A SPLINE-ASSISTED SEMIPARAMETRIC APPROACH TO NONPARAMETRIC MEASUREMENT ERROR MODELS

By Fei Jiang* and Yanyuan Ma†

University of Hong Kong and Penn State University†*

Nonparametric estimation of the probability density function of a random variable measured with error is considered to be a difficult problem, in the sense that depending on the measurement error property, the estimation rate can be as slow as the logarithm of the sample size. Likewise, nonparametric estimation of the regression function with errors in the covariate suffers the same possibly slow rate. The traditional methods for both problems are based on deconvolution, where the slow convergence rate is caused by the quick convergence to zero of the Fourier transform of the measurement error density, which, unfortunately, appears in the denominators during the construction of these methods. Using a completely different approach of spline-assisted semiparametric methods, we are able to construct nonparametric estimators of both density functions and regression mean functions that achieve the same nonparametric convergence rate as in the error free case. Other than requiring the error-prone variable distribution to be compactly supported, our assumptions are not stronger than in the classical deconvolution literatures. The performance of these methods are demonstrated through some simulations and a data example.

1. Introduction. Density estimation is a familiar problem in the nonparametric estimation literature. Generally, we observe independent and identically distributed (iid) variables $X_1, \ldots, X_n$ from a distribution with probability density function (pdf) $f_X(x)$ and nonparametric estimators such as kernel methods are available in the literature to estimate $f_X(x)$. Even when the X_i's are not directly observed, nonparametric estimation of $f_X(x)$ can still be carried out based on their surrogates. Specifically, assume that instead of observing X_i, we observe $W_i \equiv X_i + U_i$, where U_i is a mean zero random error independent of X_i and follows a distribution with pdf $f_U(u)$. This problem was studied extensively in the literature and by large the main stream approach is deconvolution (Carroll and Hall, 1988; Liu and Taylor, 1989; Stefanski and Carroll, 1990; Zhang, 1990; Fan, 1991). However, as established in these works, the deconvolution based estimator of $f_X(x)$ may converge very slowly. For example, when the error distribution $f_U(u)$

MSC 2010 subject classifications: Primary 60K35, 60K35; secondary 60K35

Keywords and phrases: errors in covariates, measurement error, semiparametrics, spline

is normal or other distributions that are "super smooth", the deconvolution estimator only converges to $f_X(x)$ at the rate of $\{\log(n)\}^{-k}$ where k is a positive constant. When the error distribution $f_U(u)$ is Laplace or other "ordinary smooth" type, the convergence rate is n^{-k}, where k is a positive constant smaller than 0.2. Here "super smooth" and "ordinary smooth" are characteristics that play an important role in determining the convergence properties of the deconvolution estimators. Because these characteristics are irrelevant in our approach, we skip the precise description and refer to Fan (1991) for details.

Given that when X_i's are observed, nonparametric density estimation usually performs better than these results, it prompts us to ask whether we could further improve the estimation performance for the error-in-variable case. However, the convergence rates described above are shown to be "optimal" in the literature as long as deconvolution is used. Thus, if better convergence is to be achieved, the estimation has to go beyond the deconvolution framework. In other words, novel approaches totally different from the existing literature will be required.

A related problem to nonparametric density estimation is nonparametric regression. Likewise, when iid observations (X_i, Y_i), $i = 1, \ldots, n$, are available, many nonparametric estimators such as kernel and spline based methods have been proposed to estimate the regression mean of Y conditional on X. Here the assumption is that $Y_i = m(X_i) + \epsilon_i$, where for simplicity, we assume ϵ_i is independent of X_i and has mean zero with density $f_\epsilon(\epsilon)$. When X_i is unavailable and instead only W_i described above is available, we encounter the problem of nonparametric regression with measurement error. The traditional way to estimate $m(x)$ is also deconvolution (Fan and Truong, 1993), resulting in the same possibly slow rate of convergence. Because the same reason that results in the slow convergence rate in the density estimation is also responsible for the nonparametric regression estimation, methods that do not rely on deconvolution have to be employed in order to achieve improvement.

Our breakthrough came when it struck us that if the functions that we want to estimate, $f_X(x)$ and $m(x)$, had been parametric, we would have much simpler problems. If these problems can be solved without using deconvolution, then we can approximate $f_X(x)$ and $m(x)$ with B-spline, and operationally we would be dealing with parametric models and deconvolution can be avoided too. Thus, we study both problems in this work with the help of a spline representation. Using spline representation in measurement error models is not entirely new, although it was mostly used in the Bayesian framework (Berry, Carroll and Ruppert, 2002; Staudenmayer, Ruppert and

Buonaccorsi, 2008). While the density estimator is relatively easy to devise, regression estimator turns out to be challenging and requires further semiparametric treatment. A second challenge in both problems is in establishing the convergence rates of the resulting estimators. The common obstacle in both estimators is the fact that it is a latent function that needs to be approximated with a spline representation, which requires unusual treatment different from the typical handling of the spline approximation. In addition, for the regression problem, we encounter further difficulties because not only the estimator, but also the estimating equation that generates the estimator, do not have closed forms. Much effort is spent in clearing these hurdles, where we resort to extensive use of bilinear operators (Conway, 1990), which are very different from typical regression spline asymptotic analysis (Ma et al., 2015). The detailed proofs are respectively in the Appendix at the end of the paper and in an online supplementary document.

The success of the B-spline-assisted procedures in achieving the better rates of convergence in these problems is encouraging. Despite of the satisfactory result, it is not immediately clear what is the intuitive reason behind the success. To rationalize the convergence rate of the B-spline based procedures, we can consider the parametric pdf and regression problems with measurement error. In these simpler problems, root-n convergence of the parameter estimation is indeed achievable. Thus, bridging a nonparametric problem to a parametric problem via B-splines, it is not unreasonable to expect a classical nonparametric convergence rate for measurement error problems as well. A critical assumption in such B-spline treatment is that the possible range of the unobserved covariates are bounded, hence we can confine to a compact set in estimating the functions of interest. This restricted the impact from the tails of the measurement error distribution, which is exactly what brings down the convergence rate of the deconvolution methods. More explicitly, in the B-spline approximation, the compact set on which we perform the estimation is built into the procedure at the very beginning and we benefit from that throughout the procedure. However, in the deconvolution procedures, the Fourier and inverse Fourier transformation steps do not take advantage of the compact set knowledge, and it automatically estimates these functions on the whole interval, which is much harder to do.

In the following, we devise the estimation procedures for both the probability density function and the regression mean function in Section 2, and summarize the theoretical properties of our estimators in Section 3. We provide simulation studies to demonstrate the rate properties of the new estimators in Section 4, and illustrate the methods in a data example in Section 5. The paper is concluded with a discussion in Section 6.

2. B-spline-assisted estimation procedures.

2.1. *Probability density function estimation.* To set the notation, we use $f_X(x)$ to denote a generic pdf function of the random variable X, and use $f_{X0}(x)$ to denote the true pdf that generates the data. We approximate $f_{X0}(x)$ on its support using B-splines (Masri and Redner, 2005). For simplicity, let the support be contained in $[0,1]$. To ensure that the density function is nonnegative and integrates to 1, we let the approximation be

$$f_X(x,\boldsymbol{\theta}) \equiv \frac{\exp\{\mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\theta}\}}{\int_0^1 \exp\{\mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\theta}\}dx},$$

where $\mathbf{B}_r(x)$ is a vector of B-spline basis functions, and $\boldsymbol{\theta}$ is the B-spline coefficient vector to be estimated. Then

$$f_W(w,\boldsymbol{\theta}) \equiv \frac{\int_0^1 \exp\{\mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\theta}\}f_U(w-x)dx}{\int_0^1 \exp\{\mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\theta}\}dx}$$

is an approximation to the pdf of $W \equiv X + U$, a surrogate of X. We then perform a simple maximum likelihood estimator (MLE) to obtain $\boldsymbol{\theta}$, i.e. we maximize

$$\begin{aligned}
&\sum_{i=1}^n \log f_W(W_i,\boldsymbol{\theta})\\
=\ &\sum_{i=1}^n \log \int_0^1 \exp\{\mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\theta}\}f_U(W_i-x)dx - n\log \int_0^1 \exp\{\mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\theta}\}dx
\end{aligned}$$

with respect to $\boldsymbol{\theta}$ to obtain $\widehat{\boldsymbol{\theta}}$, and then reconstruct $f_X(x,\widehat{\boldsymbol{\theta}})$ and use it as the estimator for $f_{X0}(x)$, i.e. $\widehat{f}_{X0}(x) = f_X(x,\widehat{\boldsymbol{\theta}})$.

While the estimation procedure for $f_{X0}(x)$ is extremely simple, it is not as straightforward to establish the large sample properties of the estimator. In Section 3, we will show that $\widehat{f}_{X0}(x)$ converges to $f_{X0}(x)$ at a nonparametric rate under mild conditions.

2.2. *Regression function estimation.* Unlike in the density estimation case, the estimation procedure in the regression model is much more complex as we now describe. The key observation is that as soon as the nonparametric function $m(x)$ is approximated with the B-spline representation, the regression function itself without measurement error is a purely parametric model (Wang and Yang, 2009), hence the idea behind Tsiatis and Ma (2004) can be adapted. Specifically, we treat the B-spline coefficients as parameters of interest, treat the unspecified distribution of X, $f_X(x)$, as nuisance parameter, and cast the problem as a semiparametric estimation problem. We can then construct the efficient score function, which relies on $f_{X0}(x)$. A key observation of Tsiatis and Ma (2004) is that by replacing $f_{X0}(x)$ with an arbitrary working model, the consistency of the estimation can still be retained. We now describe the estimation procedure in detail.

First, let $f_X^*(x)$ be a working pdf of X. Note that $f_X^*(x)$ may not be the same as $f_{X0}(x)$, that is, $f_X^*(x)$ is possibly misspecified. Following the practice of density estimation in Section 2.1, assume that the true density function $f_{X0}(x)$ has compact support contained in $[0,1]$. Therefore we only need to consider $m(x)$ on $[0,1]$. We approximate $m(x)$ using spline representation $\mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\beta}$. Define

$$\begin{aligned}\mathbf{S}^*_{\boldsymbol{\beta}}(W,Y,\boldsymbol{\beta}) &= \frac{\partial}{\partial\boldsymbol{\beta}}\log\int_0^1 f_\epsilon\{Y-\mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\beta}\}f_U(W-x)f_X^*(x)d\mu(x)\\ &= -\frac{\int_0^1 f_\epsilon'\{Y-\mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\beta}\}f_U(W-x)f_X^*(x)\mathbf{B}_r(x)d\mu(x)}{\int_0^1 f_\epsilon\{Y-\mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\beta}\}f_U(W-x)f_X^*(x)d\mu(x)}.\end{aligned}$$

As the notation suggests, $\mathbf{S}^*_{\boldsymbol{\beta}}(W,Y,\boldsymbol{\beta})$ is the score function with respect to $\boldsymbol{\beta}$ calculated from the joint pdf of (W,Y) under the working model $f_X^*(x)$ and the spline approximation. Due to the possible misspecification of $f_X^*(x)$, the mean of $\mathbf{S}^*_{\boldsymbol{\beta}}(W,Y,\boldsymbol{\beta})$ is not necessarily zero even if the mean function is exactly $\mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\beta}$. Therefore simply solving $\sum_{i=1}^n \mathbf{S}^*_{\boldsymbol{\beta}}(W_i,Y_i,\boldsymbol{\beta})=\mathbf{0}$ may generate an inconsistent estimator. The idea behind our estimator is to find a function $\mathbf{a}(x,\boldsymbol{\beta})$ so that

$$\begin{aligned}(1)\qquad & E\{\mathbf{S}^*_{\boldsymbol{\beta}}(W,Y,\boldsymbol{\beta})\mid X\}\\ &= E\left[\frac{\int_0^1 \mathbf{a}(x,\boldsymbol{\beta})f_\epsilon\{Y-\mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\beta}\}f_U(W-x)f_X^*(x)d\mu(x)}{\int_0^1 f_\epsilon\{Y-\mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\beta}\}f_U(W-x)f_X^*(x)d\mu(x)}\mid X\right],\end{aligned}$$

and then solve for $\boldsymbol{\beta}$ using the estimating equation

$$\begin{aligned}(2)\qquad & \sum_{i=1}^n\Big[\mathbf{S}^*_{\boldsymbol{\beta}}(W_i,Y_i,\boldsymbol{\beta})\\ & -\frac{\int_0^1 \mathbf{a}(x,\boldsymbol{\beta})f_\epsilon\{Y_i-\mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\beta}\}f_U(W_i-x)f_X^*(x)d\mu(x)}{\int_0^1 f_\epsilon\{Y_i-\mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\beta}\}f_U(W_i-x)f_X^*(x)d\mu(x)}\Bigg]=\mathbf{0}.\end{aligned}$$

This will guarantee a consistent estimator of $\boldsymbol{\beta}$ if the mean function is indeed $\mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\beta}$, because our construction ensures the left hand side of (2) has mean zero. Note that the right hand side of (1) is the conditional expectation of

$\mathbf{a}(X, \boldsymbol{\beta})$ calculated under the B-spline approximation and the posited model $f_X^*(x)$, hence we alternatively write it as $E^*\{\mathbf{a}(X, \boldsymbol{\beta}) \mid Y_i, W_i, \boldsymbol{\beta}\}$.

To solve for $\mathbf{a}(x, \boldsymbol{\beta})$, we discretize the integral equation (1). In particular, let $f_X^*(x) = \sum_{j=1}^L c_j I(x = x_j)$, where x_j's are the points selected on $[0, 1]$, and $c_j \geq 0, \sum_{j=1}^L c_j = 1$. Then

$$\mathbf{S}_{\boldsymbol{\beta}}^*(W, Y, \boldsymbol{\beta}) = -\frac{\sum_{j=1}^L \mathbf{B}_r(x_j) f_\epsilon'\{Y - \mathbf{B}_r(x_j)^{\mathrm{T}}\boldsymbol{\beta}\} f_U(W - x_j) c_j}{\sum_{j=1}^L f_\epsilon\{Y - \mathbf{B}_r(x_j)^{\mathrm{T}}\boldsymbol{\beta}\} f_U(W - x_j) c_j}.$$

Next, to write out the right hand side of (1) upon discretization, let $\mathbf{A}(\boldsymbol{\beta})$ be an $L \times L$ matrix with its (i, j) entry

$$\begin{aligned} A_{ij}(\boldsymbol{\beta}) &= \int \frac{f_\epsilon\{y - \mathbf{B}_r(x_j)^{\mathrm{T}}\boldsymbol{\beta}\} f_U(w - x_j) c_j}{\sum_{j=1}^L f_\epsilon\{y - \mathbf{B}_r(x_j)^{\mathrm{T}}\boldsymbol{\beta}\} f_U(w - x_j) c_j} \\ &\quad \times f_\epsilon\{y - \mathbf{B}_r(x_i)^{\mathrm{T}}\boldsymbol{\beta}\} f_U(w - x_i) d\mu(y) d\mu(w). \end{aligned}$$

Let $\mathbf{a}_i = \mathbf{a}(x_i, \boldsymbol{\beta}), \mathbf{a} = (\mathbf{a}_1, \ldots, \mathbf{a}_L)$. Further, define $\mathbf{H}_j(\boldsymbol{\beta}) = \{H_{1j}(\boldsymbol{\beta}), \ldots, H_{Lj}(\boldsymbol{\beta})\}$, where

$$\begin{aligned} H_{ij}(\boldsymbol{\beta}) &= -\int \frac{f_\epsilon'\{y - \mathbf{B}_r(x_j)^{\mathrm{T}}\boldsymbol{\beta}\} f_U(w - x_j) c_j}{\sum_{j=1}^L c_j f_\epsilon\{y - \mathbf{B}_r(x_j)^{\mathrm{T}}\boldsymbol{\beta}\} f_U(w - x_j)} \\ &\quad \times f_\epsilon\{y - \mathbf{B}_r(x_i)^{\mathrm{T}}\boldsymbol{\beta}\} f_U(w - x_i) d\mu(y) d\mu(w), \end{aligned}$$

and let $\mathbf{b}(\boldsymbol{\beta})$ be a $p \times L$ matrix, with its ith column

$$\mathbf{b}_i(\boldsymbol{\beta}) = \sum_{j=1}^L H_{ij}(\boldsymbol{\beta}) \mathbf{B}_r(x_j).$$

Then $\mathbf{b}(\boldsymbol{\beta}) = \sum_{j=1}^L \mathbf{B}_r(x_j) \mathbf{H}_j(\boldsymbol{\beta})$.

With these notations, the integral equation (1) is equivalently written as $\sum_{j=1}^L A_{ij} \mathbf{a}_j = \sum_{j=1}^L H_{ij}(\boldsymbol{\beta}) \mathbf{B}_r(x_j)$ for $i = 1, \ldots, L$, or more concisely, $\mathbf{a}\mathbf{A}^{\mathrm{T}}(\boldsymbol{\beta}) = \sum_{j=1}^L \mathbf{B}_r(x_j) \mathbf{H}_j(\boldsymbol{\beta})$. Therefore

$$\mathbf{a}(\boldsymbol{\beta}) = \sum_{k=1}^L \mathbf{B}_r(x_k) \mathbf{H}_k(\boldsymbol{\beta}) \{\mathbf{A}^{-1}(\boldsymbol{\beta})\}^{\mathrm{T}},$$

hence

$$\mathbf{a}_j(\boldsymbol{\beta}) = \sum_{k=1}^L \mathbf{B}_r(x_k) \mathbf{H}_k(\boldsymbol{\beta}) \{\mathbf{A}^{-1}(\boldsymbol{\beta})\}^{\mathrm{T}} \mathbf{e}_j,$$

where $\mathbf{e}_j$ is a length L vector with the jth component 1 and all others zero.

Thus, we have obtained $\mathbf{a}(X, \boldsymbol{\beta})$ on the discrete set $x_1, \dots x_L$ and can form

$$
\begin{aligned}
&E^*\{\mathbf{a}(X,\boldsymbol{\beta}) \mid Y, W, \boldsymbol{\beta}\} \\
&= \frac{\sum_{j=1}^{L} \mathbf{a}_j(\boldsymbol{\beta}) f_\epsilon\{Y - \mathbf{B}_r(x_j)^{\mathrm{T}}\boldsymbol{\beta}\} f_U(W - x_j) c_j}{\sum_{j=1}^{L} f_\epsilon\{Y - \mathbf{B}_r(x_j)^{\mathrm{T}}\boldsymbol{\beta}\} f_U(W - x_j) c_j} \\
&= \sum_{k=1}^{L} \mathbf{B}_r(x_k) \frac{\mathbf{H}_k(\boldsymbol{\beta}) \sum_{j=1}^{L} \{\mathbf{A}^{-1}(\boldsymbol{\beta})\}^{\mathrm{T}} \mathbf{e}_j f_\epsilon\{Y - \mathbf{B}_r(x_j)^{\mathrm{T}}\boldsymbol{\beta}\} f_U(W - x_j) c_j}{\sum_{j=1}^{L} f_\epsilon\{Y - \mathbf{B}_r(x_j)^{\mathrm{T}}\boldsymbol{\beta}\} f_U(W - x_j) c_j} \\
&= \sum_{k=1}^{L} \mathbf{B}_r(x_k) P_k(W, Y, \boldsymbol{\beta}),
\end{aligned}
$$

where

$$
P_k(W, Y, \boldsymbol{\beta}) = \frac{\mathbf{H}_k(\boldsymbol{\beta}) \sum_{j=1}^{L} \{\mathbf{A}^{-1}(\boldsymbol{\beta})\}^{\mathrm{T}} \mathbf{e}_j f_\epsilon\{Y - \mathbf{B}_r(x_j)^{\mathrm{T}}\boldsymbol{\beta}\} f_U(W - x_j) c_j}{\sum_{j=1}^{L} f_\epsilon\{Y - \mathbf{B}_r(x_j)^{\mathrm{T}}\boldsymbol{\beta}\} f_U(W - x_j) c_j}.
$$

We then obtain the estimator for $\boldsymbol{\beta}$ by solving the estimating equation (2) with the corresponding $\mathbf{a}(x, \boldsymbol{\beta})$ plugged in. Note that in all the functions that are explicitly written to depend on $\boldsymbol{\beta}$, the dependence is always through $\mathbf{B}_r(\cdot)^{\mathrm{T}}\boldsymbol{\beta}$.

We show that $\mathbf{B}_r(x)^{\mathrm{T}}\widehat{\boldsymbol{\beta}}$ converges to $m(x)$ at a nonparametric rate and we derive its estimation variance in Section 3 under mild conditions. As pointed out by an anonymous referee, instead of adopting a working model $f_X^*(x)$, we could in fact estimate $f_{X0}(x)$ using the method developed in Section 2.1 to obtain $\widehat{f}_X(x)$, and then use $\widehat{f}_X(x)$ instead of $f_X^*(x)$ to carry out the subsequent operations. We believe that this is the optimal thing to do and will generate the most efficient estimator. However, rigorously establishing the theoretical property of such practice is challenging so we leave this option to future research.

3. Asymptotic results.

3.1. *Results of probability density function estimation.* We assume the following regularity conditions.

(C1) The true density function $f_{X0}(x)$ has compact support contained in $[0, 1]$, is bounded on its support and satisfies $f_{X0}(x) \in C^q([0, 1])$, $q \geq 1$. $f_U(u)$ is bounded above. The conditional pdf of X given W, $f_{X|W}(x|w)$, is bounded.

(C2) The spline order $r \geq q$.

(C3) We define the knots $t_{-r+1} = \cdots = t_0 = 0 < t_1 < \cdots < t_N < 1 = t_{N+1} = \cdots = t_{N+r}$, where N is the number of interior knots and $[0,1]$ is divided into $N+1$ subintervals. Let $d_{\boldsymbol{\theta}} = N+r$. N satisfies $N \to \infty$, $N^{-1}n(\log n)^{-1} \to \infty$, and $Nn^{-1/(2q)} \to \infty$ when $n \to \infty$.

(C4) Let h_p be the distance between the $(p+1)$th and pth interior knots and let $h_b = \max_{r \le p \le N+r} h_p$, $h_s = \min_{r \le p \le N+r} h_p$. There exists a constant c_{h_b}, $0 < c_{h_b} < \infty$, such that

$$h_b/h_s < c_{h_b}.$$

Therefore, $h_b = O(N^{-1})$, $h_s = O(N^{-1})$.

(C5) $\boldsymbol{\theta}_0$ is a spline coefficient such that $\sup_{x\in[0,1]} |\mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\theta}_0 - \log\{f_{X0}(x)\}| = O_p(h_b^q)$. The existence of such $\boldsymbol{\theta}_0$ has been shown in De Boor (1978).

(C6) The expectation

$$E\left(\left[\frac{\int_0^1 \exp\{\mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\theta}\} f_U(W_i - x)\mathbf{B}_r(x)dx}{\int_0^1 \exp\{\mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\theta}\} f_U(W_i - x)dx} - \frac{\int_0^1 \exp\{\mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\theta}\}\mathbf{B}_r(x)dx}{\int_0^1 \exp\{\mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\theta}\}dx}\right]\right)$$

is a smooth function of $\boldsymbol{\theta}$ and has unique root for $\boldsymbol{\theta}$ in the neighborhood of $\boldsymbol{\theta}_0$.

