Nonparametric regression for a circular response with error-in-covariate

Nicholas Woolsey[1] , **Xianzheng Huang**[2]

[1]*Statistics, University of South Carolina, e-mail:* nwoolsey@email.sc.edu

[2]*Statistics, University of South Carolina, e-mail:* huang@stat.sc.edu

Abstract: This study considers regression analysis of a circular response with an error-prone linear covariate. Starting with an existing estimator of the circular regression function that assumes error-free covariate, three approaches are proposed to revise this estimator, leading to three nonparametric estimators for the circular regression function accounting for measurement error. The proposed estimators are intrinsically connected through some deconvoluting operator that is exploited differently in different estimators. Moreover, a new bandwidth selection method is developed that is more computationally efficient than an existing method well-received in the context of tuning parameter selection in the presence of measurement error. The efficacy of these new estimators and their relative strengths are demonstrated through a thorough investigation of their asymptotic properties and extensive empirical study of their finite-sample performance.

Contents

1. Introduction

Circular data are data of a periodic nature that arise in a host of applications. Examples of circular data include data on the direction of animal movements that are of interest in zoology and ecology [2, 35], and wind direction data as a subject of investigation in meteorology [24]. Times of certain events, such as hospital visits and crimes, as observed on the 24-hour clock can be transformed to circular data, so are dates of certain events on a calendar, such as births and migration. In these applications, researchers are often interested in studying the association between a circular response and linear covariates. For instance, Birkett et al. [3] considered regressing elephant's movements on rainfall; Hodel and Fieberg [21] analyzed the association between a moving animal's turn-angle and step-length; García-Portugués et al. [18] formulated a regression model for wind direction and SO_2 concentration; Gill and Hangartner

[19] developed a circular regression model for the timing of domestic terrorism events, with linear covariates such as the number of attackers and death count.

Early developments of circular-linear regression analysis led to parametric models that usually assume a von Mises distribution [28] for the circular response given covariates. Frequently referenced regression models of this type include those proposed by Gould [20], Johnson and Wehrly [23], and Fisher and Lee [16]. More recent parametric models that adopt different distributional assumptions on the circular response include those considered in [43], [36], and [33], among many others. A complication in formulating such regression models is the choice of a link function that relates predictors in the Euclidean space to the mean direction in a circular space. There is also the concern that the actual circular response may violate parametric assumptions implied by the chosen circular distribution. Nonparametric circular-linear regression models have been proposed to avoid choosing a link function or imposing stringent assumptions on the circular response. One of the first notable developments along this line is the local polynomial regression of a circular response introduced by Di Marzio et al. [12]. Meilán-Vila et al. [29] generalized this work by allowing a multivariate linear covariate.

Oftentimes covariates of scientific interest cannot be measured precisely, either due to human error or imprecise measuring instruments. For example, rainfall measurements from a rain gauge are affected by the exposure conditions of the rain gauge; the measurement accuracy of SO_2 concentration is also highly device-dependent. This motivates our study of circular regression models with a linear covariate prone to measurement error. The topic of regression analysis with error-prone covariates has been studied extensively. Comprehensive reviews of existing methodolgies for dealing with error-in-covariate can be found in [5], [17], [4], [46], and [45]. Despite the existing extensive research on regression models with error-in-covariate, parallel problems in the context of regression models for circular responses have received little attention. Our study contributes in this underinvestigated area. Concurrently, [11] considered regression models for a response that can be circular or linear, with error-prone covariates that also can be circular or linear.

Focusing on circular-linear regression models, we propose in this study three methods based on local linear estimation adapted for a circular response while accounting for covariate measurement error. To prepare for the methodology development, Section 2 briefly reviews local polynomial regression of a circular response with error-free covariates proposed by [12]. In Section 3 we develop three strategies to revise their method to account for measurement error in the linear covariate. Section 4 summarizes asymptotic properties of our proposed estimators. Bandwidth selection required for local polynomial regression is especially challenging when covariates are prone to measurement error, which is an important issue not addressed in [11]. We present in Section 5 two bandwidth selection methods utilized in our proposed estimation methods. Section 6 reports simulation studies where the proposed estimators' finite-sample performance is inspected and compared with a naive estimator and a benchmark estimator. We implement these methods in a real-life application in Section 7. Lastly, Section 8 summarizes contributions of our study and discuss extensions of the study to different relevant settings.

2. Local polynomial circular-linear regression

2.1. The circular mean function

Consider n independent bivariate observations, $\{(\Theta_j, X_j)\}_{j=1}^n$, identically distributed according to the joint distribution of (Θ, X) supported on $[-\pi, \pi) \times \mathbb{R}$. Given the linear covariate $X = x$, the distribution of the circular response Θ is specified by

$$\Theta = \{m(x) + \epsilon\} (\text{mod } 2\pi), \tag{1}$$

where ϵ is an angular error following an unspecified distribution with a circular mean of zero, and $m(x)$ is the circular mean of Θ, i.e., the circular regression function that maps from a linear space to a circular space. The circular mean of a circular random variable Θ is defined as the angle $\alpha \in [-\pi, \pi)$ that minimizes the mean cosine dissimilarity between Θ and α, $D(\alpha) = \mathbb{E}\{1 - \cos(\Theta - \alpha)\}$, which can be shown to be $\arg \min D(\alpha) = \text{atan2}[\mathbb{E}(\sin \Theta), \mathbb{E}(\cos \Theta)]$ [28, Chapter 2]. Here, $\text{atan2}[y, x]$ is the angle between the x-axis and the vector from the origin to the point (x, y) in the Cartesian coordinate system.

