{*i})\,d\mathbf{z},\\
\hat{\boldsymbol{\delta}}_{22} &= n^{-1}\sum_{i=1}^{n}\int_{[0,1]^d}\{\mathbf{X}^i - \boldsymbol{\eta}(\mathbf{z})\}\{m^0(\mathbf{z}) - \hat{m}^{\mathrm{ora}}(\mathbf{z})\}K_g^\star(\mathbf{z},\mathbf{Z}^{*i})\,d\mathbf{z},\\
\hat{\boldsymbol{\delta}}_{23} &= n^{-1}\sum_{i=1}^{n}\int_{[0,1]^d}\{\hat{\boldsymbol{\eta}}(\mathbf{z}) - \boldsymbol{\eta}(\mathbf{z})\}\{m^0(\mathbf{z}) - m^0(\mathbf{Z}^i)\}K_g^\star(\mathbf{z},\mathbf{Z}^{*i})\,d\mathbf{z},\\
\hat{\boldsymbol{\delta}}_{24} &= \int_{[0,1]^d}\{\hat{\boldsymbol{\eta}}(\mathbf{z}) - \boldsymbol{\eta}(\mathbf{z})\}\{\hat{m}^{\mathrm{ora}}(\mathbf{z}) - m^0(\mathbf{z})\}\hat{p}_{\mathbf{Z}}(\mathbf{z})\,d\mathbf{z}.
\end{aligned}$$

For $\hat{\boldsymbol{\delta}}_{21}$, a version of Lemma 3 gives $\mathrm{var}(\hat{\boldsymbol{\delta}}_{21}) = O(n^{-1}\tau(g;\beta)^4)$. By similar arguments as those leading to (4.2) and from the standard theory of kernel smoothing, we get

$$\mathrm{E}(\hat{\boldsymbol{\delta}}_{21}) = \mathrm{E}\int_{[0,1]^d}\{\boldsymbol{\eta}(\mathbf{Z}) - \boldsymbol{\eta}(\mathbf{z})\}\{m^0(\mathbf{Z}) - m^0(\mathbf{z})\}\tilde{K}_g(\mathbf{z},\mathbf{Z})\,d\mathbf{z} = O(g^2).$$

This shows

$$\hat{\boldsymbol{\delta}}_{21} = O(g^2) + O_p(n^{-1/2}\tau(g;\beta)^2). \tag{4.11}$$

Furthermore, a version of Lemma 1 entails

$$\hat{\boldsymbol{\delta}}_{22} = O_p\left(g^2h + n^{-1/2}hg^{-2\beta}\sqrt{\log n}\right) = \hat{\boldsymbol{\delta}}_{23}. \tag{4.12}$$

Finally, using similar arguments as those in deriving (4.5) we get

$$\hat{\boldsymbol{\delta}}_{24} = O_p\left(h^3 + n^{-1}h^{-1-2\beta}\tau(h,\beta)^2\log n\right). \tag{4.13}$$

The results (4.11)–(4.13) establishes (4.9). This completes the proof of Theorem 1. $\square$

## 5. Numerical Properties

We evaluated the finite sample performance of $\hat{\boldsymbol{\theta}}$ defined at (3.11). For this we considered a simulation setting similar to the one in Yu, Mammen and Park (2011) . We generated the responses $Y^i$ by

$$Y^i = 3 + \theta_1^0 X_1^i + \theta_2^0 X_2^i + m_1(Z_1^i) + m_2(Z_2^i) + \varepsilon^i, \tag{5.1}$$

where $\varepsilon^i \sim N(0,1)$, $\boldsymbol{\theta}^0 = (1.5, 0.8)^\top$ and

$$m_1(u) = \sin(2\pi(u-0.5)), \quad m_2(u) = (u-0.5) + \sin(2\pi(u-0.5)).$$

The predictor vectors $\mathbf{Z}^i$ were $N((0.5,0.5)^\top, \boldsymbol{\Gamma})$ truncated on $[0,1]^2$ with $\boldsymbol{\Gamma} = \{(1-\rho)\cdot \mathbf{I} + \rho \cdot \mathbf{1}\mathbf{1}^\top\}/4$, $\rho = 0.3$, $\mathbf{1} = (1,1)^\top$ and $\mathbf{I}$ being the identity matrix. We took $X_1^i = Z_1^i(1-2Z_2^{i2}) + \delta^i$ with $\delta$ being i.i.d. $N(0,1)$ and $X_2^i \sim N(0,1)$. We generated the contaminated predictors by