Condition (C1) contains some basic boundedness and smoothness conditions on various densities and is quite standard. The only requirement that appears nonstandard is that $f_{X0}(x)$ has compact support. This however is practically relevant in many situations as the range of the values a random variable can be may very well be bounded. Condition (C2) is a standard requirement to ensure that splines of sufficiently high order are utilized. Condition (C3) requires suitable amount of spline basis to be used according to the sample size, and Condition (C4) makes sure that the spline knots are distributed sufficiently evenly. In summary, Conditions (C2), (C3) and (C4) are standard requirements and together with Condition (C1), they ensure Condition (C5). We list Condition (C5) instead of stating it in the proof for convenience. Finally, Condition (C6) ensures that we are not in the degenerate case where the expression in (C6) is constantly zero. Under these conditions, we obtain the following properties of $\widehat{f}_{X0}(x)$.

PROPOSITION 1. *Assume Conditions (C1) – (C6) to hold. Let $\widehat{\boldsymbol{\theta}}$ maximize*

$$\begin{aligned} & \sum_{i=1}^n \log f_W(W_i, \boldsymbol{\theta}) \\ = \ & \sum_{i=1}^n \log \int_0^1 \exp\{\mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\theta}\} f_U(W_i - x)dx - n\log \int_0^1 \exp\{\mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\theta}\}dx, \end{aligned}$$

then $\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0 = o_p(1)$ *element-wise.*

PROPOSITION 2. *Assume Conditions (C1) – (C6) to hold. Let* $\widehat{\boldsymbol{\theta}}$ *maximize*

$$\begin{aligned}
&\sum_{i=1}^{n} \log f_W(W_i, \boldsymbol{\theta}) \\
=\ & \sum_{i=1}^{n} \log \int_0^1 \exp\{\mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\theta}\} f_U(W_i - x)dx - n\log \int_0^1 \exp\{\mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\theta}\}dx.
\end{aligned}$$

Then $\|\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0\|_2 = O_p\{(nh_b)^{-1/2}\}$. *Further, define*

$$\mathbf{R}_{00} = n^{-1} \sum_{i=1}^{n} \int_0^1 \left\{ \frac{f_{X0}(x) f_U(W_i - x)}{\int_0^1 f_{X0}(x) f_U(W_i - x)dx} - f_{X0}(x) \right\} \mathbf{B}_r(x)dx.$$

Then

$$\begin{aligned}
&\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0 \\
=\ & \left(E \left[\frac{\int_0^1 f_{X0}(x) f_U(W_i - x)\mathbf{B}_r(x)\mathbf{B}_r(x)^{\mathrm{T}}dx}{\int_0^1 f_{X0}(x) f_U(W_i - x)dx} \right.\right. \\
& - \frac{\left\{\int_0^1 f_{X0}(x) f_U(W_i - x)\mathbf{B}_r(x)dx\right\}^{\otimes 2}}{\left\{\int_0^1 f_{X0}(x) f_U(W_i - x)dx\right\}^2} - \int_0^1 f_{X0}(x)\mathbf{B}_r(x)\mathbf{B}_r(x)^{\mathrm{T}}dx \\
& \left.\left. + \left\{ \int_0^1 f_{X0}(x)\mathbf{B}_r(x)dx \right\}^{\otimes 2} \right] \right)^{-1} \mathbf{R}_{00}\{1 + o_p(1)\}.
\end{aligned}$$

Propositions 1 and 2 establishes the consistency and convergence rate properties of $\widehat{\boldsymbol{\theta}}$ to $\boldsymbol{\theta}_0$ defined in Condition (C5). The proof of Propositions 1 and 2 are in Supplement S.3 and S.4 respectively. We then utilize these properties to analyze the bias, variance and convergence rate of $\widehat{f}_{X0}(x)$ in Theorem 1, with its proof in Appendix A.1.

Theorem 1. *Assume Conditions (C1) – (C6) to hold. Let* $\widehat{\boldsymbol{\theta}}$ *be defined in Proposition 2 and*

$$\widehat{f}_{X0}(x) = \frac{\exp\{\mathbf{B}_r(x)^{\mathrm{T}}\widehat{\boldsymbol{\theta}}\}}{\int_0^1 \exp\{\mathbf{B}_r(x)^{\mathrm{T}}\widehat{\boldsymbol{\theta}}\}dx}.$$

Then $\sup_{x\in[0,1]} |\log\{\widehat{f}_{X0}(x)\} - \log\{f_{X0}(x)\}| = O_p\{(nh_b)^{-1/2}\}$. *Further*

$$\mathit{bias}\{\widehat{f}_{X0}(x)\} \equiv E\{\widehat{f}_{X0}(x) - f_{X0}(x)\} = O_p(h_b^{q-1/2}),$$

and

$$
\begin{aligned}
& \sqrt{nh_b}\left[\widehat{f}_{X0}(x) - f_{X0}(x) - bias\{\widehat{f}_{X0}(x)\}\right] \\
= \quad & \sqrt{nh_b}\frac{\partial}{\partial \boldsymbol{\theta}^{\mathrm{T}}}\frac{\exp\{\mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\theta}_0\}}{\int_0^1 \exp\{\mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\theta}_0\}dx}\left(E\left[\frac{\int_0^1 f_{X0}(x)f_U(W_i-x)\mathbf{B}_r(x)\mathbf{B}_r(x)^{\mathrm{T}}dx}{\int_0^1 f_{X0}(x)f_U(W_i-x)dx}\right.\right. \\
& -\frac{\left\{\int_0^1 f_{X0}(x)f_U(W_i-x)\mathbf{B}_r(x)dx\right\}^{\otimes 2}}{\left\{\int_0^1 f_{X0}(x)f_U(W_i-x)dx\right\}^2} - \int_0^1 f_{X0}(x)\mathbf{B}_r(x)\mathbf{B}_r(x)^{\mathrm{T}}dx \\
& \left.\left. + \left\{\int_0^1 f_{X0}(x)\mathbf{B}_r(x)dx\right\}^{\otimes 2}\right]\right)^{-1} \\
& \times n^{-1}\sum_{i=1}^n \int_0^1 \left\{\frac{f_{X0}(x)f_U(W_i-x)}{\int_0^1 f_{X0}(x)f_U(W_i-x)dx} - f_{X0}(x)\right\}\mathbf{B}_r(x)dx + o_p(1).
\end{aligned}
$$

Theorem 1 shows that the B-spline MLE density estimator is root-nh_b consistent to $f_{X0}(x)$, which is a standard nonparametric density estimation rate when no measurement error occurs, and is improved substantially compared to the nonparametric deconvolution method.

3.2. *Results of regression mean function estimation.* To facilitate the description of the regularity conditions and the asymptotic results, we introduce some notations. Define

$$
\begin{aligned}
P(x, W, Y, \boldsymbol{\beta}) &= \frac{\mathbf{H}(x,\boldsymbol{\beta})\sum_{j=1}^L \{\mathbf{A}^{-1}(\boldsymbol{\beta})\}^{\mathrm{T}}\mathbf{e}_j f_\epsilon\{Y - \mathbf{B}_r(x_j)^{\mathrm{T}}\boldsymbol{\beta}\}f_U(W-x_j)c_j}{\int_0^1 f_\epsilon\{Y-\mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\beta}\}f_U(W-x)f_X^*(x)d\mu(x)}, \\
\mathbf{H}(x,\boldsymbol{\beta}) &= \{H_1(x,\boldsymbol{\beta}),\ldots,H_L(x,\boldsymbol{\beta})\}, \\
H_i(x,\boldsymbol{\beta}) &= -\int \frac{f_\epsilon'\{y-\mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\beta}\}f_U(w-x)f_X^*(x)}{\int_0^1 f_\epsilon\{y-\mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\beta}\}f_U(w-x_j)f_X^*(x)d\mu(x)} \\
&\quad \times f_\epsilon\{y-\mathbf{B}_r(x_i)^{\mathrm{T}}\boldsymbol{\beta}\}f_U(w-x_i)d\mu(y)d\mu(w),
\end{aligned}
$$

and write

$$
\begin{aligned}
E^*\{\mathbf{a}(X,\boldsymbol{\beta}) \mid Y, W, \boldsymbol{\beta}\} &= \int_0^1 \mathbf{B}_r(x)P(x,W,Y,\boldsymbol{\beta})d\mu(x), \\
\mathbf{S}_{\boldsymbol{\beta}}^*(W,Y,\boldsymbol{\beta}) &= \int_0^1 \frac{\mathbf{B}_r(x)f_\epsilon'\{Y-\mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\beta}\}f_U(W-x)f_X^*(x)}{-\int_0^1 f_\epsilon\{Y-\mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\beta}\}f_U(W-x)f_X^*(x)}d\mu(x).
\end{aligned}
$$

We further define $\mathbf{S}_{\boldsymbol{\beta}}^*(W_i, Y_i, m)$, $E^*\{\mathbf{a}(X,m) \mid Y_i, W_i, m\}$, $P(x,W,Y,m)$, $P_k(W,Y,m)$ to be the resulting quantities when we replace all the appearance of $\mathbf{B}_r(\cdot)^{\mathrm{T}}\boldsymbol{\beta}$ in $\mathbf{S}_{\boldsymbol{\beta}}^*(W_i,Y_i,\boldsymbol{\beta})$, $E^*\{\mathbf{a}(X,\boldsymbol{\beta}) \mid Y_i, W_i, \boldsymbol{\beta}\}$, $P(x,W,Y,\boldsymbol{\beta})$, and

$P_k(W, Y, m)$ by $m(\cdot)$ respectively. Here $\mathbf{a}(X, m)$ is a function that satisfies

$$E[\mathbf{S}^*_{\boldsymbol{\beta}}(W_i, Y_i, m) - E^*\{\mathbf{a}(X, m)|Y_i, W_i, m\}|X, m] = \mathbf{0},$$

where the last m is used to emphasize that the calculation of the outside expectation depends on m. Note that these definitions do not conflict with the notations used before. In fact, some generalize the previous notations. We further define $S_m(Y, W, m)$ to be a linear operator on L_p whose value at $s \in L_p$ is

$$\begin{aligned} & S_m(Y, W, m)(s) \\ = & \int_0^1 \left[-\frac{f'_\epsilon\{Y - m(x)\} f_U(W - x) f^*_X(x) d\mu(x)}{\int_0^1 f_\epsilon\{Y - m(x)\} f_U(W - x) f^*_X(x) d\mu(x)} - P(x, Y, W, m) \right] s(x) dx. \end{aligned}$$

The derivative of $S_m(Y, W, m)$ is a bilinear operator on $L_p \times L_q$ with $1/p + 1/q = 1, 1 \le p, q \le \infty$, whose value at $s \in L_p, v \in L_q$ is

$$\frac{\partial S_m(Y, W, m)}{\partial m}(s, v) = \left.\frac{\partial S_m(Y, W, m + tv)(s)}{\partial t}\right|_{t=0}.$$

Further, we define a bilinear operator

$$S^2_m(Y, W, m)(s, v) = S_m(Y, W, m)(s) S_m(Y, W, m)(v).$$

From the above definition, we can see that

$$\begin{aligned} E\{S_m(Y, W, m)\}(s) = 0, \\ n^{-1} \textstyle\sum_{i=1}^n S_m(Y_i, W_i, m)(s) - E\{S_m(Y_i, W_i, m)\}(s) = o_p(1)\|s\|_p. \end{aligned}$$

Also,

$$\begin{aligned} n^{-1} \textstyle\sum_{i=1}^n \frac{\partial S_m(Y_i, W_i, m)}{\partial m}(s, v) - E\left\{\frac{\partial S_m(Y_i, W_i, m)}{\partial m}\right\}(s, v) = o_p(1)\|s\|_p\|v\|_q, \\ n^{-1} \textstyle\sum_{i=1}^n S^2_m(Y_i, W_i, m)(s, v) - E\{S^2_m(Y_i, W_i, m)\}(s, v) = o_p(1)\|s\|_p\|v\|_q, \end{aligned}$$

where $E\{S^2_m(Y_i, W_i, m)\}(s, v) = E\{S_m(Y, W, m)(s) S_m(Y, W, m)(v)\}$ and $E\{\partial S_m(Y_i, W_i, m)/\partial m\}(s, v) = E[\{\partial S_m(Y, W, m)/\partial m\}(s, v)]$.

We now list the regularity conditions.

(D1) The true density functions $f_{X0}(x)$, $f_\epsilon(\epsilon)$ are bounded on their supports. In addition, the support of $f_{X0}(x)$ is compact.

(D2) $E\{\partial S_m(Y, W, m)/\partial m\}$ is a bounded bilinear operator on $L_2 \times L_2$, $L_1 \times L_\infty$, and $L_\infty \times L_1$. $E\{S^2_m(Y_i, W_i, m)\}$ is a bounded bilinear operator on $L_2 \times L_2$.

(D3) $m(x)$ is bounded on $[0,1]$ and it satisfies $m(x) \in C^q([0,1])$, $q \geq 1$.

(D4) $\boldsymbol{\beta}_0$ is a $d_{\boldsymbol{\beta}}$ dimensional spline coefficient such that $\sup_{x\in[0,1]} |\mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\beta}_0 - m(x)| = O_p(h_b^q)$. The existence of such $\boldsymbol{\beta}_0$ has been shown in De Boor (1978).

(D5) The expectation

$$E[\mathbf{S}^*_{\boldsymbol{\beta}}(W_i, Y_i, \boldsymbol{\beta}) - E^*\{\mathbf{a}(X, \boldsymbol{\beta})|Y_i, W_i, \boldsymbol{\beta}\}]$$

has unique root for $\boldsymbol{\beta}$ in the neighborhood of $\boldsymbol{\beta}_0$. Its derivative with respect to $\boldsymbol{\beta}$ is a smooth function of $\boldsymbol{\beta}$ in the neighborhood of $\boldsymbol{\beta}_0$, with its singular values bounded and bounded away from zero. Denote the unique zero $\boldsymbol{\beta}^*$.

Conditions (D1) and (D2) contain bounded requirement on pdfs and operators and are not stringent. Condition (D1) further requires the compact support of the distribution of X. This is similar to the density estimation case and is crucial. Conditions (D3) and (D4) are regarding the mean function $m(x)$ and its spline approximation which are quite standard. Condition (D5) is a unique root requirement similar to (C6) and is used to exclude the pathologic case where the estimating equation is constantly zero.

In the following, we establish the consistency of the parameter estimation in Proposition 3 and then further analyze its convergence rate in Proposition 4. The results in these propositions are subsequently used to further establish the asymptotic properties of the estimator of the regression mean function $m(x)$ in Theorem 2. The proofs of both the propositions and the theorem are in Supplement S.5, S.6 and Appendix A.2 respectively.

Proposition 3. *Assume Conditions (C2) – (C4), (D1) – (D5) to hold. Let $\widehat{\boldsymbol{\beta}}$ satisfy*

$$\sum_{i=1}^n \mathbf{S}^*_{\boldsymbol{\beta}}(W_i, Y_i, \widehat{\boldsymbol{\beta}}) - E^*\{\mathbf{a}(X, \widehat{\boldsymbol{\beta}}) \mid Y_i, W_i, \widehat{\boldsymbol{\beta}}\} = \mathbf{0}.$$

Then $\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0 = o_p(1)$ element-wise.

Proposition 4. *Assume Conditions (C2) – (C4), (D3) – (D5) to hold. Let $\widehat{\boldsymbol{\beta}}$ satisfy*

$$\sum_{i=1}^n \mathbf{S}^*_{\boldsymbol{\beta}}(W_i, Y_i, \widehat{\boldsymbol{\beta}}) - E^*\{\mathbf{a}(X, \widehat{\boldsymbol{\beta}}) \mid Y_i, W_i, \widehat{\boldsymbol{\beta}}\} = \mathbf{0}.$$

Then $\|\widehat{\boldsymbol\beta}-\boldsymbol\beta_0\|_2 = O_p\{(nh)^{-1/2}\}$. *Further,*

$$\begin{aligned}\widehat{\boldsymbol\beta}-\boldsymbol\beta_0 &= \left\{E\left(\frac{\partial[\mathbf{S}^*_{\boldsymbol\beta}(W_i,Y_i,\boldsymbol\beta)-E^*\{\mathbf{a}(X,\boldsymbol\beta)\mid Y_i,W_i,\boldsymbol\beta\}]}{\partial\boldsymbol\beta^{\mathrm T}}\bigg|_{\mathbf{B}_r(\cdot)^{\mathrm T}\boldsymbol\beta=m(\cdot)}\right)\right\}^{-1}\\ &\quad\times\mathbf{T}_{00}\{1+o_p(1)\},\end{aligned}$$

where

$$\mathbf{T}_{00}=n^{-1}\sum_{i=1}^n \mathbf{S}^*_{\boldsymbol\beta}(W_i,Y_i,m)-E^*\{\mathbf{a}(X,m)\mid Y_i,W_i,m\}.$$

Theorem 2. *Assume Conditions (C2) – (C4), (D1) – (D5) to hold. Let* $\widehat m(x)=\mathbf{B}_r(x)^{\mathrm T}\widehat{\boldsymbol\beta}$. *Then* $\sup_{x\in[0,1]}|\widehat m(x)-m(x)| = O_p\{(nh_b)^{-1/2}\}$. *Specifically,* $bias\{\widehat m(x)\}=E\{\widehat m(x)-m(x)\}=O_p(h_b^{q-1/2})$ *and*

$$\begin{aligned}&\sqrt{nh_b}[\widehat m(x)-m(x)-bias\{\widehat m(x)\}]\\ =\;&\sqrt{nh_b}\mathbf{B}_r(x)^{\mathrm T}\\ &\times\left[-\left\{E\left(\frac{\partial[\mathbf{S}^*_{\boldsymbol\beta}(W_i,Y_i,\boldsymbol\beta)-E^*\{\mathbf{a}(X,\boldsymbol\beta)\mid Y_i,W_i,\boldsymbol\beta\}]}{\partial\boldsymbol\beta^{\mathrm T}}\bigg|_{\mathbf{B}_r(\cdot)^{\mathrm T}\boldsymbol\beta=m(\cdot)}\right)\right\}^{-1}\right.\\ &\left.\times n^{-1}\sum_{i=1}^n\mathbf{S}^*_{\boldsymbol\beta}(W_i,Y_i,m)-E^*\{\mathbf{a}(X,m)\mid Y_i,W_i,m\}\right]+o_p(1).\end{aligned}$$

Theorem 2 shows that even with misspecified $f^*_X(x)$, $\widehat m(x)$ converges to $m(x)$ with root-nh_b rate, which is the standard nonparametric rate of the regression mean function without measurement errors, and is considerably faster than deconvolution estimators.

4. Simulation studies.

4.1. *Performance of B-spline-assisted MLE.* We conduct two simulation studies to illustrate the finite sample performance of the proposed pdf and regression mean estimators. To evaluate the B-spline MLE method for estimating the density functions, we generated data sets of sample sizes from $n = 500$ to $n = 2000$, where X is generated from a beta distribution with both shape parameters equal to 4. We then generated the measurement error U from three models:

I(a): a normal distribution with mean 0, variance 0.25, denoted by $N(0, 0.25)$,

I(b): a Laplace distribution with mean 0 and scale $0.5/\sqrt{2}$, denoted by $\mathrm{Lap}(0, 0.5/\sqrt{2})$,
I(c): a uniform distribution on $[-0.25, 0.25]$, denoted by $\mathrm{Unif}(-0.25, 0.25)$.

We used cubic splines to approximate the density functions, with number of knots equal to the smallest integer larger than $1.7n^{1/5}$. In the left panel of Figure 1, we plotted the averaged root-(nh_b) maximum absolute error (MAE), calculated as $\sqrt{nh_b}\sup_x |\widehat{f}_{X0}(x) - f_{X0}(x)|$, versus the sample sizes n. Following Theorem 1, the root-(nh_b) MAE has a constant order. This translates to the curves in the plots that stabilize in a small range, which is what we observe, especially when sample size grows to larger than 700.

We also compared the B-spline MLE method with the widely used deconvolution method for density estimation. In the upper row of Figure 2, we plotted the average values of $\sup_x |\widehat{f}_{X0}(x) - f_{X0}(x)|$ based on 200 simulations for both methods at different sample sizes. We adopted the two-stage plug-in bandwidth selection method proposed in Delaigle and Gijbels (2002) in implementing the deconvolution method. Much to our surprise, results in the upper row of Figure 2 indicate that the B-spline MLE method outperforms the deconvolution method with rather significant gain in this case. We suspect that this is because the measurement errors are quite large here which caused difficulties for both methods but especially for the deconvolution method.

To further examine the performance of individual estimated pdf curves from both methods, we also plotted the estimated density curves for sample sizes 500, 1000 and 2000. Because the deconvolution method performs poorly when the measurement errors are large (see upper row of Figure 2), we reduced the error variances, and generated the three error distributions from models

$$\text{II(a): } N(0, 0.0025),\ \text{II(b): } \mathrm{Lap}(0, 0.05/\sqrt{2}),\ \text{II(c): } \mathrm{Unif}(-0.125, 0.125).$$

Under the reduced error variability, we performed the estimation and plotted the resulting density estimates and their 90% confidence bands. Figure 3 contains the results of the B-spline MLE and the deconvolution estimator using the two-stage plug-in bandwidth (Delaigle and Gijbels, 2002), at sample size 500. The results show that although not as dramatic as the large error case, the B-spline MLE is still closer to the true pdf with narrower confidence band, hence is more precise than the deconvolution method. We also provide the similar comparison results at sample sizes 1000 and 2000 in Figures S.1 and S.2 in the supplement. For a quantitative comparison, we also computed the MAE between the true pdf curve and the estimated pdf curve in the upper part of Table 1. These results show that the B-spline

MLE performs consistently better than deconvolution.

4.2. *Performance of B-spline-assisted semiparametric estimator.* We next evaluated the finite sample performance of the B-spline semiparametric mean regression method described in Section 2.2. We used sample sizes n from 500 to 2000 and generated X from a beta distribution with both shape parameters equal to 2. The true regression mean function is $m(x) = \sin(2\pi x)$ and we generated the regression model errors ϵ from a normal distribution with mean zero variance 0.25. We generated the measurement errors U from the three different distributions described in model I (a)–(c). Cubic splines were used to approximate the mean function, and the number of interior knots were proportional to $n^{1/5}$. In the right panel of Figure 1, we plotted the averaged root-(nh_b) MAE calculated via $\sqrt{nh_b}\sup_x|\widehat{m}(x) - m(x)|$ as a function of the sample size n. Similar to the density estimation experiment, the curves stabilize as sample size increases, and is largely flattened after $n = 1000$, indicating that $\sup_x|\widehat{m}(x) - m(x)|$ has order $(nh_b)^{-1/2}$.

We further compared the B-spline semiparametric method with the deconvolution method in the nonparametric mean regression model. In the lower row of Figure 2, we plotted the average $\sup_x|\widehat{m}(x) - m(x)|$ over 200 simulations for both methods. Again, in this case, with moderate to significant amount of noise, the B-spline semiparametric method greatly outperforms the deconvolution method with smaller average error.

Similar to the pdf investigation, we reduce the measurement error variability and generated the errors from model II (a)–(c) to investigate the mean function curve fitting. We plotted the mean function estimates and the 90% confidence bands for the B-spline semiparametric and deconvolution methods in Figure 4 and Figures S.3 and S.4 in the supplement for sample sizes 500, 1000 and 2000 respectively. These results indicate that the B-spline semiparametric estimator indeed outperforms the deconvolution method. Their performance difference in terms of averaged MAE between the estimated and true mean curves are further provided in the lower half of Table 1.