Because $\text{atan2}[y, x]$ depends on (x, y) only via the ratio y/x, one may infer the circular mean function $m(x)$ via inferring $m_1(x)/m_2(x)$, where $m_1(x) = \mathbb{E}(\sin \Theta | X = x)$ and $m_2(x) = \mathbb{E}(\cos \Theta | X = x)$, both functions mapping to a linear space as in the traditional regression setting.

2.2. Local polynomial estimators

Given the response data $\mathbf{\Theta} = (\Theta_1, \ldots, \Theta_n)^\top$ and covariate data $X = (X_1, \ldots, X_n)^\top$, [12] proposed kernel-based estimators for $m(x)$ in the form of

$$\tilde{m}(x) = \text{atan2}[\tilde{g}_1(x), \tilde{g}_2(x)], \tag{2}$$

where

$$\tilde{g}_1(x) = \frac{1}{n} \sum_{j=1}^n \sin(\Theta_j) \mathcal{W}(X_j - x), \quad \tilde{g}_2(x) = \frac{1}{n} \sum_{j=1}^n \cos(\Theta_j) \mathcal{W}(X_j - x), \tag{3}$$

in which $\mathcal{W}(t)$ is a weight function. For example, if $\mathcal{W}(t) = K_h(t)$, where $K_h(t) = K(t/h)/h$, $K(t)$ is a kernel function, and h is the bandwidth, then

$$\tilde{g}_1(x) = \frac{1}{n} \sum_{j=1}^n \sin(\Theta_j) K_h(X_j - x)$$

$$= \frac{\sum_{j=1}^n \sin(\Theta_j) K_h(X_j - x)}{\sum_{j=1}^n K_h(X_j - x)} \times \frac{1}{n} \sum_{j=1}^n K_h(X_j - x).$$

Now one can see that the first factor in the preceding expression is a local constant estimator [30] for $m_1(x)$, and the second factor is a traditional kernel density estimator for the probability density function of X evaluated at x, denoted by $f_X(x)$. Hence, with $\mathcal{W}(t) = K_h(t)$, $\tilde{g}_1(x)$ is a sensible estimator for $g_1(x) = m_1(x) f_X(x)$; similarly, $\tilde{g}_2(x)$ is an estimator for $g_2(x) = m_2(x) f_X(x)$.

The estimator $\tilde{m}(x)$ in (2) acknowledges that the traditional local polynomial estimation developed for a linear response is inadequate for a circular response, but remains applicable for estimating $g_1(x)$ and $g_2(x)$, both functions mapping from $\mathbb{R}$ onto another linear space. This allows estimation of the ratio $m_1(x)/m_2(x) = g_1(x)/g_2(x)$ based on two rounds of local polynomial estimation. Besides the local constant weight, [12] also considered using the following weight function that depends on all covariate data X,

$$\mathcal{W}(t) = K_h(t) \times \frac{1}{n} \sum_{k=1}^{n} \left(\frac{X_k - x}{h}\right)^2 K_h(X_k - x) - \frac{t}{h} K_h(t) \times \frac{1}{n} \sum_{k=1}^{n} \left(\frac{X_k - x}{h}\right) K_h(X_k - x). \tag{4}$$

This more complicated weight function relates to local linear estimators for the regression function of a linear response Y, $n^{-1} \sum_{j=1}^{n} Y_j \mathcal{W}^*(X_j - x)$, with the weight given by [15]

$$\begin{aligned}
&\mathcal{W}^*(X_j - x) \\
&= \mathcal{W}(X_j - x) \left[\left\{\frac{1}{n} \sum_{k=1}^{n} K_h(X_k - x)\right\} \left\{\frac{1}{n} \sum_{k=1}^{n} \left(\frac{X_k - x}{h}\right)^2 K_h(X_k - x)\right\} - \right. \\
&\left. \left\{\frac{1}{n} \sum_{k=1}^{n} \left(\frac{X_k - x}{h}\right) K_h(X_k - x)\right\}^2 \right]^{-1},
\end{aligned} \tag{5}$$

where $\mathcal{W}(X_j - x)$ is equal to (4) evaluated at $t = X_j - x$. We thus call $\mathcal{W}(X_j - x)$ the local linear weight, and refer to $\mathcal{W}^*(X_j - x)$ as the normalized local linear weight. Plugging in (3) the local linear weight, one has $\tilde{g}_\ell(x) = C(X)\tilde{m}_\ell(x)$, where $C(X)$ is the term inside the square brackets in (5), and $\tilde{m}_\ell(x)$ is a local linear estimator of $m_\ell(x)$, for $\ell = 1, 2$. Consequently, $\tilde{m}(x)$ in (2) is equal to atan2$[\tilde{m}_1(x), \tilde{m}_2(x)]$ since $\tilde{g}_1(x)/\tilde{g}_2(x) = \tilde{m}_1(x)/\tilde{m}_2(x)$. Under mild conditions that guarantee the consistency of a local linear estimator for the regression function of a linear response [15, Theorem 3.1] and the consistency of a kernel density estimator [37, Section 6.2], $\tilde{g}_\ell(x)$ consistently estimates $g_\ell(x) = m_\ell(x)f_X^2(x)\mu_2$, for $\ell = 1, 2$, where $\mu_2 = \int t^2 K(t) \, dt$. Unless otherwise stated, all integrals in this article are over the entire real line $\mathbb{R}$.