$$\mathbf{X}^{*i} = \mathbf{X}^i + \mathbf{U}^i, \quad \mathbf{Z}^{*i} = \mathbf{Z}^i + \mathbf{V}^i,$$

where $\mathbf{U}^i \sim N(\mathbf{0}, \sigma_U^2 \cdot \mathbf{I})$ with $\sigma_U = 0.3$, and $V_j^i$, $j = 1, 2$, were independent measurement errors having a double gamma difference distribution (Augustyniak and Doray (2012)) with scale parameter 1/7 and smoothness order $\beta = 0.4$.

In our setting, the noise-to-signal ratios (NSR) of $X_j^*$, $\mathrm{var}(U_j)/\mathrm{var}(X_j)$, are 0.080 and 0.090 for $j = 1$ and $j = 2$, respectively. The NSRs for $Z_j^*$ are 0.113 and 0.114. These values of the NSR were obtained by a simulation from a large size sample that were independently generated, because it is difficult to derive the exact variances of a truncated multivariate normal distribution and of its transformations.

For the bandwidth $h$ used in the smooth backfitting for $\hat{\eta}_j$ and $\hat{\xi}$ defined at (3.5) and (3.6), respectively, we took $h = C \cdot n^{-1/(5+2\beta)}$ for $C = 0.25$. The rate $n^{-1/(5+2\beta)}$ of the bandwidth is known to be optimal in nonparametric deconvolution problems, see Han and Park (2018) for example. Our choice of the constant $C = 0.25$ was based on a grid search on $[0.1, 0.3]$. One can make other choices of the bandwidth and this may give better performance, but we do not focus on bandwidth selection in this study. For the bandwidth $g$ that is used in (3.9) and (3.10) we chose $g = h^{3/2}$. This choice equalizes the bias orders $O(h^3)$ and $O(g^2)$ in Theorem 1 that arise in the two types of smoothing with our smoothed and normalized deconvolution kernel $K_h^\star$ and $K_g^\star$, respectively.

We compared our estimator $\hat{\boldsymbol{\theta}}$ with the estimator studied in Yu, Mammen and Park (2011) that ignores the measurement errors in $\mathbf{Z}^*$ as well as in $\mathbf{X}^*$. The latter estimator is defined by (2.2) but with $\tilde{\mathbf{X}}^i$ and $\tilde{Y}^i$ being replaced by $\mathbf{X}^{*i} - \hat{\boldsymbol{\eta}}(\mathbf{Z}^{*i})$ and $Y^i - \hat{\xi}(\mathbf{Z}^{*i})$, respectively, where $\hat{\eta}_j$ and $\hat{\xi}$ are constructed by using the conventional normalized kernel $K_h(\,\cdot\,,\cdot)$ and the contaminated covariate observations $Z_j^{*i}$. We call this estimator $\hat{\boldsymbol{\theta}}^{\mathrm{nve}}$. For the bandwidth $h$ in the