5. Data Example. Heavy fine particulate matter (PM2.5) air pollution has become a serious problem in China in recent years and its possible effect on respiratory diseases has been a concern in public health. Starting from 2012, Beijing Environmental Protection Bureau (BEPB) has been recording the daily PM2.5 levels in Beijing. Based on these data, Xu et al. (2016) studied the effect of PM2.5 on asthma in year 2013. Specifically, they explored the PM2.5 effect on the number of daily asthma emergency room visits (ERV) in ten hospitals in Beijing, but found no significant effect. In fact, the mean number of daily asthma ERVs even shows a decreasing trend along the increase of measured PM2.5. This contradicts with the general conclusion that PM2.5 has short term adverse effect on asthma (Fan et al., 2016).

A potential reason of this inconsistency is the errors in the PM2.5 measurements which were not taken into account in the above analysis. In fact, there are many debates on the accuracy of the PM2.5 reports, especially in the early years such as 2013. For example, we compared the daily average PM2.5 reports in 2013 from 17 ambient air quality monitoring stations and those reports from the "Mission China Beijing" website (Mission-China, 2016) maintained by the U.S. department of state, and show the two estimated pdfs of PM2.5 in Figure 5. It is clear that the estimated pdfs of PM2.5 from the two sources are very different, where the PM2.5 concentrations obtained from the BEPB yields a pdf estimate with the mode to the left of that obtained from the "Mission China Beijing", indicating a generally less severe air pollution problem. This motivates us to consider the measurement error issue in studying the effect of PM2.5 on the daily asthma ERV.

We restrict our analysis of PM2.5 in the range from 0 to 400. The largest recorded PM2.5 value is 328. The data set we analyzed contains 337 observations. In the data available to us, the ith observation contains the number of daily asthma ERVs, which we treat as response Y_i, the average PM2.5 level over 17 ambient air quality monitoring stations from BEPB, which we denote as W_i, and the PM2.5 level from "Mission China Beijing", which we write as W_{0i}. To carry out the analysis, we let the true PM2.5 value be X_i, and let W_{ki}, U_{ki} be the observed PM2.5 value and its corresponding measurement error at the kth monitoring station, $k = 1, \ldots, 17$. We assume U_{ki}'s are independent of each other and of X_i. Then $W_i = \sum_{k=1}^{17} W_{ki}/17$. Writing $\overline{U}_i = \sum_{k=1}^{17} U_{ki}/17$, we have $W_i = X_i + \overline{U}_i$. Because our preliminary analysis result in Figure 5 suggests a possible discrepancy between the measurements in W_i and in W_{0i}, we allow a potential bias term b and model $W_{0i} = b + X_i + U_{0i}$, where U_{0i} is the measurement error of W_{0i}. We assume all the U_{ki}, $k = 0, \ldots, 17$ to have the same distribution with mean zero, and we estimate b by $\widehat{b} = n^{-1}(\sum_{i=1}^{n} W_{0i} - \sum_{i=1}^{n} W_i)$, which yields the value $\widehat{b} = 132$. We further estimated the variance of $\overline{U}_i$ based on $\text{var}(\overline{U}) = \{\text{var}(W_{0i} - W_i)\}/18$. This yields $\widehat{\text{var}}(\overline{U}) = 0.008$. Further, because $\overline{U}_i$ is the average of 17 iid U_{ki}'s, it is sensible to assume that $\overline{U}_i$ has a normal distribution.

Based on the above preliminary analysis, we proceed to estimate the pdf of PM2.5, i.e. X_i, and the mean regression function of asthma ERVs conditional on PM2.5, i.e. $E(Y_i \mid X_i)$, using the B-spline-assisted MLE/semiparametric methods in Sections 2.1 and 2.2. We further implemented 100 bootstraps to estimate the asymptotic variances of the resulting estimators. We also compared the B-spline MLE and the B-spline semiparametric regression estimator with the deconvolution density and regression estimators. In implementing the B-spline approximation, we used two and three equally spaced knots respectively, and in implementing the deconvolution methods, we used bandwidth 0.05. The number of knots is chosen based on the simulation studies in Section 4. In terms of the bandwidth selection, because the crossvalidation (Stefanski and Carroll, 1990) and the plug-in (Delaigle and Gijbels, 2002) bandwidth selection algorithms tend to select small bandwidths that induce large numerical errors, we used the smallest bandwidth that produced stable results in our analysis. The upper panel in Figure 6 shows the estimated pdfs based on the B-spline MLE and the deconvolution method. Compared with the kernel estimator in the same plots which ignores the measurement errors, the B-spline MLE shows more difference than the deconvolution estimator. This is sensible given that the noise-to-signal ratio is quite large at $\mathrm{var}(\bar{U}_i)/\mathrm{var}(X_i) = 0.2$, hence measurement error is likely not ignorable. The lower panel in Figure 6 provides the estimated mean of Y_i as a function of X_i. The B-spline semiparametric estimator shows a fluctuating pattern in the range from 0 to 200, although the pattern is not significant. In the range of PM2.5 concentration larger than 200, it shows clearly an increasing trend, which agrees with the conclusion in Fan et al. (2016) that the exposure to high PM2.5 has adverse effect on asthma onset. In contrast, the relation from the deconvolution estimator is similar to that of the local linear regression estimator ignoring the measurement errors, and it is unable to detect the increasing trend of the asthma ERVs as the PM2.5 level increases.

6. Discussion. We have developed a B-spline-assisted MLE for nonparametric pdf estimation and a B-spline-assisted semiparametric estimator for nonparametric regression mean function, under the situation that the covariates are measured with errors. The performance of both procedures are superior to the widely used deconvolution methods, in terms of both their theoretical convergence rate and their numerical performance.

A key difference between our procedures and the deconvolution procedures is that we restrict our interest in estimating the functions on a compact set. Given that practically most time, the possible range of a random variable is indeed finite, hence the relevant information is indeed only functions in a finite range, we are very curious if deconvolution methods can achieve the same convergence property in such case, with possibly some modifications on the existing procedures. To this end, Hall and Qiu (2005) provides some relevant results, while we leave this general question as an open problem for researchers who specialize in deconvolution methods.

Another curious issue is regarding kernel treatment to nonparametric functions. In many nonparametric estimation problems, kernel methods and B-spline methods generate comparable results. However, our procedures critically rely on the B-spline approximation to convert the unknown functions to "parametric" ones, hence we cannot readily replace the B-spline method with kernel method. Whether or not kernel methods can be applied in nonparametric density estimation and nonparametric regression estimation problems with measurement errors to achieve the same convergence rate as B-spline based methods remains a challenging and interesting question.

The density and regression function estimation problems studied in this work is the most basic ones. In practice, various complications may occur. For example, the error distribution $f_U(u)$ may not be known and need to be estimated parametrically or nonparametrically based on repeated measurements or other instruments. If one estimates $f_U(\cdot)$ parametrically first then insert it into our procedure, it will not have first order effect. However, if $f_U(\cdot)$ is estimated nonparametrically, it will in general affect both the bias and variance of the resulting estimation of both the density and regression functions. See Delaigle and Hall (2016) for deconvolution based estimation incorporating unknown error distribution. The story may be even more interesting when U and X are correlated hence heteroscedastic measurement errors occur. These are all very interesting problems and are worth careful investigation.

Appendix.

A.1. *Proof of Theorem 1.* Note that

$$
\begin{aligned}
& \sup_{x\in[0,1]} |\log\{\widehat{f}_X(x)\} - \log\{f_{X0}(x)\}| \\
= & \sup_{x\in[0,1]} |\mathbf{B}_r(x)^{\mathrm{T}}\widehat{\boldsymbol{\theta}} - \log\int_0^1 \exp\{\mathbf{B}_r(x)^{\mathrm{T}}\widehat{\boldsymbol{\theta}}\}dx - \log\{f_{X0}(x)\}| \\
= & \sup_{x\in[0,1]} |\mathbf{B}_r(x)^{\mathrm{T}}\widehat{\boldsymbol{\theta}} - \mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\theta}_0 + \mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\theta}_0 - \log\{f_{X0}(x)\} \\
& -\log\int_0^1 \exp\{\mathbf{B}_r(x)^{\mathrm{T}}\widehat{\boldsymbol{\theta}}\}dx + \log\int_0^1 \exp\{\mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\theta}_0\}dx \\
& -\log\int_0^1 \exp\{\mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\theta}_0\}dx|
\end{aligned}
$$

$$
\begin{aligned}
&\leq \sup_{x\in[0,1]}\{\|\mathbf{B}_r(x)\|_2\|\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0\|_2\} + \sup_{x\in[0,1]}|\mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\theta}_0 - \log\{f_{X0}(x)\}| \\
&\quad +|\log\int_0^1 \exp\{\mathbf{B}_r(x)^{\mathrm{T}}\widehat{\boldsymbol{\theta}}\}dx - \log\int_0^1 \exp\{\mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\theta}_0\}dx| \\
&\quad +|\log\int_0^1 \exp\{\mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\theta}_0\}dx| \\
&= \sup_{x\in[0,1]}\{\|\mathbf{B}_r(x)\|_2\|\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0\|_2\} + \sup_{x\in[0,1]}|\mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\theta}_0 - \log\{f_{X0}(x)\}| \\
&\quad +|\frac{\partial}{\partial\boldsymbol{\theta}^{\mathrm{T}}}\log\int_0^1 \exp\{\mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\theta}^\star\}dx(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0)| + |\log\int_0^1 \exp\{\mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\theta}_0\}dx| \\
&\leq \sup_{x\in[0,1]}\{\|\mathbf{B}_r(x)\|_2\|\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0\|_2\} + \sup_{x\in[0,1]}|\mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\theta}_0 - \log\{f_{X0}(x)\}| \\
&\quad +\|\frac{\partial}{\partial\boldsymbol{\theta}^{\mathrm{T}}}\log\int_0^1 \exp\{\mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\theta}^\star\}dx\|_2\|(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0)\|_2 + |\log\int_0^1 \exp\{\mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\theta}_0\}dx| \\
&= O_p\{(nh_b)^{-1/2} + h_b^q\},
\end{aligned}
$$

where $\boldsymbol{\theta}^\star$ is the point on the line connecting $\widehat{\boldsymbol{\theta}}$ and $\boldsymbol{\theta}_0$. The last equality holds by Condition (C5), Proposition 2, and the fact that $\log\int_0^1 \exp\{\mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\theta}_0\}dx$ is a smooth function of $\mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\theta}_0$, which implies

$$
\begin{aligned}
&|\log\int_0^1 \exp\{\mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\theta}_0\}dx| \\
&= |\log\int_0^1 \exp\{\mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\theta}_0\}dx - \log\int_0^1 \exp[\log\{f_{X0}(x)\}]dx| = O_p(h_b^q)
\end{aligned}
$$

following Condition (C5). Further for any x,

$$
\begin{aligned}
\mathrm{bias}\{\widehat{f}_X(x)\} &= E\left[\frac{\exp\{\mathbf{B}_r(x)^{\mathrm{T}}\widehat{\boldsymbol{\theta}}\}}{\int_0^1 \exp\{\mathbf{B}_r(x)^{\mathrm{T}}\widehat{\boldsymbol{\theta}}\}dx}\right] - f_{X0}(x) \\
&= E\left[\frac{\exp\{\mathbf{B}_r(x)^{\mathrm{T}}\widehat{\boldsymbol{\theta}}\}}{\int_0^1 \exp\{\mathbf{B}_r(x)^{\mathrm{T}}\widehat{\boldsymbol{\theta}}\}dx}\right] - \left[\frac{\exp\{\mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\theta}_0\}}{\int_0^1 \exp\{\mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\theta}_0\}dx}\right] \\
&\quad + \left[\frac{\exp\{\mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\theta}_0\}}{\int_0^1 \exp\{\mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\theta}_0\}dx}\right] - \left[\frac{\exp[\log\{f_{X0}(x)\}]}{\int_0^1 \exp[\log\{f_{X0}(x)\}]dx}\right] \\
&= E\left(\frac{\partial}{\partial\boldsymbol{\theta}^{\mathrm{T}}}\left[\frac{\exp\{\mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\theta}_3\}}{\int_0^1 \exp\{\mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\theta}_3\}dx}\right](\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0)\right) + O_p(h_b^q), \qquad \text{(A.1)}
\end{aligned}
$$

where $\boldsymbol{\theta}_3$ is on the line connecting $\boldsymbol{\theta}_0$ and $\widehat{\boldsymbol{\theta}}$. The third equality holds by Condition (C5). Consider the expectation in (A.1), recall that from (S.8),

$$
\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0 = -\mathbf{R}_1(\boldsymbol{\theta}^*)^{-1}(\mathbf{R}_{00} + \mathbf{R}_{01}), \qquad \text{(A.2)}
$$

where $\mathbf{R}_{00}, \mathbf{R}_{01}, \mathbf{R}_1(\boldsymbol{\theta})$ are defined in Proposition 2 and (S.10) and (S.9) respectively. Further

$$
\begin{aligned}
& E\left(\frac{\partial}{\partial \boldsymbol{\theta}^{\mathrm{T}}}\left[\frac{\exp\{\mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\theta}_3\}}{\int_0^1 \exp\{\mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\theta}_3\}dx}\right]\mathbf{R}_1(\boldsymbol{\theta}^*)^{-1}\mathbf{R}_{00}\right) \\
= \; & E\left(\frac{\partial}{\partial \boldsymbol{\theta}^{\mathrm{T}}}\left[\frac{\exp\{\mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\theta}_3\}}{\int_0^1 \exp\{\mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\theta}_3\}dx}\right]\mathbf{R}_1(\boldsymbol{\theta}^*)^{-1}\right. \\
& \left.\times \int_0^1 \left\{\frac{f_{X0}(t)f_U(W_i - t)}{\int_0^1 f_{X0}(x)f_U(W_i - x)dx} - f_{X0}(t)\right\}\mathbf{B}_r(t)dt\right) \\
= \; & \int_0^1 \int \frac{\partial}{\partial \boldsymbol{\theta}^{\mathrm{T}}}\left[\frac{\exp\{\mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\theta}_3\}}{\int_0^1 \exp\{\mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\theta}_3\}dx}\right]\mathbf{R}_1(\boldsymbol{\theta}^*)^{-1} \\
& \times \left\{\frac{f_{X0}(t)f_U(W_i - t)}{\int_0^1 f_{X0}(x)f_U(W_i - x)dx} - f_{X0}(t)\right\} f_{W0}(W_i)\mathbf{B}_r(t)dW_i dt \\
= \; & \int_0^1 \int \left(\frac{\partial}{\partial \boldsymbol{\theta}^{\mathrm{T}}}\left[\frac{\exp\{\mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\theta}_3\}}{\int_0^1 \exp\{\mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\theta}_3\}dx}\right]\mathbf{R}_1(\boldsymbol{\theta}^*)^{-1}\right. \\
& \left.\times \frac{f_{X0}(t)f_U(W_i - t)}{\int_0^1 f_{X0}(x)f_U(W_i - x)dx} f_{W0}(W_i)\mathbf{B}_r(t)\right) dW_i dt \\
& - \int_0^1 \int \frac{\partial}{\partial \boldsymbol{\theta}^{\mathrm{T}}}\left[\frac{\exp\{\mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\theta}_3\}}{\int_0^1 \exp\{\mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\theta}_3\}dx}\right]\mathbf{R}_1(\boldsymbol{\theta}^*)^{-1} f_{X0}(t) f_{W0}(W_i)\mathbf{B}_r(t)dW_i dt \\
= \; & \int_0^1 \int \frac{\partial}{\partial \boldsymbol{\theta}^{\mathrm{T}}}\left[\frac{\exp\{\mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\theta}_3\}}{\int_0^1 \exp\{\mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\theta}_3\}dx}\right]\mathbf{R}_1(\boldsymbol{\theta}^*)^{-1} f_{X0}(t) f_U(W_i - t)\mathbf{B}_r(t)dW_i dt \\
& - \int_0^1 \int \frac{\partial}{\partial \boldsymbol{\theta}^{\mathrm{T}}}\left[\frac{\exp\{\mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\theta}_3\}}{\int_0^1 \exp\{\mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\theta}_3\}dx}\right]\mathbf{R}_1(\boldsymbol{\theta}^*)^{-1} f_{X0}(t) f_U(W_i - t)\mathbf{B}_r(t)dW_i dt \\
= \; & \mathbf{0}(\text{A.3})
\end{aligned}
$$

In addition, we have

$$
\begin{aligned}
& \left| E\left(\frac{\partial}{\partial \boldsymbol{\theta}^{\mathrm{T}}}\left[\frac{\exp\{\mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\theta}_3\}}{\int_0^1 \exp\{\mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\theta}_3\}dx}\right]\mathbf{R}_1(\boldsymbol{\theta}^*)^{-1}\mathbf{R}_{01}\right)\right| \\
\text{(A.4)} \quad \leq \; & E\left(\left\|\frac{\partial}{\partial \boldsymbol{\theta}^{\mathrm{T}}}\left[\frac{\exp\{\mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\theta}_3\}}{\int_0^1 \exp\{\mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\theta}_3\}dx}\right]\right\|_2 \|\mathbf{R}_1(\boldsymbol{\theta}^*)^{-1}\|_2 \|\mathbf{R}_{01}\|_2\right).
\end{aligned}
$$

First note that,

$$
\begin{aligned}
&\left\| \frac{\partial}{\partial \boldsymbol{\theta}} \left[\frac{\exp\{\mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\theta}_3\}}{\int_0^1 \exp\{\mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\theta}_3\} dx} \right] \right\|_2 \qquad \text{(A.5)}\\
= \; & \left\{ \sum_{k=1}^{d_{\boldsymbol{\theta}}} \left(\frac{\exp\{\mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\theta}_3\} B_{rk}(x)}{\int_0^1 \exp\{\mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\theta}_3\} dx} \right. \right. \\
& \left. \left. - \frac{\exp\{\mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\theta}_3\} \int_0^1 \exp\{\mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\theta}_3\} B_{rk}(x) dx}{[\int_0^1 \exp\{\mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\theta}_3\} dx]^2} \right)^2 \right\}^{1/2} \\
\le \; & \left(\sum_{k=1}^{d_{\boldsymbol{\theta}}} \left[\frac{\exp\{\mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\theta}_3\} B_{rk}(x)}{\int_0^1 \exp\{\mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\theta}_3\} dx} \right]^2 \right)^{1/2} \\
& + \left\{ \sum_{k=1}^{d_{\boldsymbol{\theta}}} \left(\int_0^1 \frac{\exp\{\mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\theta}_3\} \exp\{\mathbf{B}_r(t)^{\mathrm{T}}\boldsymbol{\theta}_3\}}{[\int_0^1 \exp\{\mathbf{B}_r(t)^{\mathrm{T}}\boldsymbol{\theta}_3\} dt]^2} B_{rk}(t) dt \right)^2 \right\}^{1/2} \\
\le \; & \left(\left[\frac{\exp\{\mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\theta}_3\}}{\int_0^1 \exp\{\mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\theta}_3\} dx} \right]^2 \sum_{k=1}^{d_{\boldsymbol{\theta}}} B_{rk}^2(x) \right)^{1/2} \\
& + \left\{ \int_0^1 \left(\frac{\exp\{\mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\theta}_3\} \exp\{\mathbf{B}_r(t)^{\mathrm{T}}\boldsymbol{\theta}_3\}}{[\int_0^1 \exp\{\mathbf{B}_r(t)^{\mathrm{T}}\boldsymbol{\theta}_3\} dt]^2} \right)^2 \sum_{k=1}^{d_{\boldsymbol{\theta}}} B_{rk}^2(t) dt \right\}^{1/2} \\
= \; & \left(\left[\frac{\exp\{\mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\theta}_3\}}{\int_0^1 \exp\{\mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\theta}_3\} dx} \right]^2 \sum_{k=1}^{d_{\boldsymbol{\theta}}} B_{rk}^2(x) \right)^{1/2} \\
& + \left\{ \int_0^1 \left(\frac{\exp\{\mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\theta}_3\} \exp\{\mathbf{B}_r(t)^{\mathrm{T}}\boldsymbol{\theta}_3\}}{[\int_0^1 \exp\{\mathbf{B}_r(t)^{\mathrm{T}}\boldsymbol{\theta}_3\} dt]^2} \right)^2 dt \sum_{k=1}^{d_{\boldsymbol{\theta}}} B_{rk}^2(t^*) \right\}^{1/2} \\
= \; & O_p(1),
\end{aligned}
$$

where t^* is the point in $(0,1)$. In the above derivation, the first inequality holds by the trianglar inequality, the second inequality holds by Cauchy-Schwarz inequality, the second equality holds by the mean value theorem, the last equality holds because

$$
\|\mathbf{B}_r(x)\|_2^2 = \sum_{k=1}^{d_{\boldsymbol{\theta}}} B_{rk}^2(x) \le \sup_k B_{rk}(x) \sum_{k=1}^{d_{\boldsymbol{\theta}}} B_{rk}(x) = \sup_k B_{rk}(x) = O_p(1).
$$

Combining (A.4), (A.5), (S.12), (S.18), we have

$$
\begin{aligned}
& \left| E\left(\frac{\partial}{\partial \boldsymbol{\theta}^{\mathrm{T}}} \left[\frac{\exp\{\mathbf{B}_r(x)^{\mathrm{T}} \boldsymbol{\theta}_3\}}{\int_0^1 \exp\{\mathbf{B}_r(x)^{\mathrm{T}} \boldsymbol{\theta}_3\} dx} \right] \mathbf{R}_1(\boldsymbol{\theta}^*)^{-1} \mathbf{R}_{01} \right) \right| \\
(\text{A.6}) \qquad = \; & O_p(h_b^{-1} d_{\boldsymbol{\theta}}^{1/2} h_b^{q+1}) = O_p(h_b^{q-1/2}).
\end{aligned}
$$

Combining (A.6), (A.2) and (A.3), we obtain

$$
E\left(\frac{\partial}{\partial \boldsymbol{\theta}^{\mathrm{T}}} \left[\frac{\exp\{\mathbf{B}_r(x)^{\mathrm{T}} \boldsymbol{\theta}_3\}}{\int_0^1 \exp\{\mathbf{B}_r(x)^{\mathrm{T}} \boldsymbol{\theta}_3\} dx} \right] (\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) \right) = O_p(h_b^{q-1/2}).
$$

Plug this result to (A.1), we have $\text{bias}\{\widehat{f}_X(x)\} = O_p(h_b^{q-1/2})$.