Following the above two examples of $\mathcal{W}(t)$, one can easily generalize the construction of $\tilde{m}(x)$ by adopting weight functions that relate to traditional local polynomial estimators of higher orders. For a more concrete exposition of methodology development and theoretical derivations, we focus on $\tilde{m}(x)$ with the local linear weight as a benchmark estimator that other estimators brought forth in the next section relate to and compare with.

3. Circular-linear regression with error-in-covariate

Suppose that, instead of X, error-contaminated covariate data $W = (W_1, \ldots, W_n)^\top$ are observed, where

$$W_j = X_j + U_j, \tag{6}$$

in which U_j is a mean-zero measurement error following a known distribution with variance σ_u^2, and U_j is independent of $(\Theta_j, X_j, \epsilon_j)$, for $j = 1, \ldots, n$. Had one ignored measurement

error, one would substitute X_j with W_j in (3) for $j = 1, \ldots, n$, leading to a naive estimator of $m(x)$, denoted by $\hat{m}^*(x)$. Because $\hat{m}^*(x)$ is merely a sensible estimator for the circular mean of Θ given $W = x$, denoted by $m^*(x)$, a function typically different from $m(x)$, so naive estimation of $m(x)$ using $\hat{m}^*(x)$ is usually misleading.

In what follows, we develop three strategies to modify $\hat{m}^*(x)$ in order to account for covariate measurement error when inferring $m(x)$. To contrast with the naive estimator $\hat{m}^*(x)$, we call $\tilde{m}(x)$ in (2) an ideal estimator of $m(x)$ to stress that $\tilde{m}(x)$ is only available in the "ideal" situation when $\sigma_u^2 = 0$ and thus $W = X$.

3.1. The deconvoluting kernel estimator

Since the naive substitution of X_j's with W_j's only occurs in the weight, we first propose to correct $\hat{m}^*(x)$ for measurement error by correcting the naive local linear weight,

$$
\mathcal{W}(W_j - x) = \frac{1}{n} K_h(W_j - x) \sum_{k=1}^{n} \left(\frac{W_k - x}{h} \right)^2 K_h(W_k - x) -
$$
$$
\frac{1}{n} \left(\frac{W_j - x}{h} \right) K_h(W_j - x) \sum_{k=1}^{n} \left(\frac{W_k - x}{h} \right) K_h(W_k - x).
$$

More specifically, we replace the above naive weight by

$$
\mathcal{L}(W_j - x) = \frac{1}{n} L_{0,h}(W_j - x) \sum_{k=1}^{n} \left(\frac{W_k - x}{h} \right)^2 L_{2,h}(W_k - x) -
$$
$$
\frac{1}{n} \left(\frac{W_j - x}{h} \right) L_{1,h}(W_j - x) \sum_{k=1}^{n} \left(\frac{W_k - x}{h} \right) L_{1,h}(W_k - x),
\tag{7}
$$

where, for $\ell = 0, 1, 2$, $L_{\ell,h}(x) = L_\ell(x/h)/h$, and

$$
L_\ell(x) = i^{-\ell} x^{-\ell} \frac{1}{2\pi} \int e^{-itx} \frac{\phi_K^{(\ell)}(t)}{\phi_U(-t/h)} dt,
\tag{8}
$$

in which $i = \sqrt{-1}$, $\phi_U(t)$ is the characteristic function of the measurement error U, $\phi_K(t)$ is the Fourier transform of the kernel, and $\phi_K^{(\ell)}(t) = (\partial^\ell / \partial t^\ell) \phi_K(t)$. In this paper, we use $\phi_A(t)$ to denote the characteristic function of A if A is a random variable, and to denote the Fourier transform of A if $A(\cdot)$ is a function. Setting $\ell = 0$ in (8) gives the deconvoluting kernel used in the deconvoluting kernel density estimator estimator for $f_X(x)$ based on W [41]. [8] generalized the deconvoluting kernel by introducing (8) for $\ell = 0, 1, 2, \ldots$, and used it to construct local polynomial estimators for the regression function of a linear response with error-in-covariate.