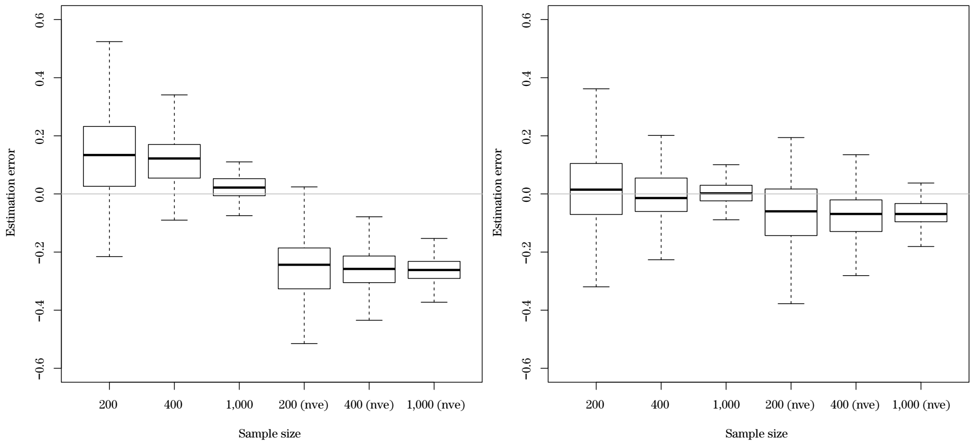


Figure 1. The first three boxplots in each panel are for $\hat{\theta}_j - \theta_j^0$ and the rest for $\hat{\theta}_j^{\mathrm{nve}} - \theta_j^0$, based on 200 pseudo samples of size $n = 200, 400$ and 1,000.

estimation of $\eta_j$ and $\xi$ based on the conventional normalized kernel $K_h(\cdot, \cdot)$, we took $h = C \cdot n^{-1/5}$ and chose $C = 0.3$ by a grid search. The rate $n^{-1/5}$ is known to be optimal in nonparametric univariate function estimation.

We computed $\hat{\boldsymbol{\theta}}$ and $\hat{\boldsymbol{\theta}}^{\mathrm{nve}}$ from $M = 200$ pseudo samples of sizes $n = 200, 400$ and 1,000. Figure 1 depicts the boxplots of the 200 values of the computed $\hat{\theta}_j$ and $\hat{\theta}_j^{\mathrm{nve}}$. We see clearly that our deconvolution-normalization kernel at (3.3) with the correction for the attenuation effect at (3.9) works quite well since the ranges and the central parts of the distributions of $\hat{\theta}_j - \theta_j^0$ shrink toward zero very quickly as the sample size increases. To the contrary, $\hat{\theta}_j^{\mathrm{nve}}$ exhibit persistent non-negligible bias.

We also computed the Monte Carlo estimates of the mean squared errors as

$$\mathrm{MSE}(\hat{\theta}_j) = \frac{1}{M} \sum_{m=1}^{M} \left( \hat{\theta}_j^{(m)} - \theta_j \right)^2,$$

where $\hat{\theta}_j^{(m)}$ denote the values of $\hat{\theta}_j$ computed from the $m$-th pseudo sample. This mean squared error is decomposed into the squared bias and the variance as $\mathrm{MSE}(\hat{\theta}_j) = \mathrm{bias}^2(\hat{\theta}_j) + \mathrm{var}(\hat{\theta}_j)$, where

$$\mathrm{bias}^2(\hat{\theta}_j) = \left( M^{-1} \sum_{m=1}^{M} \hat{\theta}_j^{(m)} - \theta_j^0 \right)^2, \ \mathrm{var}(\hat{\theta}_j) = M^{-1} \sum_{m=1}^{M} \left( \hat{\theta}_j^{(m)} - M^{-1} \sum_{m=1}^{M} \hat{\theta}_j^{(m)} \right)^2.$$

The results are contained in Table 1, where we also give the results for $\hat{\boldsymbol{\theta}}^{\mathrm{nve}}$ for comparison. For our estimator $\hat{\boldsymbol{\theta}}$ we find fast reduction in both the bias and the variance as the sample size increases. The relatively larger bias of $\hat{\theta}_1$ appears to

Table 1. Mean squared errors, squared biases and variances of $\hat{\theta}_j$ and $\hat{\theta}_j^{\text{nve}}$. Based on 200 pseudo samples of sizes $n = 200, 400$ and 1,000.