Hence,

$$
\begin{aligned}
& \sqrt{nh_b} \left[\widehat{f}_X(x) - f_{X0}(x) - \text{bias}\{\widehat{f}_X(x)\} \right] \\
= \; & \sqrt{nh_b} \left[\frac{\exp\{\mathbf{B}_r(x)^{\mathrm{T}} \widehat{\boldsymbol{\theta}}\}}{\int_0^1 \exp\{\mathbf{B}_r(x)^{\mathrm{T}} \widehat{\boldsymbol{\theta}}\} dx} - f_{X0}(x) - \text{bias}\{\widehat{f}_X(x)\} \right] \\
= \; & \sqrt{nh_b} \left[\frac{\exp\{\mathbf{B}_r(x)^{\mathrm{T}} \widehat{\boldsymbol{\theta}}\}}{\int_0^1 \exp\{\mathbf{B}_r(x)^{\mathrm{T}} \widehat{\boldsymbol{\theta}}\} dx} - \frac{\exp\{\mathbf{B}_r(x)^{\mathrm{T}} \boldsymbol{\theta}_0\}}{\int_0^1 \exp\{\mathbf{B}_r(x)^{\mathrm{T}} \boldsymbol{\theta}_0\} dx} \right. \\
& \left. + \frac{\exp\{\mathbf{B}_r(x)^{\mathrm{T}} \boldsymbol{\theta}_0\}}{\int_0^1 \exp\{\mathbf{B}_r(x)^{\mathrm{T}} \boldsymbol{\theta}_0\} dx} - f_{X0}(x) - \text{bias}\{\widehat{f}_X(x)\} \right] \\
= \; & \sqrt{nh_b} \left[\frac{\exp\{\mathbf{B}_r(x)^{\mathrm{T}} \widehat{\boldsymbol{\theta}}\}}{\int_0^1 \exp\{\mathbf{B}_r(x)^{\mathrm{T}} \widehat{\boldsymbol{\theta}}\} dx} - \frac{\exp\{\mathbf{B}_r(x)^{\mathrm{T}} \boldsymbol{\theta}_0\}}{\int_0^1 \exp\{\mathbf{B}_r(x)^{\mathrm{T}} \boldsymbol{\theta}_0\} dx} \right] + O_p(h_b^q n^{1/2}) \\
= \; & \sqrt{nh_b} \frac{\partial}{\partial \boldsymbol{\theta}^{\mathrm{T}}} \left[\frac{\exp\{\mathbf{B}_r(x)^{\mathrm{T}} \boldsymbol{\theta}_0\}}{\int_0^1 \exp\{\mathbf{B}_r(x)^{\mathrm{T}} \boldsymbol{\theta}_0\} dx} \right] \left(E\left[\frac{\int_0^1 f_{X0}(x) f_U(W_i - x) \mathbf{B}_r(x) \mathbf{B}_r(x)^{\mathrm{T}} dx}{\int_0^1 f_{X0}(x) f_U(W_i - x) dx} \right. \right. \\
& - \frac{\left\{ \int_0^1 f_{X0}(x) f_U(W_i - x) \mathbf{B}_r(x) dx \right\}^{\otimes 2}}{\left\{ \int_0^1 f_{X0}(x) f_U(W_i - x) dx \right\}^2} - \int_0^1 f_{X0}(x) \mathbf{B}_r(x) \mathbf{B}_r(x)^{\mathrm{T}} dx \\
& \left. \left. + \left\{ \int_0^1 f_{X0}(x) \mathbf{B}_r(x) dx \right\}^{\otimes 2} \right] \right)^{-1} n^{-1} \sum_{i=1}^n \int_0^1 \left\{ \frac{f_{X0}(x) f_U(W_i - x)}{\int_0^1 f_{X0}(x) f_U(W_i - x) dx} \right. \\
& \left. - f_{X0}(x) \right\} \mathbf{B}_r(x) dx + o_p(1).
\end{aligned}
$$

The last equality hold by Proposition 2 and Condition (C3). This proves the results. $\square$

A.2. *Proof of Theorem 2.*

$$
\begin{aligned}
& \sup_{x\in[0,1]} |\widehat{m}(x) - m(x)| \\
= & \sup_{x\in[0,1]} |\mathbf{B}_r(x)^{\mathrm{T}}\widehat{\boldsymbol{\beta}} - \mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\beta}_0 + \mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\beta}_0 - m(x)| \\
\leq & \sup_{x\in[0,1]} |\mathbf{B}_r(x)^{\mathrm{T}}\widehat{\boldsymbol{\beta}} - \mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\beta}_0| + \sup_{x\in[0,1]} |\mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\beta}_0 - m(x)| \\
= & O_p\{(nh_b)^{-1/2} + h_b^q\},
\end{aligned}
$$

by Proposition 2 and Condition (C5).

$$
\begin{aligned}
& \mathrm{bias}\{\widehat{m}(x)\} \\
= & E\{\widehat{m}(x) - m(x)\} \\
= & E\{\mathbf{B}_r(x)^{\mathrm{T}}\widehat{\boldsymbol{\beta}} - \mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\beta}_0\} + \{\mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\beta}_0 - m(x)\} \\
= & E\{\mathbf{B}_r(x)^{\mathrm{T}}\widehat{\boldsymbol{\beta}} - \mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\beta}_0\} + O_p(h_b^q).
\end{aligned}
\tag{A.7}
$$

The last equality holds by Condition (C5). By (S.21) and Proposition 2, we can see that

$$
\begin{aligned}
& |E\{\mathbf{B}_r(x)^{\mathrm{T}}\widehat{\boldsymbol{\beta}} - \mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\beta}_0\}| \\
= & \Bigg| E\Bigg\{ \mathbf{B}_r(x)^{\mathrm{T}} E\left(\frac{\partial[\mathbf{S}^*_{\boldsymbol{\beta}}(W_i, Y_i, \boldsymbol{\beta}) - E^*\{\mathbf{a}(X, \boldsymbol{\beta}) \mid Y_i, W_i, \boldsymbol{\beta}\}]}{\partial \boldsymbol{\beta}^{\mathrm{T}}} \Bigg|_{\mathbf{B}_r(\cdot)^{\mathrm{T}}\boldsymbol{\beta} = m(\cdot)} \right)^{-1} \\
& \times (\mathbf{T}_{00} + \mathbf{T}_{01})\{1 + o_p(1)\}\Bigg\} \Bigg| \\
= & \Bigg| E\Bigg\{ \mathbf{B}_r(x)^{\mathrm{T}} E\left(\frac{\partial[\mathbf{S}^*_{\boldsymbol{\beta}}(W_i, Y_i, \boldsymbol{\beta}) - E^*\{\mathbf{a}(X, \boldsymbol{\beta}) \mid Y_i, W_i, \boldsymbol{\beta}\}]}{\partial \boldsymbol{\beta}^{\mathrm{T}}} \Bigg|_{\mathbf{B}_r(\cdot)^{\mathrm{T}}\boldsymbol{\beta} = m(\cdot)} \right)^{-1} \\
& \times \mathbf{T}_{01}\{1 + o_p(1)\}\Bigg\} \Bigg| \\
\leq & E\Bigg\{ \Bigg\| E\left(\frac{\partial[\mathbf{S}^*_{\boldsymbol{\beta}}(W_i, Y_i, \boldsymbol{\beta}) - E^*\{\mathbf{a}(X, \boldsymbol{\beta}) \mid Y_i, W_i, \boldsymbol{\beta}\}]}{\partial \boldsymbol{\beta}^{\mathrm{T}}} \Bigg|_{\mathbf{B}_r(\cdot)^{\mathrm{T}}\boldsymbol{\beta} = m(\cdot)} \right)^{-1} \Bigg\|_2 \\
& \times \|\mathbf{B}_r(x)^{\mathrm{T}}\|_2 \|\mathbf{T}_{01}\|_2 \{1 + o_p(1)\}\Bigg\} \\
= & O_p(h_b^{-1} h_b^{q+1/2}) = O_p(h_b^{q-1/2}).
\end{aligned}
$$

The second equality holds because $E(\mathbf{T}_{00}) = 0$. The fourth line holds by the fact that $\sum_{k=1}^{d_{\boldsymbol{\beta}}}\{u_k B_{rk}(x)\}^2 = M\sum_{k=1}^{d_{\boldsymbol{\beta}}} B_{rk}(x) = M$ for some finite constant

M and hence $\|\mathbf{B}_r(x)^{\mathrm{T}}\|_2 = O(1)$,

$$\left\| E\left(\frac{\partial[\mathbf{S}^*_{\boldsymbol{\beta}}(W_i, Y_i, \boldsymbol{\beta}) - E^*\{\mathbf{a}(X, \boldsymbol{\beta}) \mid Y_i, W_i, \boldsymbol{\beta}\}]}{\partial \boldsymbol{\beta}^{\mathrm{T}}} \Big|_{\mathbf{B}_r(\cdot)^{\mathrm{T}}\boldsymbol{\beta} = m(\cdot)} \right)^{-1} \right\|_2 = O_p(h_b^{-1})$$

by (S.28), and $\|\mathbf{T}_{01}\|_2 \le d_{\boldsymbol{\beta}}^{1/2} \|\mathbf{T}_{01}\|_\infty = O_p(h_b^{q+1/2})$ by (S.25). Plug this result to (A.7), we have

$$\mathrm{bias}\{\widehat{m}(x)\} = O_p(h_b^{q-1/2}).$$

Hence

$$\begin{aligned}
& \sqrt{nh_b}[\widehat{m}(x) - m(x) - \mathrm{bias}\{\widehat{m}(x)\}] \\
= \; & \sqrt{nh_b}\mathbf{B}_r(x)^{\mathrm{T}} \\
& \times \left[- \left\{ E\left(\frac{\partial[\mathbf{S}^*_{\boldsymbol{\beta}}(W_i, Y_i, \boldsymbol{\beta}) - E^*\{\mathbf{a}(X, \boldsymbol{\beta}) \mid Y_i, W_i, \boldsymbol{\beta}\}]}{\partial \boldsymbol{\beta}^{\mathrm{T}}} \Big|_{\mathbf{B}_r(\cdot)^{\mathrm{T}}\boldsymbol{\beta} = m(\cdot)} \right) \right\}^{-1} \right. \\
& \left. \times \frac{1}{n} \sum_{i=1}^n \mathbf{S}^*_{\boldsymbol{\beta}}(W_i, Y_i, m) - E^*\{\mathbf{a}(X, m) \mid Y_i, W_i, m\} \right] - \sqrt{nh_b}\mathrm{bias}\{\widehat{m}(x)\} \\
= \; & \sqrt{nh_b}\mathbf{B}_r(x)^{\mathrm{T}} \\
& \times \left[- \left\{ E\left(\frac{\partial[\mathbf{S}^*_{\boldsymbol{\beta}}(W_i, Y_i, \boldsymbol{\beta}) - E^*\{\mathbf{a}(X, \boldsymbol{\beta}) \mid Y_i, W_i, \boldsymbol{\beta}\}]}{\partial \boldsymbol{\beta}^{\mathrm{T}}} \Big|_{\mathbf{B}_r(\cdot)^{\mathrm{T}}\boldsymbol{\beta} = m(\cdot)} \right) \right\}^{-1} \right. \\
& \left. \times n^{-1} \sum_{i=1}^n \mathbf{S}^*_{\boldsymbol{\beta}}(W_i, Y_i, m) - E^*\{\mathbf{a}(X, m) \mid Y_i, W_i, m\} \right] + O_p(n^{1/2}h_b^q) \\
= \; & \sqrt{nh_b}\mathbf{B}_r(x)^{\mathrm{T}} \\
& \times \left[- \left\{ E\left(\frac{\partial[\mathbf{S}^*_{\boldsymbol{\beta}}(W_i, Y_i, \boldsymbol{\beta}) - E^*\{\mathbf{a}(X, \boldsymbol{\beta}) \mid Y_i, W_i, \boldsymbol{\beta}\}]}{\partial \boldsymbol{\beta}^{\mathrm{T}}} \Big|_{\mathbf{B}_r(\cdot)^{\mathrm{T}}\boldsymbol{\beta} = m(\cdot)} \right) \right\}^{-1} \right. \\
& \left. \times n^{-1} \sum_{i=1}^n \mathbf{S}^*_{\boldsymbol{\beta}}(W_i, Y_i, m) - E^*\{\mathbf{a}(X, m) \mid Y_i, W_i, m\} \right] + o_p(1),
\end{aligned}$$

The second equality holds by the fact that $\mathrm{bias}\{\widehat{m}(x)\} = O_p(h_b^{q-1/2})$. The last equality holds by Condition (C3). This proves the result. □

Acknowledgment. We thank Yang Li for providing the daily asthma emergency room visits data and the referees for helpful comments.

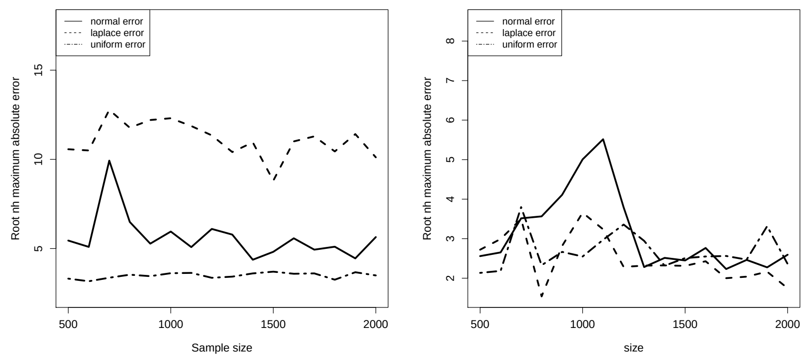


FIG 1. *Performance of B-spline MLE pdf estimation (upper) and B-spline semiparametric mean estimation (lower). Results based on 200 simulations.*

FIG 2. *Comparison of pdf (upper) and mean (lower) estimators based on the B-spline MLE or B-spline semiparametric estimator (solid) and the deconvolution (dashed) method, when measurement errors are norm (left), Laplace (middle) and uniform (right) respectively. Average maximum absolute error* $\sup_x |\widehat{f}_X(x) - f_X(x)|$ *(upper) or* $\sup_x |\widehat{m}_X(x) - m_X(x)|$ *(lower) is computed based on 200 simulations at sample sizes from 500 to 2000.*

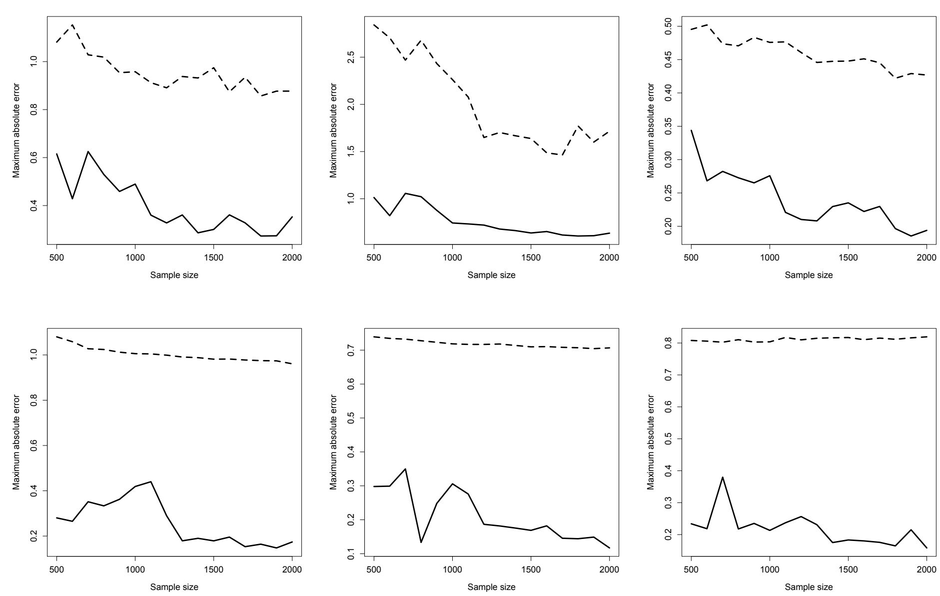


FIG 3. *B-spline MLE density estimation (left) and deconvolution estimation (right) from 200 simulations: The solid lines represent the true functions and the dash lines represent the estimated functions and their 90% confidence bands. The first row to third row are the results for model II (a)–(c) respectively. Sample size 500.*

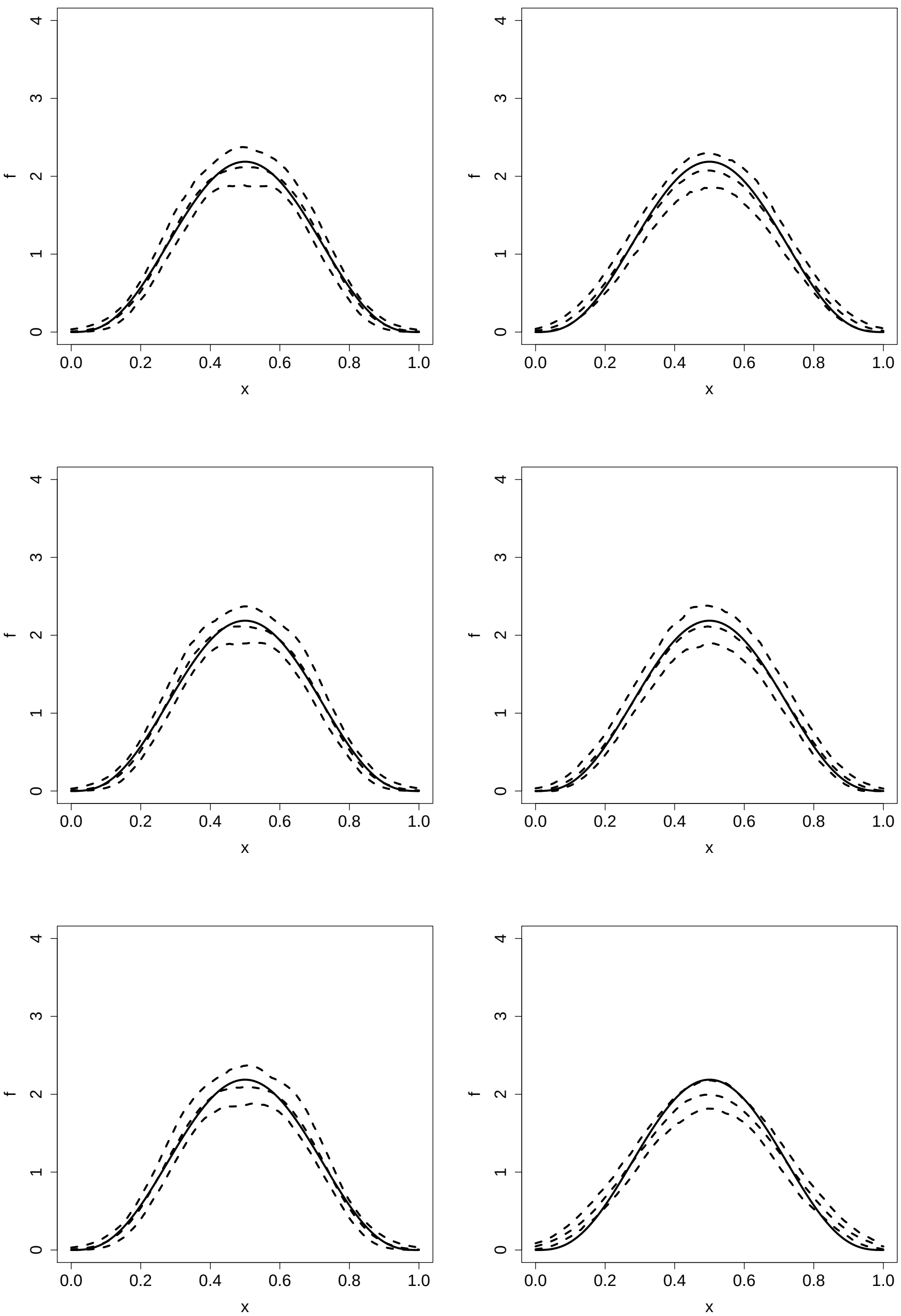


FIG 4. *B-spline semiparametric regression estimation (left) and deconvolution estimation (right) from 200 simulations: The solid lines represent the true functions and the dash lines represent the estimated functions and their 90% confidence bands. The first row to third row are the results for model II (a)–(c), respectively. Sample size 500.*

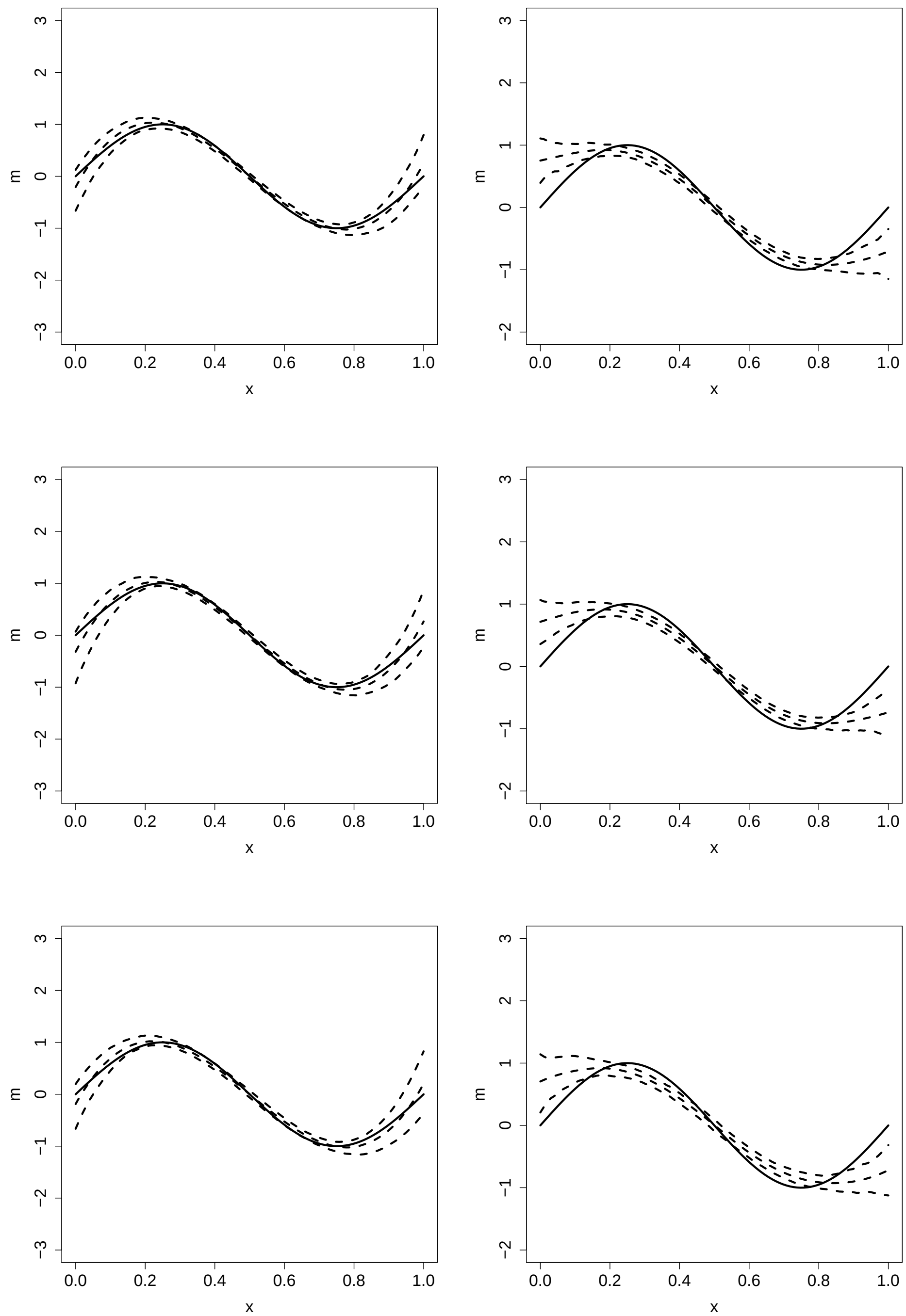


Fig 5. *The estimated pdf of PM2.5 without considering measurement error, based on data from Beijing Environmental Protection Bureau (solid line) and "Mission China" website (dash-line).*

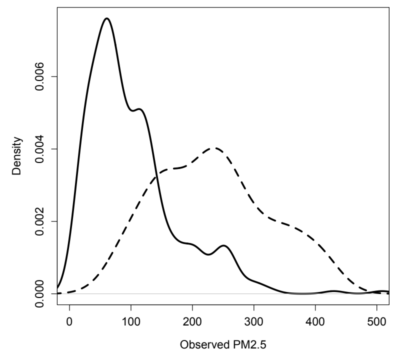


Table 1
Performance comparison between B-spline MLE/semiparametric estimator and deconvolution. Mean of the maximum absolute differences are reported. Results based on average over 200 simulations.

	pdf estimation: $E\{\sup_x \lvert\widehat{f}_{X0}(x) - f_{X0}(x)\rvert\}$					
	B-spline MLE			Deconvolution		
	$n = 500$	$n = 1000$	$n = 2000$	$n = 500$	$n = 1000$	$n = 2000$
model II(a)	0.204	0.160	0.109	0.238	0.212	0.160
model II(b)	0.213	0.153	0.110	0.230	0.197	0.158
model II(c)	0.222	0.184	0.124	0.315	0.242	0.230
	mean estimation: $E\{\sup_x \lvert\widehat{m}(x) - m(x)\rvert\}$					
	B-spline semiparametric			Deconvolution		
	$n = 500$	$n = 1000$	$n = 2000$	$n = 500$	$n = 1000$	$n = 2000$
model II(a)	0.370	0.263	0.175	0.908	0.796	0.762
model II(b)	0.425	0.264	0.163	0.857	0.828	0.779
model II(c)	0.414	0.291	0.219	0.880	0.832	0.801

References.