Under suitable conditions (to be stated in Section 4), [8] showed that

$$
\mathbb{E}\{(W - x)^\ell L_{\ell,h}(W - x)|X\} = (X - x)^\ell K_h(X - x), \text{ for } \ell = 0, 1, 2.
\tag{9}
$$

Because $\{(X - x)^\ell K_h(X - x), \ell = 0, 1, 2\}$ are building blocks of the local linear weight $\mathcal{W}(X_j - x)$, (9) sheds light on the rationale behind the new weight $\mathcal{L}(W_j - x)$ in (7): instead of naively using $\{(W - x)^\ell K_h(W - x), \ell = 0, 1, 2\}$ to substitute these building blocks, one uses $\{(W - x)^\ell L_{\ell,h}(W - x), \ell = 0, 1, 2\}$ as unbiased estimators of them. In this line of arguments,

as in (9), we treat X, or X_j's, as if they were unknown parameters, and we estimate functions of them using $\boldsymbol{W}$.

Using the new weight $\mathcal{L}(W_j - x)$ in (7), we have our first proposed estimator of $m(x)$ referred to as the deconvoluting kernel estimator, $\hat{m}_{\mathrm{DK}}(x) = \mathrm{atan2}[\hat{g}_{1,\mathrm{DK}}(x), \hat{g}_{2,\mathrm{DK}}(x)]$, where $\hat{g}_{1,\mathrm{DK}}(x) = n^{-1} \sum_{j=1}^{n} \sin(\Theta_j)\mathcal{L}(W_j - x)$ and $\hat{g}_{2,\mathrm{DK}}(w) = n^{-1} \sum_{j=1}^{n} \cos(\Theta_j)\mathcal{L}(W_j - x)$ are estimators for $g_1(x) = m_1(x)f_X^2(x)\mu_2$ and $g_2(x) = m_2(x)f_X^2(x)\mu_2$, respectively. Instead of correcting the naive local linear weight for measurement error as we do here, the estimator proposed by Di Marzio et al. [11, see Section 4.1.2] in the same context results from using (8) with $\ell = 0$ to correct the naive local constant weight.

3.2. The complex error estimator

Following a similar direction leading to $\hat{m}_{\mathrm{DK}}(x)$, we propose a second strategy where one corrects the naive normalized local linear weight, as the naive counterpart of $\mathcal{W}^*(X_j - x)$ in (5), for measurement error. This new strategy is motivated by the result in Lemma 3.1 given next, which we prove in Appendix A. Denote by $\mathbb{C}$ the set of complex numbers, by $\mathbb{N}$ the set of natural numbers, and define $\mathbb{N}_0 = \mathbb{N} \cup \{0\}$.

Lemma 3.1. *For an entire function $g : \mathbb{C}^n \to \mathbb{C}$, if $\mathbb{E}(\boldsymbol{V}^{\boldsymbol{\ell}}) = 0$, $\forall \boldsymbol{\ell} \in \mathbb{N}_0^n$ such that $|\boldsymbol{\ell}| > 0$, then $\mathbb{E}\{g(\boldsymbol{t} + \boldsymbol{V})\} = g(\boldsymbol{t})$, where $\boldsymbol{t} = (t_1, \dots, t_n)^\top \in \mathbb{R}^n$, and $\boldsymbol{V} = (V_1, \dots, V_n)^\top$ is an $n \times 1$ random vector. Here, for $\boldsymbol{\ell} = (\ell_1, \dots, \ell_n) \in \mathbb{N}_0^n$, $\boldsymbol{V}^{\boldsymbol{\ell}} = V_1^{\ell_1} \dots V_n^{\ell_n}$, and $|\boldsymbol{\ell}| = \ell_1 + \dots + \ell_n$.*

This result generalizes a similar result for a univariate entire function $g(\cdot)$ that was exploited in [40], [32], and [39] in measurement error problems. An entire function is a complex-valued function that is complex differentiable on the whole complex plane [26]. By Lemma 3.1, if measurement errors $\{U_j\}_{j=1}^n$ in (6) are independent normal errors, then $\mathbb{E}\{g(\boldsymbol{X} + \boldsymbol{U} + i\sigma_u\boldsymbol{Z})|\boldsymbol{X}\} = g(\boldsymbol{X})$, where $\boldsymbol{U} = (U_1, \dots, U_n)^\top$ and $\boldsymbol{Z} = (Z_1, \dots, Z_n)^\top$, in which $\{Z_j\}_{j=1}^n$ are independent standard normal errors that are also independent of $\boldsymbol{X}$ and $\boldsymbol{U}$. This is because, now with $\boldsymbol{V} = \boldsymbol{U} + i\sigma_u\boldsymbol{Z}$, $\mathbb{E}(\boldsymbol{V}^{\boldsymbol{\ell}}) = \prod_{j=1}^n \mathbb{E}\{(U_j + i\sigma_u Z_j)^{\ell_j}\} = 0$ for all $\boldsymbol{\ell} \in \mathbb{N}_0^n$ as long as $|\boldsymbol{\ell}| \neq 0$ in this case [40]. In conclusion, with normal measurement errors in (6), $g(\boldsymbol{W} + i\sigma_u\boldsymbol{Z})$ is an unbiased estimator of $g(\boldsymbol{X})$.