| sample size & criterion | | $\hat{\theta}_j$ | | $\hat{\theta}_j^{\text{nve}}$ | |
|---|---|---|---|---|---|
| | | $j = 1$ | $j = 2$ | $j = 1$ | $j = 2$ |
| 200 | MSE | 0.0437 | 0.0199 | 0.0753 | 0.0172 |
| | Sq. Bias | 0.0195 | 0.0003 | 0.0618 | 0.0043 |
| | Variance | 0.0242 | 0.0196 | 0.0135 | 0.0129 |
| 400 | MSE | 0.0206 | 0.0071 | 0.0728 | 0.0122 |
| | Sq. Bias | 0.0127 | 0.0000 | 0.0676 | 0.0054 |
| | Variance | 0.0079 | 0.0071 | 0.0052 | 0.0068 |
| 1,000 | MSE | 0.0023 | 0.0016 | 0.0713 | 0.0066 |
| | Sq. Bias | 0.0005 | 0.0000 | 0.0694 | 0.0044 |
| | Variance | 0.0018 | 0.0016 | 0.0019 | 0.0022 |

come from the dependence of the corresponding predictor $X_1$ on the predictors $Z_1$ and $Z_2$ in the nonparametric part of the simulation model. For the naive estimator $\hat{\boldsymbol{\theta}}^{\text{nve}}$ that ignores the measurement errors, we find that there exist intrinsic biases which do not vanish even as the sample size increases.

We also examined what happens if one uses our deconvolution profiling procedure when the covariates are not actually contaminated. For this, we used $(\mathbf{X}^i, \mathbf{Z}^i, Y^i)$ in the construction of $\hat{\boldsymbol{\theta}}$ and $\hat{\boldsymbol{\theta}}^{\text{nve}}$. In this case, $\hat{\boldsymbol{\theta}}^{\text{nve}}$ is an 'oracle' estimator that utilizes the knowledge of no measurement errors in the observed covariates. For $\hat{\boldsymbol{\theta}}$, we took $\mathbf{X}^{*i} = \mathbf{X}^i$, $\mathbf{Z}^{*i} = \mathbf{Z}^i$, $\boldsymbol{\Sigma}_{\mathbf{U}} = (0.3)^2 \cdot \mathbf{I}$ in (3.9). For this estimator we also used the deconvolution-normalization kernels $K_h^\star(\,\cdot\,,\cdot)$ and $K_g^\star(\,\cdot\,,\cdot)$, in (3.9) and (3.11), constructed as if there were measurement errors $V_j^i$ having a double gamma difference distribution with scale parameter 1/7 and smoothness order $\beta = 0.4$. As expected, the MSE properties of $\hat{\boldsymbol{\theta}}^{\text{nve}}$ were superior to those of $\hat{\boldsymbol{\theta}}$ in this case. However, our deconvolution profiling method still worked well in terms of consistent estimation. We found that $\text{MSE}(\hat{\theta}_1) + \text{MSE}(\hat{\theta}_2) = 0.0236$ and 0.0154 for $n = 400$ and $n = 1{,}000$, respectively, while they were 0.0050 and 0.0021 for $\hat{\boldsymbol{\theta}}^{\text{nve}}$.

## Supplementary Materials

Supplement to "Estimation of Errors-in-Variables Partially Linear Additive Models". The supplement contains proofs of Lemmas 1 – 4.

## Acknowledgment

Research of Eun Ryung Lee was supported by the National Research Foundation of Korea (NRF) grant funded by the Korea government (MSIP) (No. NRF-2016R1C1B1011874). Research of Byeong U. Park and Kyunghee Han was supported by the National Research Foundation of Korea (NRF) grants funded by the Korea government (MSIP) (No. NRF-2015R1A2A1A05001753 and No. NRF-2014R1A4A1007895).

## References

Augustyniak, M. and Doray, L. G. (2012). Inference for a leptokurtic symmetric family of distributions represented by the difference of two gamma variates. *Journal of Statistical Computation and Simulation* **82**, 1621–1634.

Bickel, P. J., Klaassen, C. A. J., Ritov, Y. and Wellner, J. A. (1993). *Efficient and Adaptive Estimation for Semiparametric Models*. John Hopkins University Press.

Carroll, R. J., Delaigle, A. and Hall, P. (2007). Non-parametric regression estimation from data contaminated by a mixture of Berkson and classical errors. *Journal of the Royal Statistical Society. Series B. Statistical Methodology.* **69**, 859–878.