Berry, S. M., Carroll, R. J. and Ruppert, D. (2002). Bayesian smoothing and regression splines for measurement error problems. *Journal of the American Statistical Association* **97** 160–169.

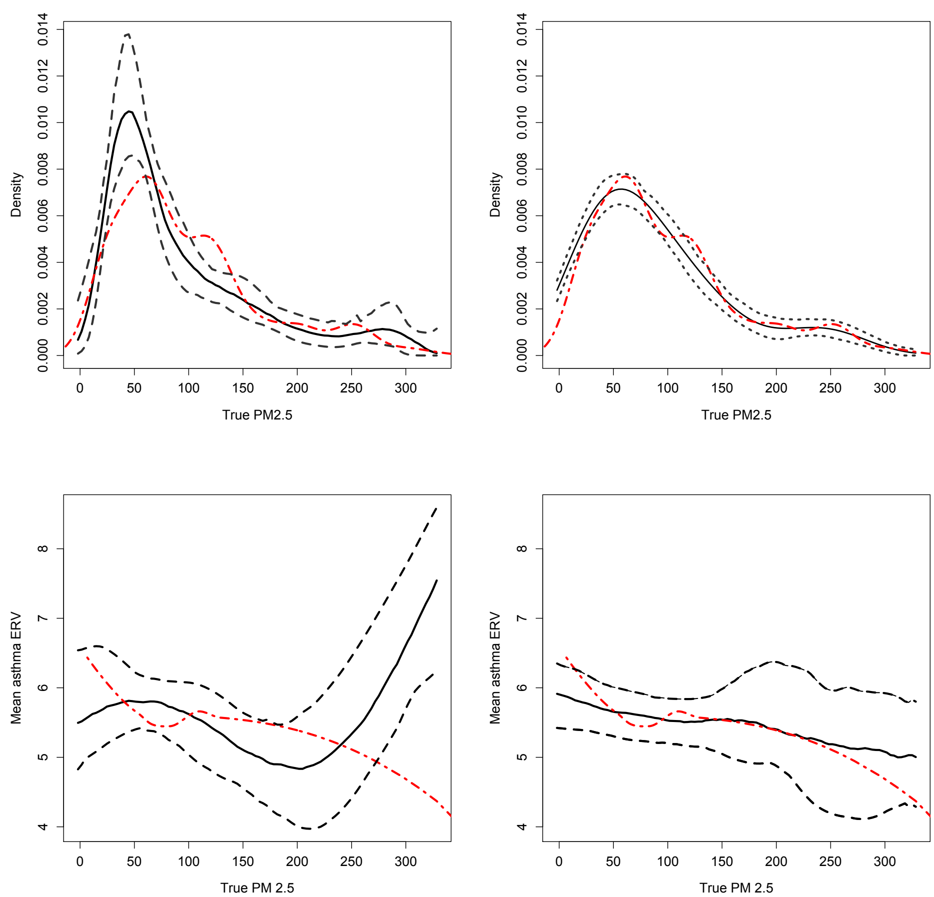
Carroll, R. J. and Hall, P. (1988). Optimal rates of convergence for deconvolving a density. *Journal of the American Statistical Association* **83** 1184-1186.

Conway, J. B. (1990). *A Course in Functional Analysis*. Springer-Verlag, New York.

De Boor, C. (1978). *A practical guide to splines* **27**. Springer-Verlag, New York.

Delaigle, A. and Gijbels, I. (2002). Estimation of integrated squared density derivatives from a contaminated sample. *Journal of the Royal Statistical Society: Series B* **64** 869–886.

FIG 6. *The B-spline MLE (upper left) and deconvolution (upper right) pdf estimators and their 95% confidence bands of PM2.5. The B-spline semiparametric (bottom left) and deconvolution (bottom right) regression estimators and their 95% confidence bands. The red line is the naive estimator ignoring measurement errors.*



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

Fan, J. (1991). On the Optimal Rates of Convergence for Nonparametric Deconvolution Problems. *Annals of Statistics* **19** 1257-1272.

Fan, J. and Truong, Y. K. (1993). Nonparametric Regression with Errors in Variables. *Annals of Statistics* **21** 1900-1925.

Fan, J., Li, S., Fan, C., Bai, Z. and Yang, K. (2016). The impact of PM2.5 on asthma emergency department visits: a systematic review and meta-analysis. *Environmental Science and Pollution Research* **23** 843–850.

Hall, P. and Qiu, P. (2005). Discrete-transform approach to deconvolution problems. *Biometrika* **92** 135–148.

Liu, M. C. and Taylor, R. L. (1989). A consistent nonparametric density estimator for the deconvolution problem. *Canadian Journal of Statistics* **17** 427-438.

Ma, S., Carroll, R. J., Liang, H. and Xu, S. (2015). Estimation and Inference in Generalized Additive Coefficient Models for Nonlinear Interactions with High-dimensional Covariates. *Annals of Statistics* **43** 2102–2131.

Masri, R. and Redner, R. A. (2005). Convergence rates for uniform B-spline density estimators on bounded and semi-infinite domains. *Nonparametric Statistics* **17** 555–582.

Mission-China (2016). Accessed:09-30. http://www.stateair.net/web/historical/1/1.html.

Staudenmayer, J., Ruppert, D. and Buonaccorsi, J. P. (2008). Density Estimation in the Presence of Heteroscedastic Measurement Error. *Journal of the American Statistical Association* **103** 726–736.

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

Tsiatis, A. A. and Ma, Y. (2004). Locally Efficient Semiparametric Estimators for Functional Measurement Error Models. *Biometrika* **91** 835-848.

Wang, L. and Yang, L. (2009). Spline estimation of single-index models. *Statistica Sinica* 765–783.

Xu, Q., Li, X., Wang, S., Wang, C., Huang, F., Gao, Q., Wu, L., Tao, L., Guo, J., Wang, W. et al. (2016). Fine particulate air pollution and hospital emergency room visits for respiratory disease in urban areas in Beijing, China, in 2013. *PloS one* **11** e0153099.

Zhang, C. H. (1990). Fourier methods for estimating mixing densities and distributions. *Annals of Statistics* **18** 806-830.

Department of Statistics
University of Hongkong
Pok Fu Lam, Hongkong
E-mail: feijiang@hku.hk

Department of Statistics
Penn State University
University Park, PA 16802
E-mail: yzm63@psu.edu

SUPPLEMENTARY DOCUMENT FOR "A SPLINE-ASSISTED SEMIPARAMETRIC APPROACH TO NONPARAMETRIC MEASUREMENT ERROR MODELS"

By Fei Jiang* and Yanyuan Ma†

University of Hong Kong and Penn State University†*

S.1. Definitions. Let $\mathcal{H}$ be the separable Hilbert space of square integrable functions on $[0,1]$, and let $<\phi_1,\phi_2>\equiv\int_0^1\phi_1(x)\phi_2(x)dx$ denote the inner product of the functions ϕ_1,ϕ_2 in $\mathcal{H}$. Let $\mathcal{B}\subset\mathcal{H}$ be the space spanned by the B-spline bases. Then the L_2 norm $\|\cdot\|_2$ is the norm induced by the inner product. We denote the functional sup norm as $\|\cdot\|_\infty$, and the L_p norm as $\|\cdot\|_p$.

In addition, for any random function $g(x)$ with the ith observation $g_i(x)$, we define $\Gamma(g)$ to be a second moment based linear operator such that

$$\begin{aligned}\Gamma(g)\phi(x) &\equiv \int_0^1 E\{g_i(x)g_i(s)\}\phi(s)ds\\ &= E\{<g_i,\phi>g_i(x)\}\end{aligned}$$

while its empirical version can be written as

$$\Gamma_n(g)\phi(x)\equiv n^{-1}\sum_{i=1}^n <g_i,\phi>g_i(x).$$

Using these definitions, we can write

$$\begin{aligned}<\Gamma(g)B_{rk},B_{rl}> &= E\left(\int_0^1 B_{rk}(x)g_i(x)dx\int_0^1 B_{rl}(x)g_i(x)dx\right),\\ <\Gamma_n(g)B_{rk},B_{rl}> &= n^{-1}\sum_{i=1}^n\int_0^1 B_{rk}(x)g_i(x)dx\int_0^1 B_{rl}(x)g_i(x)dx.\end{aligned}$$

Define $\mathbf{C}(g)$ as a $d_{\boldsymbol\theta}\times d_{\boldsymbol\theta}$ matrix with its (k,l) element $<\Gamma(g)B_{rk},B_{rl}>$, and $\widehat{\mathbf{C}}(g)$ as a $d_{\boldsymbol\theta}\times d_{\boldsymbol\theta}$ matrix matrix with its (k,l) element $<\Gamma_n(g)B_{rk},B_{rl}>$.

S.2. Lemmas and their Proofs.

<u>Lemma</u> 1. *There is a constant $D_r > 0$ such that for each spline $\sum_{k=1}^{d_{\boldsymbol{\theta}}} c_k B_{rk}$, and for each $1 \le p \le \infty$*

$$D_r \|\mathbf{c}'\|_p \le \|\sum_{k=1}^{d_{\boldsymbol{\theta}}} c_k B_{rk}\|_p \le \|\mathbf{c}'\|_p,$$

where $\mathbf{c}' = \{c_k\{(t_k - t_{k-r})/r\}^{1/p}, k = 1, \ldots, d_{\boldsymbol{\theta}}\}^{\mathrm{T}}$.

Proof: This is a direct consequence of Theorem 5.4.2 on page 145 in DeVore and Lorentz (1993). $\square$

<u>Lemma</u> 2. *Let $\Gamma(g)$ be a strictly positive definite operator on $\mathcal{B}$. Let $\mathbf{u}$ be a $d_{\boldsymbol{\theta}}$-dimensional vector with $\|\mathbf{u}\|_2 = 1$. There exist positive constants D_1, D_2, D_3, D_4 such that*

$$D_1 h_b \le \mathbf{u}^{\mathrm{T}} \mathbf{C}(g) \mathbf{u} \le D_2 h_b,$$
$$D_2^{-1} h_b^{-1} \le \mathbf{u}^{\mathrm{T}} \mathbf{C}(g)^{-1} \mathbf{u} \le D_1^{-1} h_b^{-1},$$

and

$$D_3 h_b \le \mathbf{u}^{\mathrm{T}} \widehat{\mathbf{C}}(g) \mathbf{u} \le D_4 h_b,$$
$$D_4^{-1} h_b^{-1} \le \mathbf{u}^{\mathrm{T}} \widehat{\mathbf{C}}(g)^{-1} \mathbf{u} \le D_3^{-1} h_b^{-1}$$

in probability.

Proof: First note that by the Cauchy-Schwartz inequality, Lemma 1 and Condition (A6), we have

$$\begin{aligned}
& E\left[\left\{\mathbf{u}^{\mathrm{T}} \int_0^1 \mathbf{B}_r(x) g_i(x) dx\right\}^2\right] \\
\le\ & E\left[\int_0^1 \left\{\sum_{k=1}^{d_{\boldsymbol{\theta}}} u_k B_{rk}(x)\right\}^2 dx \int_0^1 g_i(x)^2 dx\right] \\
=\ & \int_0^1 \left\{\sum_{k=1}^{d_{\boldsymbol{\theta}}} u_k B_{rk}(x)\right\}^2 dx E\left\{\int_0^1 g_i(x)^2 dx\right\} \\
\le\ & \|\mathbf{u}'\|_2^2 O(1) \\
=\ & O(h_b),
\end{aligned}$$

where $\mathbf{u}' = \{u_k\{(t_k - t_{k-r})/r\}^{1/2}, k = 1, \ldots, d_{\boldsymbol{\theta}}\}^{\mathrm{T}}$, whose L_2 norm is of order $O(h_b^{1/2})$. Thus we have $\mathbf{u}^{\mathrm{T}}\mathbf{C}(\boldsymbol{\beta})\mathbf{u} \le D_2 h_b$ for some positive constant $D_2 < \infty$.

As shown in (28) in Cardot, Ferraty and Sarda (2003), since the eigenvalues of the covariance operator $\Gamma(g)$ are strictly positive definite on $\mathcal{B}$, there is a positive constant D such that

$$< \Gamma(g)\phi, \phi > \ge D\|\phi\|^2, \text{ for } \phi \in \mathcal{B}.$$

Note that $\mathbf{B}_r^{\mathrm{T}}\mathbf{u} \in \mathcal{B}$, so we have

$$\mathbf{u}^{\mathrm{T}}\mathbf{C}(g)\mathbf{u} =< \Gamma(g)\mathbf{B}_r^{\mathrm{T}}\mathbf{u}, \mathbf{B}_r^{\mathrm{T}}\mathbf{u} > \ge D\|\mathbf{B}_r^{\mathrm{T}}\mathbf{u}\|_2 \ge D_1\|\mathbf{u}\|^2 h_b = D_1 h_b$$

for a positive constant D_1 by Lemma 1 and Condition (A4). Therefore, $D_1 h_b \le \mathbf{u}^{\mathrm{T}}\mathbf{C}(g)\mathbf{u} \le D_2 h_b$. And so $D_2^{-1}h_b^{-1} \le \mathbf{u}^{\mathrm{T}}\mathbf{C}^{-1}(g)\mathbf{u} \le D_1^{-1}h_b^{-1}$. Further, with Theorem 1.19 in Chatelin (1983), we have

$$\|\widehat{\mathbf{C}}(g) - \mathbf{C}(g)\|_2 \le \sup_{1\le l\le d_{\boldsymbol{\theta}}} \sum_{k=1}^{d_{\boldsymbol{\theta}}} \|\Gamma_n - \Gamma\|_2 | < B_{rk}, B_{rl} > |$$

As shown in Cardot, Ferraty and Sarda (2003), Lemma 5.3 in Cardot, Ferraty and Sarda (1999) implies $\|\Gamma_n - \Gamma\|_2 = o_p(n^{(h_b-1)/2})$. Further, by the property of B-spline basis, we have when $|k-l| > r$, $B_{rk}B_{rl} = 0$. Therefore, $\sup_{1\le l\le d_{\boldsymbol{\theta}}} \sum_{k=1}^{d_{\boldsymbol{\theta}}} | < B_{rk}, B_{rl} > | = O(h_b)$, which implies

$$\|\widehat{\mathbf{C}}(g) - \mathbf{C}(g)\|_2 \le o_p(h_b n^{(h_b-1)/2}). \tag{S.1}$$

Now because $h_b < 1$, combine with the result that $D_1 h_b \le \mathbf{u}^{\mathrm{T}}\mathbf{C}(g)\mathbf{u} \le D_2 h_b$, by the triangular inequality we obtain

$$\begin{aligned}
\mathbf{u}^{\mathrm{T}}\widehat{\mathbf{C}}(g)\mathbf{u} &= \mathbf{u}^{\mathrm{T}}\mathbf{C}(g)\mathbf{u} + \mathbf{u}^{\mathrm{T}}\{\widehat{\mathbf{C}}(g) - \mathbf{C}(g)\}\mathbf{u} \\
&\le D_2 h_b + \|\mathbf{u}\|_2\|\{\widehat{\mathbf{C}}(g) - \mathbf{C}(g)\}\mathbf{u}\|_2 \\
&\le D_2 h_b + \|\{\widehat{\mathbf{C}}(g) - \mathbf{C}(g)\}\|_2\|\mathbf{u}\|_2 \\
&= D_2 h_b + o_p(h_b)
\end{aligned}$$

and

$$\mathbf{u}^{\mathrm{T}}\widehat{\mathbf{C}}(g)\mathbf{u} = \mathbf{u}^{\mathrm{T}}\mathbf{C}(g)\mathbf{u} + \mathbf{u}^{\mathrm{T}}\{\widehat{\mathbf{C}}(g) - \mathbf{C}(g)\}\mathbf{u}$$

$$
\begin{aligned}
&\geq D_1 h_b - \|\mathbf{u}\|_2 \|\{\widehat{\mathbf{C}}(g) - \mathbf{C}(g)\}\mathbf{u}\|_2 \\
&\geq D_1 h_b - \|\{\widehat{\mathbf{C}}(g) - \mathbf{C}(g)\}\|_2 \|\mathbf{u}\|_2 \\
&= D_1 h_b + o_p(h_b).
\end{aligned}
$$

Thus, $D_3 h_b \leq \mathbf{u}^{\mathrm{T}}\widehat{\mathbf{C}}(g)\mathbf{u} \leq D_4 h_b$ in probability for some positive constant $D_3, D_4 < \infty$. And so $D_4^{-1} h_b \leq \mathbf{u}^{\mathrm{T}}\widehat{\mathbf{C}}(g)^{-1}\mathbf{u} \leq D_3^{-1} h_b^{-1}$ in probability. This proves the result.

$\square$

Lemma 3. *Assume $C_i(\cdot)$ is a continuous random function of $x \in [0,1]$. At each x, $|E\{C_i(x)\}| < \infty$. $\|C_i(\cdot)\|_2 < \infty$ a.s.. Then*

$$|n^{-1}\sum_{i=1}^{n}\int_0^1 B_{rk}(x)C_i(t)dx| = O_p(h_b)$$

if $E\{C_i(x)\} \neq 0$ and

$$|n^{-1}\sum_{i=1}^{n}\int_0^1 B_{rk}(x)C_i(x)dx| = O_p\{\sqrt{h_b n^{-1} log(n)}\}$$

if $E\{C_i(x)\} = 0$.

Proof: By the Bernstein's inequality in Bosq (1998), we have

$$
\begin{aligned}
&\left|1/n\sum_{i=1}^{n}\int_0^1 B_{rk}(x)C_i(x)dx - E\left\{\int_0^1 B_{rk}(x)C_i(x)dx\right\}\right| \\
&= O_p\left[\left[\sum_{i=1}^{n} E\left\{1/n\int_0^1 B_{rk}(x)C_i(x)dx\right\}^2 \mathrm{log}n\right]^{1/2}\right] \\
&= O_p\{\sqrt{h_b n^{-1}\log(n)}\}.
\end{aligned}
$$

The last equality holds from Lemma 1 by choosing $g_i = C_i$. Now if $E\{C_i(x)\} \neq 0$, then because B_{rk} is positive in the interval (t_{k-r}, t_k), and is 0 otherwise (page 88 in De Boor (1978))

$$
\begin{aligned}
\left|E\left\{\int_0^1 B_{rk}(x)C_i(x)dx\right\}\right| &= \left|\int_0^1 B_{rk}(x)E\{C_i(x)\}dx\right| \\
&= |E\{C_i(\xi)\}\int_0^1 B_{rk}(x)dx| \\
&= |E\{C_i(\xi)\}|\int_0^1 B_{rk}(x)dx
\end{aligned}
$$

$$\leq D_9(t_k - t_{k-r})$$
$$= O_p(h_b). \tag{S.2}$$

where ξ is a point in the interval $[0,1]$, D_9 is a finite constant. The second equality holds by the assumption that $C_i(\cdot)$ is continuous function in x and the mean value theorem. The inequality holds because the support of B_{rk} is the interval (t_{k-r}, t_k) and $|E\{C_i(x)\}| < \infty$ for any $x \in [0,1]$. Therefore, by Condition (A4) that $N^{-1}n(\log n)^{-1} \to \infty$ we have

$$\|n^{-1}\sum_{i=1}^{n}\int_0^1 B_{rk}(x)C_i(x)dx\|_2 = O_p(h_b)$$

for $E\{C_i(x)\} \neq 0$ and

$$\|n^{-1}\sum_{i=1}^{n}\int_0^1 B_{rk}(x)C_i(x)dx\|_2 = \sqrt{h_b n^{-1}\log(n)}$$

for $E\{C_i(x)\} = 0$. This proves the results. □

S.3. Proof of Proposition 1. First note that for any $h(x)$, we have

$$E\left(\frac{\int_0^1 \exp[\log\{f_{X0}(x)\}]f_U(W_i - x)h(x)dx}{\int_0^1 \exp[\log\{f_{X0}(x)\}]f_U(W_i - x)dx} - \frac{\int_0^1 \exp[\log\{f_{X0}(x)\}]h(x)dx}{\int_0^1 \exp[\log\{f_{X0}(x)\}]dx}\right) \tag{S.3}$$
$$= \int\left\{\frac{\int_0^1 f_{X0}(x)f_U(y_i - x)h(x)dx}{\int_0^1 f_{X0}(x)f_U(y_i - x)dx} - \int_0^1 f_{X0}(x)h(x)dx\right\} \times \int_0^1 f_{X0}(x)f_U(w_i - x)dxd\mu(y_i)$$
$$= \int_0^1\int f_{X0}(x)f_U(w_i - x)h(x)d\mu(w_i)dx - \int_0^1 f_{X0}(x)h(x)dx$$
$$= 0, \tag{S.4}$$

hence

$$E\left[\frac{\int_0^1 \exp[\log\{f_{X0}(x)\}]f_U(W_i - x)B_{rk}(x)dx}{\int_0^1 \exp[\log\{f_{X0}(x)\}]f_U(W_i - x)dx} - \frac{\int_0^1 \exp[\log\{f_{X0}(x)\}]B_{rk}(x)dx}{\int_0^1 \exp[\log\{f_{X0}(x)\}]dx}\right] = 0$$

for each $B_{rk}(x)$, $k = 1, \dots, d_{\boldsymbol{\theta}}$. Therefore, by Condition (C5), we have

$$E\left[\frac{\int_0^1 \exp\{\mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\theta}_0\} f_U(W_i - x) B_{rk}(x)dx}{\int_0^1 \exp\{\mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\theta}_0\} f_U(W_i - x)dx}\right. \\ \left. - \frac{\int_0^1 \exp\{\mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\theta}_0\} B_{rk}(x)dx}{\int_0^1 \exp\{\mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\theta}_0\}dx}\right] = O_p(h_b^q),$$

which suggests that

$$n^{-1}\sum_{i=1}^{n}\left[\frac{\int_0^1 \exp\{\mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\theta}_0\} f_U(W_i - x) B_{rk}(x)dx}{\int_0^1 \exp\{\mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\theta}_0\} f_U(W_i - x)dx}\right. \\ \left. - \frac{\int_0^1 \exp\{\mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\theta}_0\} B_{rk}(x)dx}{\int_0^1 \exp\{\mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\theta}_0\}dx}\right] = o_p(1).$$