Now return to our goal of formulating a new weight using $\boldsymbol{W}$ that estimates $\mathcal{W}^*(X_j - x)$ unbiasedly. Since (5) depends on all true covariate data, we re-write $\mathcal{W}^*(X_j - x)$ as $\mathcal{W}^*(X_j - x; \boldsymbol{X})$, viewed as a function of $\boldsymbol{X}$, and we estimate this function in its entirety. In other words, we have

$$\mathcal{W}^*(t; \boldsymbol{X}) = \mathcal{W}(t) \left[\left\{ \frac{1}{n} \sum_{k=1}^n K_h(X_k - x) \right\} \left\{ \frac{1}{n} \sum_{k=1}^n \left(\frac{X_k - x}{h} \right)^2 K_h(X_k - x) \right\} - \right.$$
$$\left. \left\{ \frac{1}{n} \sum_{k=1}^n \left(\frac{X_k - x}{h} \right) K_h(X_k - x) \right\}^2 \right]^{-1},$$

where $\mathcal{W}(t)$ is given in (4), and t depends on $\boldsymbol{X}$ via one of its entry. Assuming normal measurement errors and that $\mathcal{W}^*(X_j - x; \boldsymbol{X})$ is an entire function, by Lemma 3.1, $\mathcal{W}^*(W_j^* - x; \boldsymbol{W}^*)$ is an unbiased estimator of $\mathcal{W}^*(X_j - x; \boldsymbol{X})$, where $W_j^* = W_j + i\sigma_u Z_j$, for $j = 1, \dots, n$, and $\boldsymbol{W}^* = (W_1^*, \dots, W_n^*)^\top$. Recognizing that both $\boldsymbol{W}$ and $\boldsymbol{Z}$ contribute to the variability of

this unbiased estimator, we employ an average of it across B realizations of $\mathbf{Z}$ to obtain a less variable unbiased estimator for $\mathcal{W}^*(X_j - x)$, leading to the new weight,

$$\mathcal{L}^*(W_j - x) = \frac{1}{B^*} \sum_{b=1}^{B^*} \mathcal{W}^*(W_{j,b}^* - x; \mathbf{W}_b^*), \tag{10}$$

where $W_{j,b}^* = W_j + i\sigma_u Z_{j,b}$, for $j = 1, \ldots, n$, $\mathbf{W}_b^* = (W_{1,b}^*, \ldots, W_{n,b}^*)^\top$, for $b = 1, \ldots, B$, and $\{Z_{j,b}, \ j = 1, \ldots, n\}_{b=1}^B$ are independent standard normal errors.

Using the new weight in (10) yields our second proposed estimator of $m(x)$ referred to as the complex error estimator, $\hat{m}_{\mathrm{CE}}(x) = \mathrm{atan2}[\hat{g}_{1,\mathrm{CE}}(x), \hat{g}_{2,\mathrm{CE}}(x)]$, where $\hat{g}_{1,\mathrm{CE}}(x) = n^{-1} \sum_{j=1}^n \sin(\Theta_j) \mathcal{L}^*(W_j - x)$ estimates $m_1(x)$, and $\hat{g}_{2,\mathrm{CE}}(w) = n^{-1} \sum_{j=1}^n \cos(\Theta_j) \mathcal{L}^*(W_j - x)$ estimate $m_2(x)$. In a separate study reported in Woolsey and Huang [44], we extended this idea to account for non-normal measurement error by introducing hypercomplex quantities.

3.3. The one-step correction estimator

The common thread running through the first two strategies of correcting $\hat{m}^*(x)$ for measurement error is to adopt some weight that considers measurement error. Each new weight leads to non-naive estimators for some functions $g_1(x)$ and $g_2(x)$ such that $g_1(x)/g_2(x) = m_1(x)/m_2(x)$. Deviating from this theme of weight correction, we now propose a third strategy that exploits a one-step correction of a naive estimator for $g_\ell(x)$ as a whole, for $\ell \in \{1, 2\}$. This one-step correction is motivated by Lemma 3.2 that we prove in Appendix B. A similar result was also utilized by [22] for local polynomial regression of a linear response with error-in-covariate.

Lemma 3.2. *Define $g_\ell(x) = m_\ell(x) f_X(x)$ and $g_\ell^*(x) = m_\ell^*(x) f_W(x)$, for $\ell = 1, 2$, where $m_1^*(x) = \mathbb{E}(\sin \Theta | W = x)$, $m_2^*(x) = \mathbb{E}(\cos \Theta | W = x)$, and $f_W(x)$ is the probability density of W evaluated at x. Then, assuming all integrals are well-defined,*

$$g_\ell(x) = \frac{1}{2\pi} \int e^{-itx} \frac{\phi_{g_\ell^*}(t)}{\phi_U(t)} \, dt, \text{ for } \ell = 1, 2.$$