Carroll, R. J., Delaigle, A. and Hall, P. (2009). Nonparametric prediction in measurement error models. *Journal of American Statistical Association* **104**, 993–1003.

Carroll, R. J. and Hall, P. (1988). Optimal rates of convergence for deconvolving a density. *Journal of American Statistical Association* **83**, 1184–1186.

Delaigle, A. (2016). Peter Hall's main contributions to deconvolution. *The Annals of Statistics* **44**, 1854–1866.

Delaigle, A., Fan, J. and Carroll, R. J. (2009). A design-adaptive local polynomial estimator for the errors-in-variables problem. *Journal of the Americal Statistical Association* **104**, 348–359.

Delaigle, A. and Hall, P. (2016). Methodology for non-parametric deconvolution when the error distribution is unknow. *Journal of the Royal Statistical Society. Series B. Statistical Methodology.* **78**, 231–252.

Delaigle, A., Hall, P. and Meister, A. (2008). On deconvolution with repeated measurements. *The Annals of Statistics* **36**, 665–685.

Delaigle, A., Hall, P. and Müller, H.-G. (2007). Accelerated convergence for nonparametric regression with coarsened predictors. *The Annals of Statistics* **35**, 2639–2653.

Delaigle, A., Hall, P. and Qiu, P. (2006). Nonparametric methods for solving the Berkson errors-in-variables problem. *Journal of the Royal Statistical Society. Series B. Statistical Methodology.* **68**, 201–220.

Fan, J. and Truong, Y. K. (1993). Nonparametric regression with errors in variables. *The Annals of Statistics* **21**, 1900–1925.

Han, K. and Park, B. U. (2018). Smooth backfitting for errors-in-variables additive models. The Annals of Statistics, in print.

Lee, Y. K., Mammen, E. and Park, B. U. (2010). Backfitting and smooth backfitting for additive quantile models. *The Annals of Statistics* **38**, 2857–2883.

Lee, Y. K., Mammen, E. and Park, B. U. (2012). Flexible generalized varying coefficient regression models. *The Annals of Statistics* **40**, 1906–1933.

Liang, H., Härdle, W. and Carroll, R. J. (1999). Estimation in a semiparametric partially linear errors-in-variables model. *The Annals of Statistics* **27**, 1519–1535.

Linton, O. and Nielsen, J. P. (1995). A kernel method of estimating structured nonparametric regression based on marginal integration. *Biometrika* **82**, 93–100.

Mammen, E., Linton, O. and Nielsen, J. P. (1999). The existence and asymptotic properties of a backfitting projection algorithm under weak conditions. *The Annals of Statistics* **27**, 1443–1490.

Opsomer, J. D. and Ruppert, D. (1997). Fitting a bivariate additive model by local polynomial regression. *The Annals of Statistics* **25** 186–211.

Stefanski, L. A. and Carroll, R. J. (1990). Deconvoluting kernel density estimators. *Statistics* **21**, 169–184.

Yu, K., Park, B. U. and Mammen, E. (2008). Smooth backfitting in generalized additive models. *The Annals of Statistics.* **36**, 228–260.

Yu, K., Mammen, E. and Park, B. U. (2011). Semi-parametric regression: Efficiency gains from modeling the nonparametric part. *Bernoulli* **17**, 736–748.

Zhu, L. and Cui, H. (2003). A semi-parametric regression model with errors in variables. *Scandinavian Journal of Statistics* **30**, 429–442.

Department of Statistics, Sungkyunkwan University, 25-2 Sungkyunkwan-ro, Jongno-gu, Seoul 03063, Korea.

E-mail: silverryuee@gmail.com

Department of Statistics, Mathematical & Sciences Building 4118, University of California, Davis, CA 95616, U.S.A.

E-mail: kyunghee.stat@gmail.com

Department of Statistics, Seoul National University, 1 Gwanak-ro, Gwanak-gu, Seoul 08826, Korea.

E-mail: bupark@stats.snu.ac.kr

(Received February 2017; accepted May 2017)