Further, since

$$n^{-1}\sum_{i=1}^{n}\left[\frac{\int_0^1 \exp\{\mathbf{B}_r(x)^{\mathrm{T}}\widehat{\boldsymbol{\theta}}\} f_U(W_i - x) B_{rk}(x)dx}{\int_0^1 \exp\{\mathbf{B}_r(x)^{\mathrm{T}}\widehat{\boldsymbol{\theta}}\} f_U(W_i - x)dx}\right. \\ \left. - \frac{\int_0^1 \exp\{\mathbf{B}_r(x)^{\mathrm{T}}\widehat{\boldsymbol{\theta}}\} B_{rk}(x)dx}{\int_0^1 \exp\{\mathbf{B}_r(x)^{\mathrm{T}}\widehat{\boldsymbol{\theta}}\}dx}\right] = 0, \tag{S.5}$$

we have

$$n^{-1}\sum_{i=1}^{n}\left[\frac{\int_0^1 \exp\{\mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\theta}_0\} f_U(W_i - x) \mathbf{B}_r(x)dx}{\int_0^1 \exp\{\mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\theta}_0\} f_U(W_i - x)dx}\right. \\ \left. - \frac{\int_0^1 \exp\{\mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\theta}_0\} \mathbf{B}_r(x)dx}{\int_0^1 \exp\{\mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\theta}_0\}dx}\right] \\ -n^{-1}\sum_{i=1}^{n}\left[\frac{\int_0^1 \exp\{\mathbf{B}_r(x)^{\mathrm{T}}\widehat{\boldsymbol{\theta}}\} f_U(W_i - x) \mathbf{B}_r(x)dx}{\int_0^1 \exp\{\mathbf{B}_r(x)^{\mathrm{T}}\widehat{\boldsymbol{\theta}}\} f_U(W_i - x)dx}\right. \\ \left. - \frac{\int_0^1 \exp\{\mathbf{B}_r(x)^{\mathrm{T}}\widehat{\boldsymbol{\theta}}\} \mathbf{B}_r(x)dx}{\int_0^1 \exp\{\mathbf{B}_r(x)^{\mathrm{T}}\widehat{\boldsymbol{\theta}}\}dx}\right] = o_p(1) \tag{S.6}$$

element-wise. By Condition (C6), as a function of $\boldsymbol{\theta}$,

$$E\left(\left[\frac{\int_0^1 \exp\{\mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\theta}\} f_U(W_i - x) \mathbf{B}_r(x)dx}{\int_0^1 \exp\{\mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\theta}\} f_U(W_i - x)dx} - \frac{\int_0^1 \exp\{\mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\theta}\} \mathbf{B}_r(x)dx}{\int_0^1 \exp\{\mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\theta}\}dx}\right]\right)$$

has its derivative with respect to $\boldsymbol{\theta}$ being a nonsingular matrix in the neighborhood of its zero. Thus,

$$n^{-1}\sum_{i=1}^{n}\left[\frac{\int_0^1 \exp\{\mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\theta}\} f_U(W_i - x) \mathbf{B}_r(x)dx}{\int_0^1 \exp\{\mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\theta}\} f_U(W_i - x)dx} - \frac{\int_0^1 \exp\{\mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\theta}\} B_{rk}(x)dx}{\int_0^1 \exp\{\mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\theta}\}dx}\right],$$

also has nonsingular derivative with respect to $\boldsymbol{\theta}$ and hence is an invertible function of $\boldsymbol{\theta}$ in the neighborhood. Therefore, from (S.5) and (S.6), by the continuous mapping theorem, we have $\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0 = o_p(1)$ element-wise. This proves the results. □

S.4. Proof of Proposition 2. Proof: First note by using (S.3) at $h(x) = \delta(x - x_0)$ and then set $x_0 = x$, we have

$$E\left(\frac{\exp[\log\{f_{X0}(x)\}]f_U(W_i - x)}{\int_0^1 \exp[\log\{f_{X0}(x)\}]f_U(W_i - x)dx} - \frac{\exp[\log\{f_{X0}(x)\}]}{\int_0^1 \exp[\log\{f_{X0}(x)\}]dx}\right) = 0. \tag{S.7}$$

Since $\widehat{\boldsymbol{\theta}}$ is the maximum likelihood estimator, we have

$$\begin{aligned}
\mathbf{0} &= \frac{\partial}{\partial \boldsymbol{\theta}}\left[n^{-1}\sum_{i=1}^n \log \int_0^1 \exp\{\mathbf{B}_r(x)^{\mathrm{T}}\widehat{\boldsymbol{\theta}}\}f_U(W_i - x)dx \right.\\
&\quad \left. -\log \int_0^1 \exp\{\mathbf{B}_r(x)^{\mathrm{T}}\widehat{\boldsymbol{\theta}}\}dx\right]\\
&= \mathbf{R}_0 + \mathbf{R}_1(\boldsymbol{\theta}_0)(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0) + o_p\{\|\mathbf{R}_1(\boldsymbol{\theta}_0)(\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0)\|_2\}, \qquad \text{(S.8)}
\end{aligned}$$

where for any $\boldsymbol{\theta}$,

$$\begin{aligned}
\mathbf{R}_1(\boldsymbol{\theta}) &= \frac{\partial^2}{\partial \boldsymbol{\theta}\partial \boldsymbol{\theta}^{\mathrm{T}}}\left[n^{-1}\sum_{i=1}^n \log \int_0^1 \exp\{\mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\theta}\}f_U(W_i - x)dx \right.\\
&\quad \left. -\log \int_0^1 \exp\{\mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\theta}\}dx\right], \qquad \text{(S.9)}
\end{aligned}$$

and $\boldsymbol{\theta}^*$ is on the line connecting $\boldsymbol{\theta}_0$ and $\widehat{\boldsymbol{\theta}}$, and

$$\begin{aligned}
\mathbf{R}_0 &= n^{-1}\sum_{i=1}^n \frac{\int_0^1 \exp\{\mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\theta}_0\}f_U(W_i - x)\mathbf{B}_r(x)dx}{\int_0^1 \exp\{\mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\theta}_0\}f_U(W_i - x)dx}\\
&\quad - \frac{\int_0^1 \exp\{\mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\theta}_0\}\mathbf{B}_r(x)dx}{\int_0^1 \exp\{\mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\theta}_0\}dx}\\
&= \mathbf{R}_{00} + \mathbf{R}_{01},
\end{aligned}$$

where

$$\begin{aligned}
\mathbf{R}_{00} &\equiv n^{-1}\sum_{i=1}^n \int_0^1 \left(\frac{\exp[\log\{f_{X0}(x)\}]f_U(W_i - x)}{\int_0^1 \exp[\log\{f_{X0}(x)\}]f_U(W_i - x)dx}\right.\\
&\quad \left. - \frac{\exp[\log\{f_{X0}(x)\}]}{\int_0^1 \exp[\log\{f_{X0}(x)\}]dx}\right)\mathbf{B}_r(x)dx
\end{aligned}$$

$$= n^{-1}\sum_{i=1}^{n}\int_0^1 \left\{\frac{f_{X0}(x)f_U(W_i-x)}{\int_0^1 f_{X0}(x)f_U(W_i-x)dx} - f_{X0}(x)\right\}\mathbf{B}_r(x)dx$$

and

$$\begin{aligned}
\text{(S.10)}\qquad \mathbf{R}_{01} \equiv\; & n^{-1}\sum_{i=1}^{n}\int_0^1 \left[\frac{\exp\{\mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\theta}_0\}f_U(W_i-x)}{\int_0^1 \exp\{\mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\theta}_0\}f_U(W_i-x)dx}\right.\\
&\left. - \frac{\exp\{\mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\theta}_0\}}{\int_0^1 \exp\{\mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\theta}_0\}dx}\right]\mathbf{B}_r(x)dx\\
& - n^{-1}\sum_{i=1}^{n}\int_0^1 \left(\frac{\exp[\log\{f_{X0}(x)\}]f_U(W_i-x)}{\int_0^1 \exp[\log\{f_{X0}(x)\}]f_U(W_i-x)dx}\right.\\
&\left. - \frac{\exp[\log\{f_{X0}(x)\}]}{\int_0^1 \exp[\log\{f_{X0}(x)\}]dx}\right)\mathbf{B}_r(x)dx.
\end{aligned}$$

Now

$$\begin{aligned}
\|\mathbf{R}_{00}\|_\infty =\; & \sup_k \left| n^{-1}\sum_{i=1}^{n}\int_0^1 \left(\frac{\exp[\log\{f_{X0}(x)\}]f_U(W_i-x)}{\int_0^1 \exp[\log\{f_{X0}(x)\}]f_U(W_i-x)dx}\right.\right.\\
&\left.\left. - \frac{\exp[\log\{f_{X0}(x)\}]}{\int_0^1 \exp[\log\{f_{X0}(x)\}]dx}\right) B_{rk}(x)dx\right|\\
\text{(S.11)}\qquad =\; & O_p\{\sqrt{h_b n^{-1}\log(n)}\}.
\end{aligned}$$

The last equality holds by Lemma 3 in the supplementary material with

$$C_i(x) = \frac{f_{X0}(x)f_U(W_i-x)}{\int_0^1 f_{X0}(x)f_U(W_i-x)dx} - \frac{f_{X0}(x)}{\int_0^1 f_{X0}(x)dx},$$

which satisfies $E\{C_i(x)\} = 0$ by (S.7), and $\|C_i(x)\|_2 < \infty$ by Condition (C1). Further,

$$\begin{aligned}
\text{(S.12)}\qquad & \|\mathbf{R}_{01}\|_\infty\\
=\; & \sup_k \left[\left| n^{-1}\sum_{i=1}^{n}\int_0^1 \left\{\left[\frac{\exp\{\mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\theta}_0\}f_U(W_i-x)}{\int_0^1 \exp\{\mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\theta}_0\}f_U(W_i-x)dx}\right.\right.\right.\right.\\
&\left. - \frac{\exp\{\mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\theta}_0\}}{\int_0^1 \exp\{\mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\theta}_0\}dx}\right] - \left(\frac{\exp[\log\{f_{X0}(x)\}]f_U(W_i-x)}{\int_0^1 \exp\{f_{X0}(x)\}f_U(W_i-x)dx}\right.\\
&\left.\left.\left.\left. - \frac{\exp[\log\{f_{X0}(x)\}]}{\int_0^1 \exp[\log\{f_{X0}(x)\}]dx}\right)\right\} B_{rk}(x)dx\right|\right]\\
=\; & h_b^q \sup_k \left[\left| n^{-1}\sum_{i=1}^{n} h_b^{-q}\int_0^1 \left\{\left[\frac{\exp\{\mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\theta}_0\}f_U(W_i-x)}{\int_0^1 \exp\{\mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\theta}_0\}f_U(W_i-x)dx}\right.\right.\right.\right.
\end{aligned}$$

$$
\begin{aligned}
& -\frac{\exp\{\mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\theta}_0\}}{\int_0^1 \exp\{\mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\theta}_0\}dx}\Bigg] - \Bigg(\frac{\exp[\log\{f_{X0}(x)\}]f_U(W_i-x)}{\int_0^1 \exp[\log\{f_{X0}(x)\}]f_U(W_i-x)dx} \\
& -\frac{\exp[\log\{f_{X0}(x)\}]}{\int_0^1 \exp[\log\{f_{X0}(x)\}]dx}\Bigg)\Bigg\} B_{rk}(x)dx\Bigg|\Bigg] \\
= \;& O_p(h_b^{q+1}).
\end{aligned}
$$

The last equality holds by using Lemma 3 in the supplementary material. with

$$
\begin{aligned}
& C_i(x) \\
= \;& h_b^{-q}\Bigg\{\Bigg[\frac{\exp\{\mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\theta}_0\}f_U(W_i-x)}{\int_0^1 \exp\{\mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\theta}_0\}f_U(W_i-x)dx} - \frac{\exp\{\mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\theta}_0\}}{\int_0^1 \exp\{\mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\theta}_0\}dx}\Bigg] \\
& - \Bigg(\frac{\exp[\log\{f_{X0}(x)\}]f_U(W_i-x)}{\int_0^1 \exp[\log\{f_{X0}(x)\}]f_U(W_i-x)dx} - \frac{\exp[\log\{f_{X0}(x)\}]}{\int_0^1 \exp[\log\{f_{X0}(x)\}]dx}\Bigg)\Bigg\},
\end{aligned}
$$

which satisfies $\|C_i(x)\|_2 < \infty$. Combining (S.11) and (S.12) we have $\mathbf{R}_0 = \mathbf{R}_{00}\{1 + o_p(1)\}$ by Conditions (C3) and (C4). Further,

$$
\begin{aligned}
& E(\mathbf{R}_{00}^{\mathrm{T}}\mathbf{R}_{00}) \\
\leq \;& n^{-1}E\left[\left\{\sup_{\xi}\left|\frac{f_{X0}(\xi)f_U(W_i-\xi)}{\int_0^1 f_{X0}(x)f_U(W_i-x)dx} - f_{X0}(\xi)\right|\right\}^2 \sum_{l=k}^{d_{\boldsymbol{\theta}}}\left\{\int_0^1 B_{rk}(x)dx\right\}^2\right] \\
= \;& O_p(n^{-1}h_b).
\end{aligned}
$$

where ξ is a point in $(0,1)$. The last equality holds because

$$
\sup_{\xi}\left|\frac{f_{X0}(\xi)f_U(W_i-\xi)}{\int_0^1 f_{X0}(x)f_U(W_i-x)dx} - f_{X0}(\xi)\right|
$$

is bounded by Condition (C1). Further,

$$
\begin{aligned}
\sum_{k=1}^{d_{\boldsymbol{\theta}}}\left\{\int_0^1 B_{rk}(x)dx\right\}^2 &= \sum_{k=1}^{d_{\boldsymbol{\theta}}} D_k(t_k - t_{k-r})^2 \\
&= \sup_k |t_k - t_{k-r}|D \\
&= O_p(h_b),
\end{aligned}
$$

for some finite constant D_k and D. The second equality holds because the support of B_{rk} is the interval (t_{k-r}, t_k). The last equality holds because D_k is bounded above since $B_{rk}(x) < 1$ for any x. Therefore,

$$
\|\mathbf{R}_0\|_2 = O_p(n^{-1/2}h_b^{1/2}). \tag{S.13}
$$

To analyze $\mathbf{R}_1(\boldsymbol{\theta}_0)$, we first have that for any $\mathbf{u}$ such that $\|\mathbf{u}\|_2 = 1$, from Lemma 1 in the supplementary material,

$$
\begin{aligned}
&\mathbf{u}^{\mathrm{T}} \frac{\int_0^1 \mathbf{B}_r(x)\mathbf{B}_r^{\mathrm{T}}(x) \exp\{\mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\theta}_0\} f_U(W_i - x)dx}{\int_0^1 \exp\{\mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\theta}_0\} f_U(W_i - x)dx}\mathbf{u} \\
= & \left\| \mathbf{B}_r(x)^{\mathrm{T}} \frac{\mathbf{u}\exp\{\mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\theta}_0/2\} f_U^{1/2}(W_i - x)}{[\int_0^1 \exp\{\mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\theta}_0\} f_U(W_i - x)dx]^{1/2}} \right\|_2^2 \\
= & O_p\left[h_b \left\| \frac{\mathbf{u}\exp\{\mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\theta}_0/2\} f_U^{1/2}(W_i - x)}{[\int_0^1 \exp\{\mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\theta}_0\} f_U(W_i - x)dx]^{1/2}} \right\|_2^2 \right] \\
= & O_p(h_b),
\end{aligned}
$$

and

$$
\begin{aligned}
\mathbf{u}^{\mathrm{T}} \frac{\int_0^1 \mathbf{B}_r(x)\mathbf{B}_r^{\mathrm{T}}(x) \exp\{\mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\theta}_0\}dx}{\int_0^1 \exp\{\mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\theta}_0\}dx}\mathbf{u} &= \left\| \mathbf{B}_r(x)^{\mathrm{T}} \frac{\mathbf{u}\exp\{\mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\theta}_0/2\}}{[\int_0^1 \exp\{\mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\theta}_0\}dx]^{1/2}} \right\|_2^2 \\
&= O\left[h_b \left\| \frac{\mathbf{u}\exp\{\mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\theta}_0/2\}}{[\int_0^1 \exp\{\mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\theta}_0\}dx]^{1/2}} \right\|_2^2 \right] \\
&= O(h_b).
\end{aligned}
$$

Similarly

$$
\begin{aligned}
&\mathbf{u}^{\mathrm{T}} \frac{\int_0^1 \mathbf{B}_r(x) \exp\{\mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\theta}_0\} f_U(W_i - x)dx \int_0^1 \mathbf{B}_r^{\mathrm{T}}(x) \exp\{\mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\theta}_0\}dx}{\int_0^1 \exp\{\mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\theta}_0\} f_U(W_i - x)dx \int_0^1 \exp\{\mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\theta}_0\}dx}\mathbf{u} \\
\leq & \left[\frac{\int_0^1 \mathbf{u}^{\mathrm{T}}\mathbf{B}_r(x) \exp\{\mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\theta}_0\} f_U(W_i - x)dx}{\int_0^1 \exp\{\mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\theta}_0\} f_U(W_i - x)dx} \right]^2 \\
& + \left[\frac{\int_0^1 \mathbf{B}_r^{\mathrm{T}}(x)\mathbf{u} \exp\{\mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\theta}_0\}dx}{\int_0^1 \exp\{\mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\theta}_0\}dx} \right]^2,
\end{aligned}
$$

while

$$
\begin{aligned}
&\left[\frac{\int_0^1 \mathbf{B}_r^{\mathrm{T}}(x)\mathbf{u} \exp\{\mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\theta}_0\} f_U(W_i - x)dx}{\int_0^1 \exp\{\mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\theta}_0\} f_U(W_i - x)dx} \right]^2 \\
\leq & \left\| \frac{\mathbf{B}_r^{\mathrm{T}}(x)\mathbf{u} \exp\{\mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\theta}_0\} f_U(W_i - x)}{\int_0^1 \exp\{\mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\theta}_0\} f_U(W_i - x)dx} \right\|_1^2 \\
\leq & O_p\left[h_b \left\| \frac{\mathbf{u} \exp\{\mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\theta}_0\} f_U(W_i - x)}{\int_0^1 \exp\{\mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\theta}_0\}dx f_U(W_i - x)} \right\|_1 \right]^2 \\
\leq & O_p(h_b^2),
\end{aligned}
$$

and

$$
\begin{aligned}
\left[\frac{\int_0^1 \mathbf{B}_r^{\mathrm{T}}(x)\mathbf{u}\exp\{\mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\theta}_0\}dx}{\int_0^1 \exp\{\mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\theta}_0\}dx}\right]^2 &\leq \left\|\frac{\mathbf{B}_r^{\mathrm{T}}(x)\mathbf{u}\exp\{\mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\theta}_0\}}{\int_0^1 \exp\{\mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\theta}_0\}dx}\right\|_1^2 \\
&\leq O\left[h_b\left\|\frac{\mathbf{u}\exp\{\mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\theta}_0\}}{\int_0^1 \exp\{\mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\theta}_0\}dx}\right\|_1\right]^2 \\
&\leq O(h_b^2).
\end{aligned}
$$

Combining these results, we have

$$
\begin{aligned}
&\left|\mathbf{u}^{\mathrm{T}}\frac{\partial^2 f_W(w,\boldsymbol{\theta}_0)/\partial\boldsymbol{\theta}_0\partial\boldsymbol{\theta}_0^{\mathrm{T}}}{f_W(w,\boldsymbol{\theta}_0)}\mathbf{u}\right| \\
= &\left|\mathbf{u}^{\mathrm{T}}\left[\frac{\int_0^1 \mathbf{B}_r(x)\mathbf{B}_r^{\mathrm{T}}(x)\exp\{\mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\theta}_0\}f_U(w-x)dx}{\int_0^1 \exp\{\mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\theta}_0\}f_U(w-x)dx}\right.\right. \\
&-\frac{\int_0^1 \mathbf{B}_r(x)\exp\{\mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\theta}_0\}f_U(w-x)dx\int_0^1 \mathbf{B}_r^{\mathrm{T}}(x)\exp\{\mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\theta}_0\}dx}{\int_0^1 \exp\{\mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\theta}_0\}f_U(w-x)dx\int_0^1 \exp\{\mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\theta}_0\}dx} \\
&-\frac{\int_0^1 \mathbf{B}_r(x)\exp\{\mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\theta}_0\}dx\int_0^1 \mathbf{B}_r^{\mathrm{T}}(x)\exp\{\mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\theta}_0\}f_U(w-x)dx}{\int_0^1 \exp\{\mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\theta}_0\}dx\int_0^1 \exp\{\mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\theta}_0\}f_U(w-x)dx} \\
&-\frac{\int_0^1 \mathbf{B}_r(x)\mathbf{B}_r^{\mathrm{T}}(x)\exp\{\mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\theta}_0\}dx}{\int_0^1 \exp\{\mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\theta}_0\}dx} \\
&\left.\left.+2\frac{\int_0^1 \mathbf{B}_r(x)\exp\{\mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\theta}_0\}dx\int_0^1 \mathbf{B}_r^{\mathrm{T}}(x)\exp\{\mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\theta}_0\}dx}{\{\int_0^1 \exp\{\mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\theta}_0\}dx\}^2}\right]\mathbf{u}\right| \\
= &\ O_p(h_b),
\end{aligned}
$$

and

$$
\begin{aligned}
&E\left[\left\{\mathbf{u}^{\mathrm{T}}\frac{\partial^2 f_W(W_i,\boldsymbol{\theta}_0)/\partial\boldsymbol{\theta}_0\partial\boldsymbol{\theta}_0^{\mathrm{T}}}{f_W(W_i,\boldsymbol{\theta}_0)}\mathbf{u}\right\}^2\right] \\
= &\ E\left\{\left(\mathbf{u}^{\mathrm{T}}\left[\frac{\int_0^1 \mathbf{B}_r(x)\mathbf{B}_r^{\mathrm{T}}(x)\exp\{\mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\theta}_0\}f_U(W_i-x)dx}{\int_0^1 \exp\{\mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\theta}_0\}f_U(W_i-x)dx}\right.\right.\right. \\
&-\frac{\int_0^1 \mathbf{B}_r(x)\exp\{\mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\theta}_0\}f_U(W_i-x)dx\int_0^1 \mathbf{B}_r^{\mathrm{T}}(x)\exp\{\mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\theta}_0\}dx}{\int_0^1 \exp\{\mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\theta}_0\}f_U(W_i-x)dx\int_0^1 \exp\{\mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\theta}_0\}dx} \\
&-\frac{\int_0^1 \mathbf{B}_r(x)\exp\{\mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\theta}_0\}dx\int_0^1 \mathbf{B}_r^{\mathrm{T}}(x)\exp\{\mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\theta}_0\}f_U(W_i-x)dx}{\int_0^1 \exp\{\mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\theta}_0\}dx\int_0^1 \exp\{\mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\theta}_0\}f_U(W_i-x)dx} \\
&-\frac{\int_0^1 \mathbf{B}_r(x)\mathbf{B}_r^{\mathrm{T}}(x)\exp\{\mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\theta}_0\}dx}{\int_0^1 \exp\{\mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\theta}_0\}dx} \\
&\left.\left.\left.+2\frac{\int_0^1 \mathbf{B}_r(x)\exp\{\mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\theta}_0\}dx\int_0^1 \mathbf{B}_r^{\mathrm{T}}(x)\exp\{\mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\theta}_0\}dx}{\{\int_0^1 \exp\{\mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\theta}_0\}dx\}^2}\right]\mathbf{u}\right)^2\right\}
\end{aligned}
$$

$$= \quad O_p(h_b^2).$$

Now using Bernstein's inequality, for $\epsilon = h_b/\log(n)$,

$$
\begin{aligned}
& \mathrm{pr}\left[\mathbf{u}^{\mathrm{T}}\mathbf{R}_1(\boldsymbol{\theta}_0)\mathbf{u} - \mathbf{u}^{\mathrm{T}}E\{\mathbf{R}_1(\boldsymbol{\theta}_0)\}\mathbf{u} > \epsilon\right] \\
\leq\ & \exp\left[\frac{-n^2\epsilon^2/2}{nE(\mathbf{u}^{\mathrm{T}}[\partial^2 f_W(W_i,\boldsymbol{\theta}_0)/\partial\boldsymbol{\theta}_0\partial\boldsymbol{\theta}_0^{\mathrm{T}}/f_W(W_i,\boldsymbol{\theta}_0) - E\{\mathbf{R}_1(\boldsymbol{\theta}_0)\}]\mathbf{u})^2 + nh_b\epsilon/3}\right] \\
\leq\ & \exp\left(\frac{-n\epsilon^2/2}{E[\{\mathbf{u}^{\mathrm{T}}\partial^2 f_W(W_i,\boldsymbol{\theta}_0)/\partial\boldsymbol{\theta}_0\partial\boldsymbol{\theta}_0^{\mathrm{T}}/f_W(W_i,\boldsymbol{\theta}_0)\mathbf{u}\}^2] + h_b\epsilon/3}\right) \\
\leq\ & \exp\left(\frac{-n\epsilon^2/2}{h_b^2 + h_b\epsilon/3}\right) \to 0
\end{aligned}
$$

when $n \to \infty$. Thus,

$$|\mathbf{u}^{\mathrm{T}}\mathbf{R}_1(\boldsymbol{\theta}_0)\mathbf{u} - \mathbf{u}^{\mathrm{T}}E\{\mathbf{R}_1(\boldsymbol{\theta}_0)\}\mathbf{u}| = o_p(h_b). \tag{S.14}$$