Motivated by Lemma 3.2, we propose to estimate $g_\ell(x)$ by first obtaining a naive estimator of it that essentially estimates $g_\ell^*(x)$, then transforming the naive estimator using the integral transform suggested in Lemma 3.2. In this article, we use a local linear estimator denoted by $\hat{m}_\ell^*(x)$ to estimate $m_\ell^*(x)$, for $\ell = 1, 2$, and use the traditional kernel density estimator to estimate $f_W(x)$, denoted by $\hat{f}_W(x)$, leading to a naive estimator of $g_\ell(x)$ given by $\hat{g}_\ell^*(x) = \hat{m}_\ell^*(x) \hat{f}_W(x)$. Our third proposed estimator of $m(x)$ referred to as the one-step correction estimator is then given by $\hat{m}_{\mathrm{OS}}(x) = \mathrm{atan2}[\hat{g}_{1,\mathrm{OS}}(x), \hat{g}_{2,\mathrm{OS}}(x)]$, where

$$\hat{g}_{\ell,\mathrm{OS}}(x) = \frac{1}{2\pi} \int e^{-itx} \frac{\phi_{\hat{g}_\ell^*}(t)}{\phi_U(t)} \, dt \tag{11}$$

is an estimator of $g_\ell(x) = m_\ell(x) f_X(x)$, for $\ell = 1, 2$. Henceforth, we use $\mathcal{T}_U(A)(x)$ to refer to the integral transform in Lemma 3.2 of a function $A : \mathbb{R} \to \mathbb{R}$, i.e.,

$$\mathcal{T}_U(A)(x) = \frac{1}{2\pi} \int e^{-itx} \frac{\phi_A(t)}{\phi_U(t)} dt. \tag{12}$$

Then (11) is equivalent to $\hat{g}_{\ell,\mathrm{OS}}(x) = \mathcal{T}_U(\hat{g}_\ell^*)(x)$, for $\ell = 1, 2$.

3.4. Comparisons between proposed estimators

With three estimators of the circular regression function $m(x)$ accounting for measurement error proposed, some remarks on comparisons between them are in order. Among the three, $\hat{m}_{\mathrm{CE}}(x)$ stands out as the only one that is developed under the assumption of normal measurement error. Both $\hat{m}_{\mathrm{DK}}(x)$ and $\hat{m}_{\mathrm{OS}}(x)$ are practically applicable as long as $\phi_U(t)$ never vanishes since $\hat{m}_{\mathrm{DK}}(x)$ depends on the integral in (8), and $\hat{m}_{\mathrm{OS}}(x)$ hinges on the integral in (11), both integrals involve $\phi_U(\cdot)$ at the denominator of the integrand.

By construction, $\hat{m}_{\mathrm{OS}}(x)$ corrects naive estimation for measurement error at a higher level in the sense that, unlike $\hat{m}_{\mathrm{DK}}(x)$ that accounts for measurement error by correcting the naive weight, $\hat{m}_{\mathrm{OS}}(x)$ corrects naive estimation of $g_\ell(x)$ via transforming the naive estimator as a whole. Similar to $\hat{m}_{\mathrm{DK}}(x)$, $\hat{m}_{\mathrm{CE}}(x)$ results from correcting the normalized naive weight, but we correct it in its entirety instead of correcting it term-by-term as done in deriving the new weight for $\hat{m}_{\mathrm{DK}}(x)$. In summary, we correct the naive estimator $\hat{m}^*(x)$ at an increasingly higher level as we progress from the first non-naive estimator to the third one in Section 3.1–3.3, i.e., from $\hat{m}_{\mathrm{DK}}(x)$ to $\hat{m}_{\mathrm{CE}}(x)$, then to $\hat{m}_{\mathrm{OS}}(x)$.

On the other hand, $\hat{m}_{\mathrm{DK}}(x)$ and $\hat{m}_{\mathrm{OS}}(x)$ exploit integral transforms of a similar form, with the former using (8) to correct a naive weight, while the latter utilizing (11) to correct a naive estimator of $g_\ell(x)$. Although no integral transform is involved in $\hat{m}_{\mathrm{CE}}(x)$, it intrinsically relates to a similar integral transform because of Lemma 3.1 that motivates it. To see this connection, consider the special case with $n = 1$ in Lemma 3.1 and let $V = U + A$, where U is the measurement error. The key there to deriving an unbiased estimator of $g(X)$ based on W is to find a random variable A such that all moments of $V = U + A$ are equal to zero, because then one has $\mathbb{E}\{g(W + A)|X\} = \mathbb{E}\{g(X + U + A)|X\} = \mathbb{E}\{g(X + V)|X\} = g(X)$ according to Lemma 3.1. If one further assumes $A \perp U$, then $\phi_V(t) = \phi_U(t)\phi_A(t)$, and thus $\phi_A(t) = \phi_V(t)/\phi_U(t)$. By the Fourier inversion theorem, the probability density function of A is $f_A(a) = (2\pi)^{-1} \int e^{-ita}\phi_V(t)/\phi_U(t)\, dt$, which is an integral transform similar to (8) and (11). Depending on the distribution of U, there may not exist such a V (with all moments equal to zero); consequently, the so-obtained $f_A(a)$ may not be well-defined. But, it has been shown that, if $U \sim N(0, \sigma_u^2)$, then letting $A = i\sigma_u Z$ gives rise to a desired V, with $Z \sim N(0, 1)$ and $Z \perp (U, X)$ [40]. In conclusion, all three proposed estimators have connections with some integral transform relating to deconvolution. After comparing the three proposed estimators in regard to assumptions, constructions, and rationales, we present more formally technical conditions imposed on these estimators next, under which we investigate their asymptotic properties.