Further,

$$
\begin{aligned}
& E\{\mathbf{R}_1(\boldsymbol{\theta}_0)\} \\
=\ & E\left\{\frac{\partial^2 f_W(W_i,\boldsymbol{\theta}_0)/\partial\boldsymbol{\theta}_0\partial\boldsymbol{\theta}_0^{\mathrm{T}}}{f_W(W_i,\boldsymbol{\theta}_0)}\right\} - E\left[\left\{\frac{\partial \log f_W(W_i,\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}_0}\right\}^{\otimes 2}\right] \\
=\ & \int\left\{\frac{\partial^2 f_W(w,\boldsymbol{\theta}_0)/\partial\boldsymbol{\theta}_0\partial\boldsymbol{\theta}_0^{\mathrm{T}}}{f_W(w,\boldsymbol{\theta}_0)}\right\} f_{W0}(w)dw - E\left[\left\{\frac{\partial \log f_W(W_i,\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}_0}\right\}^{\otimes 2}\right] \\
=\ & O_p(h_b^{q-1}) - E\left[\left\{\frac{\partial \log f_W(W_i,\boldsymbol{\theta}_0)}{\partial\boldsymbol{\theta}_0}\right\}^{\otimes 2}\right],
\end{aligned} \tag{S.15}
$$

where $f_{W0}(w)$ is the true density of W. The last equality holds by Condition (C5) so that

$$
\begin{aligned}
& \left\| \int\left\{\frac{\partial^2 f_W(w,\boldsymbol{\theta}_0)/\partial\boldsymbol{\theta}_0\partial\boldsymbol{\theta}_0^{\mathrm{T}}}{f_W(w,\boldsymbol{\theta}_0)}\right\} f_{W0}(w)dw \right. \\
& \left. - \int\left\{\frac{\partial^2 f_W(w,\boldsymbol{\theta}_0)/\partial\boldsymbol{\theta}_0\partial\boldsymbol{\theta}_0^{\mathrm{T}}}{f_W(w,\boldsymbol{\theta}_0)}\right\} f_W(w,\boldsymbol{\theta}_0)dw \right\|_2 = O_p(h_b^{q-1})
\end{aligned}
$$

and the fact that $\int f_W(w,\boldsymbol{\theta})dw = 1$ so

$$
\begin{aligned}
& \int\left\{\frac{\partial^2 f_W(w,\boldsymbol{\theta}_0)/\partial\boldsymbol{\theta}_0\partial\boldsymbol{\theta}_0^{\mathrm{T}}}{f_W(w,\boldsymbol{\theta}_0)}\right\} f_W(w,\boldsymbol{\theta}_0)dw \\
=\ & \frac{\partial^2 \int f_W(w,\boldsymbol{\theta}_0)dw}{\partial\boldsymbol{\theta}_0\partial\boldsymbol{\theta}_0^{\mathrm{T}}} \\
=\ & \mathbf{0}.
\end{aligned}
$$

Now

$$E\left[\left\{\frac{\partial \log f_W(W_i,\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\right\}^{\otimes 2}\right] \tag{S.16}$$
$$= E\left(\left[\frac{\int_0^1 \exp\{\mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\theta}\}f_U(W_i-x)\mathbf{B}_r(x)dx}{\int_0^1 \exp\{\mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\theta}\}f_U(W_i-x)dx} - \frac{\int_0^1 \exp\{\mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\theta}\}\mathbf{B}_r(x)dx}{\int_0^1 \exp\{\mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\theta}\}dx}\right]^{\otimes 2}\right).$$

First,

$$E\left(\left[\frac{\int_0^1 \exp\{\mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\theta}\}f_U(W_i-x)\mathbf{B}_r(x)dx}{\int_0^1 \exp\{\mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\theta}\}f_U(W_i-x)dx} - \frac{\int_0^1 \exp\{\mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\theta}\}\mathbf{B}_r(x)dx}{\int_0^1 \exp\{\mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\theta}\}dx}\right]^{\otimes 2}\right) \neq 0$$

by Condition (C6). Also, it is easy to see that,

$$E\left(\left[\frac{\int_0^1 \exp\{\mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\theta}\}f_U(W_i-x)\mathbf{B}_r(x)dx}{\int_0^1 \exp\{\mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\theta}\}f_U(W_i-x)dx} - \frac{\int_0^1 \exp\{\mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\theta}\}\mathbf{B}_r(x)dx}{\int_0^1 \exp\{\mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\theta}\}dx}\right]^{\otimes 2}\right)$$

is $\mathbf{C}(g)$ defined in Section S.1 in the supplementary material with

$$g_i(x) = \frac{\exp\{\mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\theta}\}f_U(W_i-x)}{\int_0^1 \exp\{\mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\theta}\}f_U(W_i-x)dx} - \frac{\exp\{\mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\theta}\}}{\int_0^1 \exp\{\mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\theta}\}dx}.$$

First note that $\Gamma(g)$ is a strictly positive definite operator on $\mathcal{B}$. We prove this by contradiction. Note that $\Gamma(g)$ is obviously positive definite. Assume $\Gamma(g)$ is not strictlyy positive definite, then it has a zero eigenvalue when operating on $\mathcal{B}$, hence the matix

$$< \Gamma(g)\mathbf{B}_r, \mathbf{B}_r > = E\left(\left[\frac{\int_0^1 \exp\{\mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\theta}\}f_U(W_i-x)\mathbf{B}_r(x)dx}{\int_0^1 \exp\{\mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\theta}\}f_U(W_i-x)dx} - \frac{\int_0^1 \exp\{\mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\theta}\}\mathbf{B}_r(x)dx}{\int_0^1 \exp\{\mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\theta}\}dx}\right]^{\otimes 2}\right)$$

has a zero eigenvalue. By (S.15) and (S.16), this implies $E\{\mathbf{R}_1(\boldsymbol{\theta}_0)\}$ is not be invertible, which contradicts with Condition (C6) that the score function has unique solution around the truth. Hence by using Lemma 2 in the supplementary material, for $\mathbf{u}$ such that $\|\mathbf{u}\|_2 = 1$, we have that there exist constants $C > c > 0$, such that

$$ch_b \leq \mathbf{u}^{\mathrm{T}} E\left(\left[\frac{\int_0^1 \exp\{\mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\theta}\}f_U(W_i-x)\mathbf{B}_r(x)dx}{\int_0^1 \exp\{\mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\theta}\}f_U(W_i-x)dx}\right.\right. \tag{S.17}$$

$$-\frac{\int_0^1 \exp\{\mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\theta}\}\mathbf{B}_r(x)dx}{\int_0^1 \exp\{\mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\theta}\}dx}\Bigg]^{\otimes 2}\Bigg)\mathbf{u} \le Ch_b.$$

Combining the results in (S.14), (S.15), (S.16) and (S.17), there exist constants $0 < c_1 < C_1$, such that

$$-C_1 h_b \le \mathbf{u}^{\mathrm{T}}\mathbf{R}_1(\boldsymbol{\theta}_0)\mathbf{u} \le -c_1 h_b$$

in probability, and in turn

$$-c_1^{-1}h_b^{-1} \le \mathbf{u}^{\mathrm{T}}\mathbf{R}_1(\boldsymbol{\theta}_0)^{-1}\mathbf{u} \le -C_1^{-1}h_b^{-1} \tag{S.18}$$

in probability. Therefore, red $\|\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0\|_2 = O_p(\|\mathbf{R}_1(\boldsymbol{\theta}_0)^{-1}\mathbf{R}_0\|_2) \le O_p\{\|\mathbf{R}_1(\boldsymbol{\theta}_0)^{-1}\|_2\|\mathbf{R}_0\|_2\} = O_p\{(nh_b)^{-1/2}\}$. Therefore, for any vector $\mathbf{a}$ with $\|\mathbf{a}\|_2 = O_p(1)$ we have $|\mathbf{a}^{\mathrm{T}}(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0)| \le \|\mathbf{a}\|_2\|(\widehat{\boldsymbol{\theta}}-\boldsymbol{\theta}_0)\|_2 = O_p\{(nh_b)^{-1/2}\}$.

Further, expanding the second derivative, we can write

$$\begin{aligned}
&\mathbf{R}_1(\boldsymbol{\theta}_0)\\
= \;& E\Bigg(\frac{\int_0^1 \exp[\log\{f_{X0}(x)\}]f_U(W_i-x)\mathbf{B}_r(x)\mathbf{B}_r(x)^{\mathrm{T}}dx}{\int_0^1 \exp[\log\{f_{X0}(x)\}]f_U(W_i-x)dx}\\
&-\frac{\left[\int_0^1 \exp[\log\{f_{X0}(x)\}]f_U(W_i-x)\mathbf{B}_r(x)dx\right]^{\otimes 2}}{\left[\int_0^1 \exp[\log\{f_{X0}(x)\}]f_U(W_i-x)dx\right]^2}\\
&-\frac{\int_0^1 \exp[\log\{f_{X0}(x)\}]\mathbf{B}_r(x)\mathbf{B}_r(x)^{\mathrm{T}}dx}{\int_0^1 \exp[\log\{f_{X0}(x)\}]dx}\\
&+\frac{\left[\int_0^1 \exp[\log\{f_{X0}(x)\}]\mathbf{B}_r(x)dx\right]^{\otimes 2}}{\left[\int_0^1 \exp[\log\{f_{X0}(x)\}]dx\right]^2}\Bigg)\{1+o_p(1)\}\\
= \;& E\Bigg[\frac{\int_0^1 f_{X0}(x)f_U(W_i-x)\mathbf{B}_r(x)\mathbf{B}_r(x)^{\mathrm{T}}dx}{\int_0^1 f_{X0}(x)f_U(W_i-x)dx}\\
&-\frac{\left\{\int_0^1 f_{X0}(x)f_U(W_i-x)\mathbf{B}_r(x)dx\right\}^{\otimes 2}}{\left\{\int_0^1 f_{X0}(x)f_U(W_i-x)dx\right\}^2}\\
&-\int_0^1 f_{X0}(x)\mathbf{B}_r(x)\mathbf{B}_r(x)^{\mathrm{T}}dx + \left\{\int_0^1 f_{X0}(x)\mathbf{B}_r(x)dx\right\}^{\otimes 2}\Bigg]\{1+o_p(1)\}.
\end{aligned}$$

Hence we can write

$$\widehat{\boldsymbol{\theta}} - \boldsymbol{\theta}_0$$

$$
\begin{aligned}
= \quad & \Bigg(E\Bigg[\frac{\int_0^1 f_{X0}(x) f_U(W_i - x)\mathbf{B}_r(x)\mathbf{B}_r(x)^{\mathrm{T}} dx}{\int_0^1 f_{X0}(x) f_U(W_i - x) dx} \\
& - \frac{\left\{\int_0^1 f_{X0}(x) f_U(W_i - x)\mathbf{B}_r(x) dx\right\}^{\otimes 2}}{\left\{\int_0^1 f_{X0}(x) f_U(W_i - x) dx\right\}^2} \\
& - \int_0^1 f_{X0}(x)\mathbf{B}_r(x)\mathbf{B}_r(x)^{\mathrm{T}} dx \\
& + \left\{\int_0^1 f_{X0}(x)\mathbf{B}_r(x) dx\right\}^{\otimes 2}\Bigg]\Bigg)^{-1} \mathbf{R}_{00}\{1 + o_p(1)\}.
\end{aligned}
$$

This proves the results. □

S.5. Proof of Proposition 3. First note that by the definition of $\mathbf{a}(\mathbf{X}, m)$, we have

$$E[\mathbf{S}^*_{\boldsymbol\beta}(W_i, Y_i, m) - E^*\{\mathbf{a}(X, m) \mid Y_i, W_i, m\} \mid \mathbf{X}, m] = \mathbf{0}.$$

This leads to

$$E[\mathbf{S}^*_{\boldsymbol\beta}(W_i, Y_i, \boldsymbol\beta_0) - E^*\{\mathbf{a}(X, \boldsymbol\beta_0) \mid Y_i, W_i, \boldsymbol\beta_0\}] = o_p(1) \tag{S.19}$$

element-wise by Condition (D4). By Condition (D5), $E[\mathbf{S}^*_{\boldsymbol\beta}(W_i, Y_i, \boldsymbol\beta) - E^*\{\mathbf{a}(X, \boldsymbol\beta) \mid Y_i, W_i, \boldsymbol\beta\}]$ is invertible near $\boldsymbol\beta^*$ and its inverse function is a one-to-one function with bounded first derivative in the neighborhood of zero. Therefore $\|\boldsymbol\beta^* - \boldsymbol\beta_0\|_2 = o_p(1)$. On the other hand, since

$$n^{-1}\sum_{i=1}^n \mathbf{S}^*_{\boldsymbol\beta}(W_i, Y_i, \widehat{\boldsymbol\beta}) - E^*\{\mathbf{a}(X, \widehat{\boldsymbol\beta}) \mid Y_i, W_i, \widehat{\boldsymbol\beta}\} = \mathbf{0}, \tag{S.20}$$

we obtain

$$E[\mathbf{S}^*_{\boldsymbol\beta}(W_i, Y_i, \widehat{\boldsymbol\beta}) - E^*\{\mathbf{a}(X, \widehat{\boldsymbol\beta}) \mid Y_i, W_i, \widehat{\boldsymbol\beta}\}] = o_p(1).$$

Using the same argument regarding $\boldsymbol\beta_0$ in (S.19), we obtain $\|\widehat{\boldsymbol\beta} - \boldsymbol\beta^*\|_2 = o_p(1)$. Thus, $\|\widehat{\boldsymbol\beta} - \boldsymbol\beta_0\|_2 = o_p(1)$. □

S.6. Proof of Proposition 4. By the Taylor expansion, we have

$$\mathbf{0} \quad = \quad \sum_{i=1}^n \mathbf{S}^*_{\boldsymbol\beta}(W_i, Y_i, \widehat{\boldsymbol\beta}) - E^*\{\mathbf{a}(X, \widehat{\boldsymbol\beta}) \mid Y_i, W_i, \widehat{\boldsymbol\beta}\}$$

$$= \mathbf{T}_0 + \mathbf{T}_1(\boldsymbol{\beta}^*)(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta}_0), \tag{S.21}$$

where $\boldsymbol{\beta}^*$ is a point on the line connecting $\boldsymbol{\beta}_0$ and $\widehat{\boldsymbol{\beta}}$, and for any $\boldsymbol{\beta}$

$$\mathbf{T}_1(\boldsymbol{\beta}) = n^{-1}\sum_{i=1}^{n}\frac{\partial[\mathbf{S}^*_{\boldsymbol{\beta}}(W_i, Y_i, \boldsymbol{\beta}) - E^*\{\mathbf{a}(X, \boldsymbol{\beta}) \mid Y_i, W_i, \boldsymbol{\beta}\}]}{\partial\boldsymbol{\beta}^{\mathrm{T}}},$$

and

$$\begin{aligned}\mathbf{T}_0 &= n^{-1}\sum_{i=1}^{n}\mathbf{S}^*_{\boldsymbol{\beta}}(W_i, Y_i, \boldsymbol{\beta}_0) - E^*\{\mathbf{a}(X, \boldsymbol{\beta}_0) \mid Y_i, W_i, \boldsymbol{\beta}_0\}\\ &= \mathbf{T}_{00} + \mathbf{T}_{01},\end{aligned}$$

where

$$\mathbf{T}_{00} = n^{-1}\sum_{i=1}^{n}\mathbf{S}^*_{\boldsymbol{\beta}}(W_i, Y_i, m) - E^*\{\mathbf{a}(X, m) \mid Y_i, W_i, m\},$$

and

$$\begin{aligned}\mathbf{T}_{01} &= n^{-1}\sum_{i=1}^{n}\mathbf{S}^*_{\boldsymbol{\beta}}(W_i, Y_i, \boldsymbol{\beta}_0) - E^*\{\mathbf{a}(X, \boldsymbol{\beta}_0) \mid Y_i, W_i, \boldsymbol{\beta}_0\}\\ &\quad - \left[n^{-1}\sum_{i=1}^{n}\mathbf{S}^*_{\boldsymbol{\beta}}(W_i, Y_i, m) - E^*\{\mathbf{a}(X, m) \mid Y_i, W_i, m\}\right].\end{aligned}$$

By the definition of $\mathbf{S}^*_{\boldsymbol{\beta}}(W_i, Y_i, m)$, we can write

$$\begin{aligned}\|\mathbf{T}_{00}\|_\infty &= \sup_k\left|n^{-1}\sum_{i=1}^{n}\int_0^1\left[-\frac{f'_\epsilon\{Y_i - m(x)\}f_U(W_i - x)f^*_X(x)}{\int_0^1 f_\epsilon\{Y_i - m(x)\}f_U(W_i - x)f^*_X(x)d\mu(x)}\right.\right.\\ &\quad \left.\left.-P(x, W_i, Y_i, m)\right]B_{rk}(x)d\mu(x)\right|. \end{aligned}\tag{S.22}$$

Now define

$$C_{1i}(x) = \left[-\frac{f'_\epsilon\{Y_i - m(x)\}f_U(W_i - x)f^*_X(x)}{\int_0^1 f_\epsilon\{Y_i - m(x)\}f_U(W_i - x)f^*_X(x)d\mu(x)} - P(x, W_i, Y_i, m)\right],$$

by the Bernstein's inequality in Bosq (1998), we have

$$\begin{aligned}&\left|1/n\sum_{i=1}^{n}\int_0^1 B_{rk}(x)C_{1i}(x)dx - E\left\{\int_0^1 B_{rk}(x)C_{1i}(x)dx\right\}\right|\\ &= O_p\left(\left[\sum_{i=1}^{n}E\left\{1/n\int_0^1 B_{rk}(x)C_{1i}(x)dx\right\}^2 \log n\right]^{1/2}\right)\end{aligned}$$

$$= O_p\{\sqrt{h_b n^{-1}\log(n)}\}. \tag{S.23}$$

The last equality holds from Lemma 2 in the supplementary material, by choosing $g_i(x) = C_{1i}(x)$ and setting $\mathbf{u}$ with $u_k = 1$ and $u_l = 0$, for $l \neq k$. Now

$$\begin{aligned}
& E[\mathbf{S}^*_{\boldsymbol{\beta}}(W_i, Y_i, m) - E^*\{\mathbf{a}(X, m) \mid Y_i, W_i, m\}] \\
= \ & E\left(\int_0^1 \left[-\frac{f'_\epsilon\{Y - m(x)\}f_U(W - x)f^*_X(x)}{\int_0^1 f_\epsilon\{Y - m(x)\}f_U(W - x)f^*_X(x)dx} - P(x, W, Y, m)\right]\mathbf{B}_r(x)dx\right) \\
= \ & \mathbf{0},
\end{aligned}$$

which further implies $E\left\{\int_0^1 B_{rk}(x)C_{1i}(x)dx\right\} = 0$ for all k. Plug it into (S.23), and combine with (S.22), we obtain

$$\|\mathbf{T}_{00}\|_\infty = O_p\{\sqrt{h_b n^{-1}\log(n)}\}. \tag{S.24}$$

Further,

$$\begin{aligned}
& \|\mathbf{T}_{01}\|_\infty \\
= \ & \sup_k \Bigg| n^{-1}\sum_{i=1}^n \int_0^1 \Bigg(\Bigg[-\frac{f'_\epsilon\{Y_i - \mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\beta}_0\}f_U(W_i - x)f^*_X(x)}{\int_0^1 f_\epsilon\{Y_i - \mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\beta}_0\}f_U(W_i - x)f^*_X(x)d\mu(x)} \\
& -P(x, W_i, Y_i, \boldsymbol{\beta}_0)\Bigg] - \Bigg[-\frac{f'_\epsilon\{Y_i - m(x)\}f_U(W_i - x)f^*_X(x)}{\int_0^1 f_\epsilon\{Y_i - m(x)\}f_U(W_i - x)f^*_X(x)d\mu(x)} \\
& -P(x, W_i, Y_i, m)\Bigg]\Bigg) B_{rk}(x)d\mu(x)\Bigg| \\
= \ & \sup_k \left| n^{-1}\sum_{i=1}^n \frac{\partial S_m(Y_i, W_i, m)}{\partial m}(B_{rk}, \mathbf{B}_r^{\mathrm{T}}\boldsymbol{\beta} - m)\right|\{1 + o_p(1)\} \\
= \ & O_p(1)\sup_k\left\{\int_0^1 B_{rk}(x)dx\right\}\left\{\sup_{x\in[0,1]}|\mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\beta} - m(x)|\right\}\{1 + o_p(1)\} \\
= \ & O_p(h_b^{q+1}),
\end{aligned} \tag{S.25}$$

The second equality holds by the Taylor expansion with respect to $\mathbf{B}_r(\cdot)^{\mathrm{T}}\boldsymbol{\beta}$. The third equality holds by Condition (D2). The last equality holds by Condition (D4) and the fact that the support of B_{rk} is the interval (t_{k-r}, t_k). Combining (S.24) and (S.25) we have $\mathbf{T}_0 = \mathbf{T}_{00}\{1 + o_p(1)\}$ by Conditions (C3) and (C4). Further,

$$E(\mathbf{T}_{00}^{\mathrm{T}}\mathbf{T}_{00})$$

$$
\begin{aligned}
&= n^{-1}E\left\{\left(\sum_{l=1}^{d_{\boldsymbol{\beta}}}\int_0^1\left[-\frac{f'_\epsilon\{Y_i-m(x)\}f_U(W_i-x)f_X^*(x)}{\int_0^1 f_\epsilon\{Y_i-m(x)\}f_U(W_i-x)f_X^*(x)d\mu(x)}\right.\right.\right.\\
&\quad \left.\left.\left. -P(x,W_i,Y_i,m)\right]B_{rl}(x)dx\right)^2\right\}\\
&\le n^{-1}G_1\sum_{l=1}^{d_\beta}\left\{\int_0^1 B_{rl}(x)dx\right\}^2\\
&= O_p(n^{-1}h_b),
\end{aligned}
$$

for some constant G_1. The second equality holds by Condition (D2) and the fact that,

$$
\begin{aligned}
&\sum_{l=1}^{d_{\boldsymbol{\beta}}}\left\{\int_0^1 B_{rl}(x)dx\right\}^2\\
&= \sum_{l=1}^{d_{\boldsymbol{\beta}}} D_{1l}(t_l-t_{l-r})^2\\
&\le \sup_l |t_l-t_{l-r}|D_1\\
&= O_p(h_b),
\end{aligned}
$$

for some finite constant D_{1l} and D_1. The second equality holds because the support of B_{rl} is the interval (t_{l-r},t_l). The last equality holds because D_{1l} are bounded above since $B_{rl}(x)<1$ for any x. Therefore,

$$
\|\mathbf{T}_0\|_2 = O_p(n^{-1/2}h_b^{1/2}). \tag{S.26}
$$

Further for arbitrary vector $\mathbf{u}$ with $\|\mathbf{u}\|_2=1$, we can write $\mathbf{T}_1(\boldsymbol{\beta}^*)$ as

$$
\begin{aligned}
|\mathbf{u}^{\mathrm{T}}\mathbf{T}_1(\boldsymbol{\beta}^*)\mathbf{u}| &= \left|n^{-1}\sum_{i=1}^n\frac{\partial S_m(Y_i,W_i,m^*)}{\partial m^*}\left(\sum_{k=1}^{d_{\boldsymbol{\beta}}}u_kB_{rk},\sum_{k=1}^{d_{\boldsymbol{\beta}}}u_kB_{rk}\right)\right|\\
&= \{c+o_p(1)\}\|\sum_{k=1}^{d_{\boldsymbol{\beta}}}u_kB_{rk}\|_2^2
\end{aligned}
$$

for a positive constant c, where $m^*(\cdot)=\mathbf{B}_r(\cdot)^{\mathrm{T}}\boldsymbol{\beta}^*$. The last equality holds by Condition (D2) and the fact that $n^{-1}\sum_{i=1}^n\partial S_m(Y_i,W_i,m^*)/\partial m^*$ is a non-singular operator on $\mathcal{B}$. To show the non-singularity, we prove by contradiction. Recall that

$$
S_m(Y,W,m^*)(s)
$$

$$= \int_0^1 \left[-\frac{f'_\epsilon\{Y - m^*(x)\} f_U(W-x) f_X^*(x) d\mu(x)}{\int_0^1 f_\epsilon\{Y - m^*(x)\} f_U(W-x) f_X^*(x) d\mu(x)} - P(x, Y, W, m^*) \right] s(x) dx.$$

and

$$\begin{aligned} & \mathbf{S}^*_{\boldsymbol\beta}(W_i, Y_i, \boldsymbol\beta^*) - E^*\{\mathbf{a}(X, \boldsymbol\beta^*)|Y_i, W_i, \boldsymbol\beta^*\} \\ = & \int_0^1 \mathbf{B}_r(x) \left[-\frac{f'_\epsilon\{Y - m^*(x)\} f_U(W-x) f_X^*(x) d\mu(x)}{\int_0^1 f_\epsilon\{Y - m^*(x)\} f_U(W-x) f_X^*(x) d\mu(x)} - P(x, Y, W, m^*) \right] dx. \end{aligned}$$

Suppose $E\{\partial S_m(Y_i, W_i, m^*)/\partial m^*\}$ is singular on $\mathcal{B}$, then

$$\begin{aligned} & E\left\{ \frac{\partial \mathbf{S}^*_{\boldsymbol\beta}(W_i, Y_i, \boldsymbol\beta^*) - E^*\{\mathbf{a}(X, \boldsymbol\beta^*)|Y_i, W_i, \boldsymbol\beta^*\}}{\partial \boldsymbol\beta^{*\mathrm{T}}} \right\} \\ = & E\left\{ \int_0^1 \mathbf{B}_r(x) \frac{\partial}{\partial m^*(x)} \left[-\frac{f'_\epsilon\{Y - m^*(x)\} f_U(W-x) f_X^*(x) d\mu(x)}{\int_0^1 f_\epsilon\{Y - m^*(x)\} f_U(W-x) f_X^*(x) d\mu(x)} - P(x, Y, W, m^*) \right] \right. \\ & \left. \mathbf{B}_r(x)^{\mathrm{T}} dx \right\} \end{aligned}$$

is a singular matrix. However by Condition (D5) we know that the score function

$$E[\mathbf{S}^*_{\boldsymbol\beta}(W_i, Y_i, \boldsymbol\beta) - E^*\{\mathbf{a}(X, \boldsymbol\beta)|Y_i, W_i, \boldsymbol\beta\}]$$

has a unique solution around $\boldsymbol\beta_0$, and hence

$$E\left\{ \frac{\partial \mathbf{S}^*_{\boldsymbol\beta}(W_i, Y_i, \boldsymbol\beta^*) - E^*\{\mathbf{a}(X, \boldsymbol\beta^*)|Y_i, W_i, \boldsymbol\beta^*\}}{\partial \boldsymbol\beta^{*\mathrm{T}}} \right\}$$

cannot be singular. Therefore, $E\{\partial S_m(Y_i, W_i, m^*)/\partial m\}$ is indeed a non-singular operator on $\mathcal{B}$.. Further,
$n^{-1}\sum_{i=1}^n \{\partial S_m(Y_i, W_i, m^*)/\partial m^*\}$ is a consistent estimator of $E\{\partial S_m(Y_i, W_i, m^*)/\partial m^*\}$, and hence
$n^{-1}\sum_{i=1}^n \{\partial S_m(Y_i, W_i, m^*)/\partial m^*\}$ is also a non-singular operator on $\mathcal{B}$ for sufficiently large n. Further by using Lemme 1 in the supplementary material, we have $g_3 h_b^{1/2} = D_r \|\mathbf{u}'\|_2 \le \|\sum_{k=1}^{d_{\boldsymbol\beta}} u_k B_{rk}\|_2 \le \|\mathbf{u}'\|_2 = G_3 h_b^{1/2}$, where $\mathbf{u}' = \{u_k\{(t_k - t_{k-r})/r\}^{1/2}, k = 1, \ldots, d_{\boldsymbol\beta}\}^{\mathrm{T}}$ and g_3, G_3 are finite constants. These together with Lemma 2 imply

$$c_9 h_b \le \mathbf{u}^{\mathrm{T}} \mathbf{T}_1(\boldsymbol\beta^*) \mathbf{u} \le C_9 h_b, \tag{S.27}$$

and

$$C_9^{-1} h_b^{-1} \le \mathbf{u}^{\mathrm{T}} \mathbf{T}_1(\boldsymbol\beta^*)^{-1} \mathbf{u} \le c_9^{-1} h_b^{-1}, \tag{S.28}$$

in probability. Combining with (S.21), (S.26), we have

$$\|\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0\|_2 \le \|\mathbf{T}_1(\boldsymbol{\beta}^*)^{-1}\|_2\|\mathbf{T}_0\|_2 = O_p\{(nh_b)^{-1/2}\}.$$

By the consistency of $\mathbf{B}_r(x)^{\mathrm{T}}\boldsymbol{\beta}^*$ to $m(x)$, we have

$$\begin{aligned}
& \mathbf{T}_1(\boldsymbol{\beta}^*) \\
= \ & E\left(\left.\frac{\partial[\mathbf{S}^*_{\boldsymbol{\beta}}(W_i,Y_i,\boldsymbol{\beta})-E^*\{\mathbf{a}(X,\boldsymbol{\beta})\mid Y_i,W_i,\boldsymbol{\beta}\}]}{\partial\boldsymbol{\beta}^{\mathrm{T}}}\right|_{\mathbf{B}_r(\cdot)^{\mathrm{T}}\boldsymbol{\beta}=m(\cdot)}\right)\{1+o_p(1)\}.
\end{aligned}$$

Therefore, we can write

$$\begin{aligned}
& \widehat{\boldsymbol{\beta}}-\boldsymbol{\beta}_0 \\
= \ & -E\left(\left.\frac{\partial[\mathbf{S}^*_{\boldsymbol{\beta}}(W_i,Y_i,\boldsymbol{\beta})-E^*\{\mathbf{a}(X,\boldsymbol{\beta})\mid Y_i,W_i,\boldsymbol{\beta}\}]}{\partial\boldsymbol{\beta}^{\mathrm{T}}}\right|_{\mathbf{B}_r(\cdot)^{\mathrm{T}}\boldsymbol{\beta}=m(\cdot)}\right)^{-1} \\
& \times\mathbf{T}_{00}\{1+o_p(1)\}.
\end{aligned}$$

This proves the results. □

References.

Bosq, D. (1998). *Nonparametric statistics for stochastic processes : estimation and prediction D. Bosq. Lecture notes in statistics: 110.* Springer, New York.

Cardot, H., Ferraty, F. and Sarda, P. (1999). Functional linear model. *Statistics & Probability Letters* **45** 11–22.

Cardot, H., Ferraty, F. and Sarda, P. (2003). Spline estimators for the functional linear model. *Statistica Sinica* **13** 571–592.

Chatelin, F. (1983). *Spectral Approximation of Linear Operators.* SIAM, New York.

De Boor, C. (1978). *A practical guide to splines* **27**. Springer-Verlag, New York.

DeVore, R. A. and Lorentz, G. G. (1993). *Constructive approximation* **303**. Springer Science & Business Media, New York.

Department of Statistics
University of Hongkong
Pok Fu Lam, Hongkong
E-mail: feijiang@hku.hk

Department of Statistics
Penn State University
University Park, PA 16802
E-mail: yzm63@psu.edu

Fig S.1: B-spline MLE density estimation (left) and deconvolution estimation (right) from 200 simulations: The solid lines represent the true functions and the dash lines represent the estimated functions and their 90% confidence bands. The first row to third row are the results for model II (a)–(c) respectively. Sample size 1000.

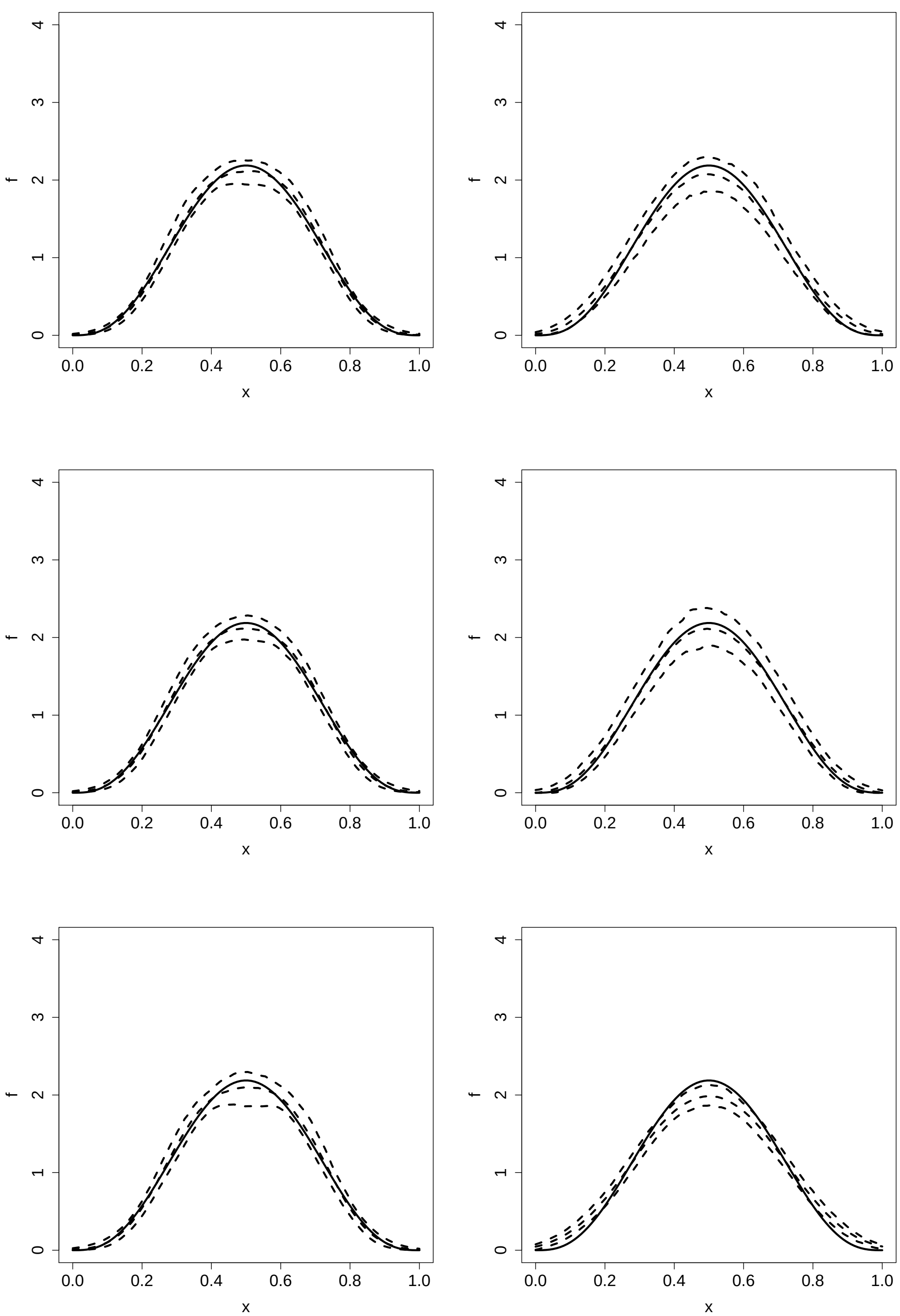


Fig S.2: B-spline MLE density estimation (left) and deconvolution estimation (right) from 200 simulations: The solid lines represent the true functions and the dash lines represent the estimated functions and their 90% confidence bands. The first row to third row are the results for model II (a)–(c) respectively. Sample size 2000.

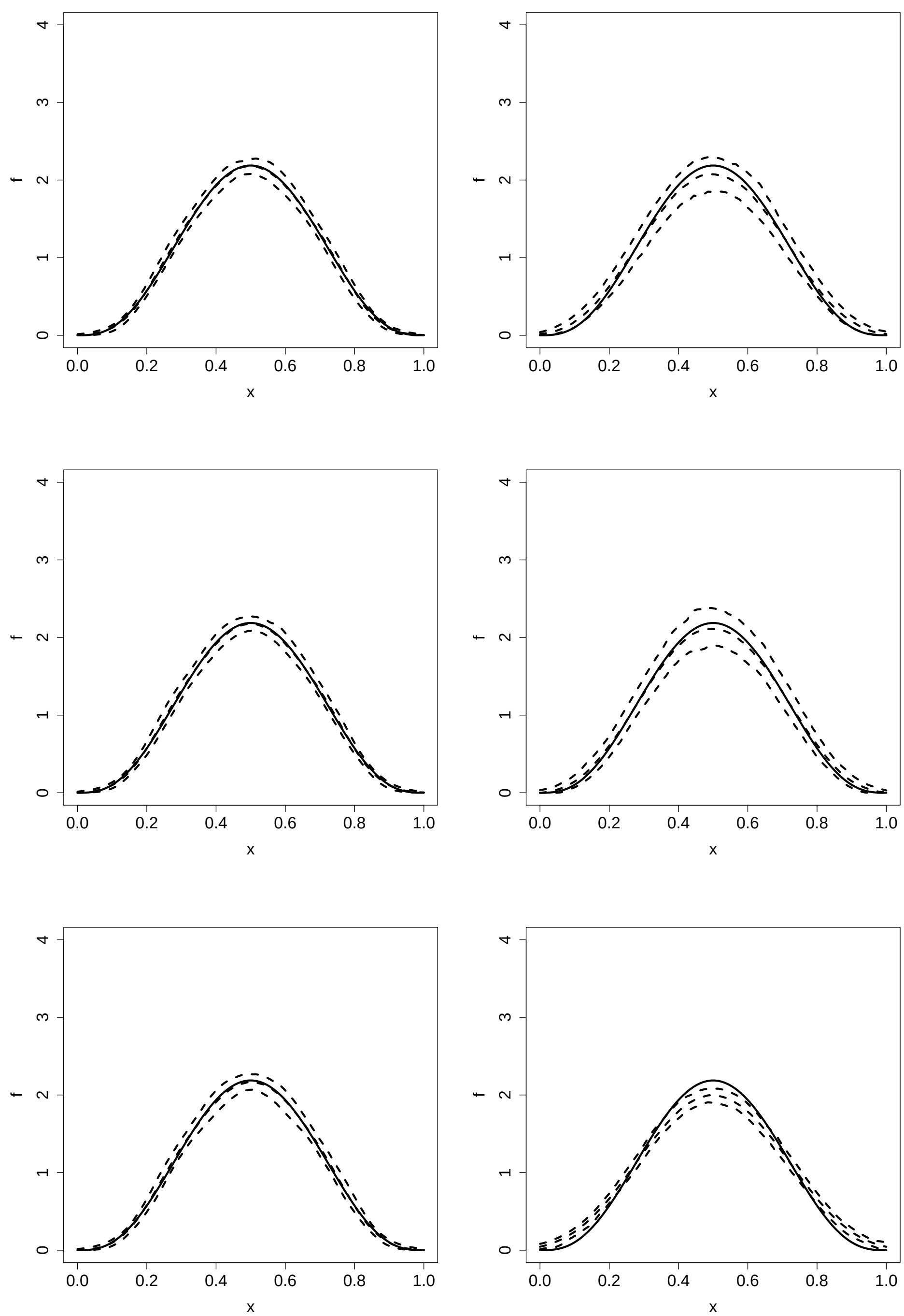


Fig S.3: B-spline semiparametric regression estimation (left) and deconvolution estimation (right) from 200 simulations: The solid lines represent the true functions and the dash lines represent the estimated functions and their 90% confidence bands. The first row to third row are the results for model II (a)–(c), respectively. Sample size 1000.

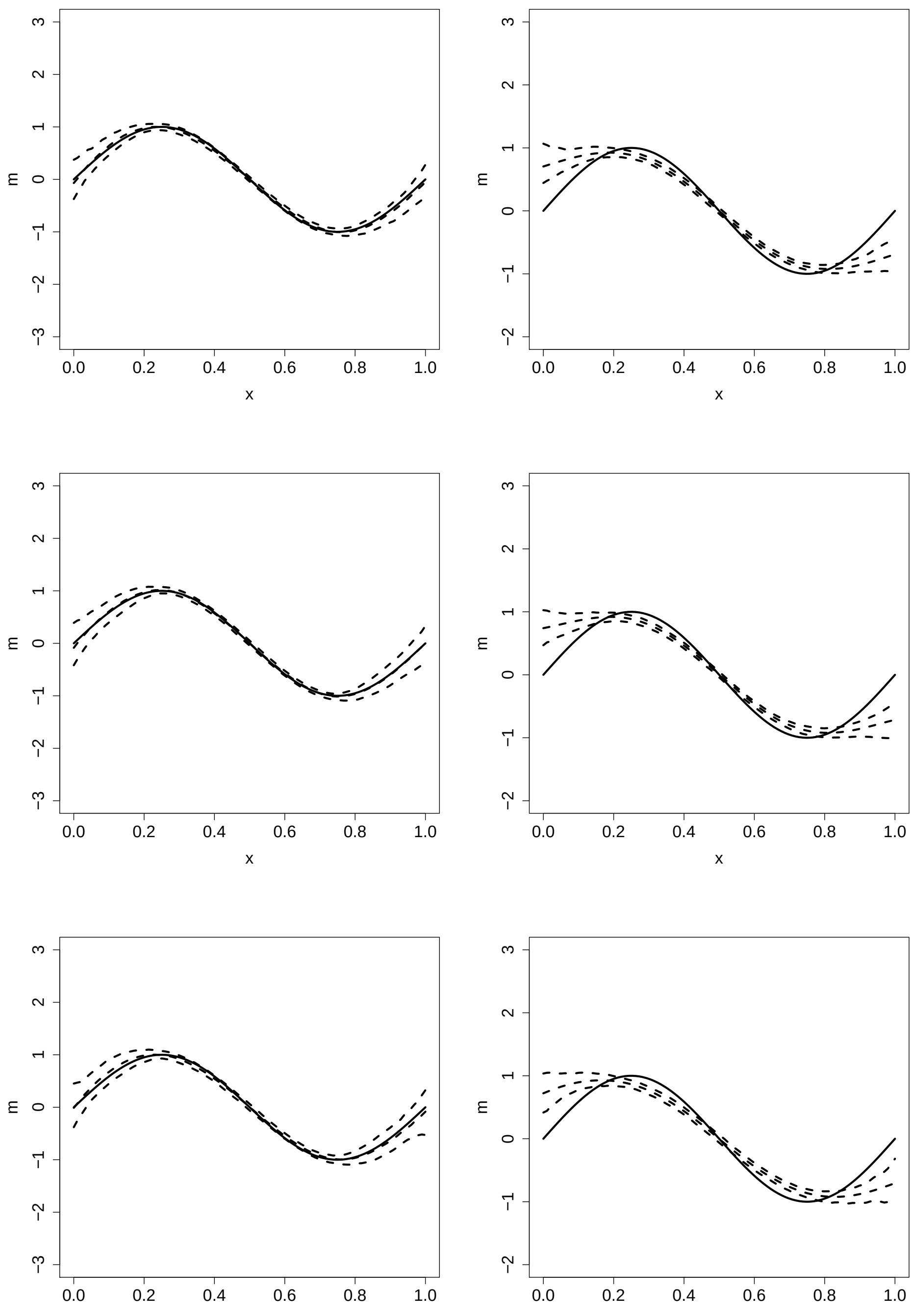


Fig S.4: B-spline semiparametric regression estimation (left) and deconvolution estimation (right) from 200 simulations: The solid lines represent the true functions and the dash lines represent the estimated functions and their 90% confidence bands. The first row to third row are the results for model II (a)–(c), respectively. Sample size 2000.