4. Asymptotic analysis

4.1. Conditions and notations

When studying properties of their proposed local polynomial estimators with error in covariates, [8] imposed different sets of conditions on the kernel $K(t)$ depending on the measurement error distribution characterized by $\phi_U(t)$. These conditions are also needed for our proposed estimators depending on the smoothness of U [14]. More specifically, if $\lim_{t \to \infty} t^\beta \phi_U(t) = c$ and $\lim_{t \to \infty} t^{\beta+1} \phi_U'(t) = -c\beta$ for some constants $c > 0$ and $\beta > 1$, then we say that (the

distribution of) U is ordinary smooth of order β. The Laplace distribution is an example of ordinary smooth distribution of order 2. If, for some positive constants d_0, d_1, γ, and β, and some real-valued constants β_0 and β_1, $d_0|t|^{\beta_0}\exp(-|t|^{\beta}/\gamma) \leq |\phi_U(t)| \leq d_1|t|^{\beta_1}\exp(-|t|^{\beta}/\gamma)$, as $t \to \infty$, then we say that U is super smooth of order β. A normal distribution is super smooth of order 2.

Because all proposed estimators are of the form $\hat{m}_{\cdot}(x) = \text{atan2}[\hat{g}_{1,\cdot}(x), \hat{g}_{2,\cdot}(x)]$, where "$\cdot$" generically refers to an acronym ("DK", "CE", or "OS") relating to a proposed estimator, we first study asymptotic properties of $\hat{g}_{1,\cdot}(x)$ and $\hat{g}_{2,\cdot}(x)$. We then use these findings to establish the asymptotic moments and distribution of $\hat{m}_{\cdot}(x)$ through a Taylor expansion of the atan2 function. Conditions frequently referenced in theorems presented next are listed below. Similar conditions are included in [8] and [22] in their asymptotic analyses.

> **Condition O**: For $\ell = 0, 1, 2, 3$, $\|\phi_K^{(\ell)}(t)\|_\infty < \infty$ and $\int (|t|^\beta + |t|^{\beta-1})|\phi_K^{(\ell)}(t)|dt < \infty$.
> For $0 \leq k, \ell \leq 2$, $\int |t|^{2\beta}|\phi_K^{(k)}(t)||\phi_K^{(\ell)}(t)|dt < \infty$, and $\|\phi_U'(t)\|_\infty < \infty$.
> **Condition S**: For $\ell = 0, 1, 2$, $\|\phi_K^{(\ell)}(t)\|_\infty < \infty$; $\phi_K(t)$ is supported on $[-1, 1]$.

The first set of conditions is imposed for asymptotic analysis when U is o̲rdinary smooth, and the latter set is imposed when U is s̲uper smooth. For example, the characteristic function of a mean-zero Laplace distribution with variance σ_u^2, $\phi_U(t) = 1/(1+\sigma_u^2 t^2/2)$, satisfies **Condition O**. As for the kernel $K(\cdot)$, the Gaussian kernel satisfies constraints under **Condition O** because, with $\phi_K(t) = e^{-t^2/2}$, $\phi_K(t)$ is continuously differentiable with bounded derivatives, and both integral constraints relate to integrals that define moments of normal or half normal distributions, which are well-defined at all orders. An example of kernels satisfying **Condition S** that we use in our simulation experiments has its Fourier transform given by $\phi_K(t) = (1 - t^2)^3 \mathbf{1}_{\{-1 \leq t \leq 1\}}$. Lastly, for $\ell \in \mathbb{N}_0$, define $\mu_\ell = \int t^\ell K(t)dt$ and $\nu_\ell = \int t^\ell K^2(t)dt$.

4.2. Asymptotic results

We derive the asymptotic bias and variance of $\hat{m}_{\text{DK}}(x)$, $\hat{m}_{\text{CE}}(x)$, and $\hat{m}_{\text{OS}}(x)$ in Appendices D, E, and F, respectively. For ease of comparison, we subsume results relating to the asymptotic bias of these estimators in Theorem 4.1, followed by results regarding their asymptotic variance summarized in Theorem 4.2.

Theorem 4.1. *When U is ordinary smooth of order β, then under Condition O, if $nh^{1+2\beta} \to \infty$ as $n \to \infty$ and $h \to 0$,*

(i)

$$
Bias\{\hat{m}_{DK}(x)|X\}
$$
$$
= \frac{h^2\mu_2}{2}\left\{m^{(2)}(x) + 2m'(x)\frac{m_1(x)m_1'(x) + m_2(x)m_2'(x)}{m_1^2(x) + m_2^2(x)}\right\} + o(h^2)
$$
$$
+ \frac{f_W(x)}{nh^{1+2\beta}f_X^2(x)}\left[\frac{\{\eta(0,2) - \eta(1,1)\}\{m_1^*(x)m_2(x) - m_1(x)m_2^*(x)\}}{\mu_2\{m_1^2(x) + m_2^2(x)\}}\right.
$$
$$
\left. - \eta(0,0)\frac{m_1(x)m_2(x)\{\xi_1^*(x) - \xi_2^*(x)\} + \psi^*(x)\{m_2^2(x) - m_1^2(x)\}}{\{m_1^2(x) + m_2^2(x)\}^2}\right]
$$

$$+ o_p \left(\frac{1}{nh^{1+2\beta}} \right),$$

where $\xi_1^*(x) = \mathbb{E}\{(\sin\Theta)^2 | W = x\}$, $\xi_2^*(x) = \mathbb{E}\{(\cos\Theta)^2 | W = x\}$, $\psi^*(x) = \mathbb{E}(\sin\Theta\cos\Theta | W = x)$, *and, for* $k, \ell \in \{0, 1, 2\}$, $\eta(k, \ell) = \{i^{-(k+\ell)}/(2\pi c^2)\}$ $\int t^{2\beta} \phi_K^{(k)}(t) \phi_K^{(\ell)}(-t) dt$, *in which c is the positive constant appearing in the definition of ordinary smooth U;*

(ii)

$$Bias\{\hat{m}_{OS}(x) | \mathbf{X}\}$$

$$= \frac{h^2 \mu_2}{2} \left[\frac{m_2(x) \mathcal{T}_U(M_1)(x) - m_1(x) \mathcal{T}_U(M_2)(x)}{f_X(x) \{m_1^2(x) + m_2^2(x)\}} \right] + o\left(h^2\right)$$

$$+ \frac{\eta(0,0) f_W(x)}{nh^{1+2\beta} f_X^2(x) \{m_1^2(x) + m_2^2(x)\}^2} \left[m_1(x) m_2(x) \{\sigma_1^{*2}(x) - \sigma_2^{*2}(x)\} \right.$$

$$\left. + \{m_2^2(x) - m_1^2(x)\} \{\psi^*(x) - m_1^*(x) m_2^*(x)\} \right] + o_p \left(\frac{1}{nh^{1+2\beta}} \right),$$

where $\mathcal{T}_U$ *is the integral transform defined in* [(12)](12), $M_\ell(x) = m_\ell^*(x) f_W^{(2)}(x) + m_\ell^{*(2)}(x) f_W(x)$, *for* $\ell = 1, 2$, $\sigma_1^{*2}(x) = Var(\sin\Theta | W = x)$, *and* $\sigma_2^{*2}(x) = Var(\cos\Theta | W = x)$.

When U is super smooth of order β, if $nh^{1-2\beta_2} \exp(-2h^{-\beta}/\gamma) \to \infty$ as $n \to \infty$ and $h \to 0$, where $\beta_2 = \beta_0 I(\beta_0 < 0.5)$, then, under Condition S,

(i)

$$Bias\{\hat{m}_{DK}(x) | \mathbf{X}\}$$

$$= \frac{h^2 \mu_2}{2} \left\{ m^{(2)}(x) + 2m'(x) \frac{m_1(x) m_1'(x) + m_2(x) m_2'(x)}{m_1^2(x) + m_2^2(x)} \right\} + o(h^2)$$

$$+ O_p \left(\frac{\exp(2h^{-\beta}/\gamma)}{nh^{1-2\beta_2}} \right);$$

(ii)

$$Bias\{\hat{m}_{OS}(x) | \mathbf{X}\}$$

$$= \frac{h^2 \mu_2}{2} \left[\frac{m_2(x) \mathcal{T}_U(M_1)(x) - m_1(x) \mathcal{T}_U(M_2)(x)}{f_X(x) \{m_1^2(x) + m_2^2(x)\}} \right] + o(h^2)$$

$$+ O_p \left(\frac{\exp(2h^{-\beta}/\gamma)}{nh^{1-2\beta_2}} \right);$$

(iii) more specifically assuming $U \sim N(0, \sigma_u^2)$, and thus $\beta = 2$, $\beta_2 = 0$, and $\gamma = 2/\sigma_u^2$, then, if B/n tends to a positive constant as $n \to \infty$,

$$Bias\{\hat{m}_{CE}(x) | \mathbf{X}\}$$

$$= \frac{h^2 \mu_2}{2} \left\{ m^{(2)}(x) + 2m'(x) \frac{m_1(x) m_1'(x) + m_2(x) m_2'(x)}{m_1^2(x) + m_2^2(x)} \right\}$$

$$+ \frac{h^2}{f_X(x) \left\{ m_1^2(x) + m_2^2(x) \right\}^2} \Big[2(1 - \mu_2) f_X'(x) m_1(x) m_2(x) \left\{ m_1(x) m_1'(x) \right.$$

$$-m_2(x) m_2'(x) \right\} + \mu_2 \left\{ f_X(x) - f_X'(x) \right\} \left\{ m_1^2(x) - m_2^2(x) \right\} \left\{ m_1(x) m_2'(x) \right.$$

$$+m_1'(x) m_2(x) \right\} \Big] + o(h^2) + O_p \left(\frac{\exp(\sigma_u^2 h^{-2})}{nh} \right).$$

According to Theorem 4.1, the dominating bias of each proposed estimator consists of a term of order $O(h^2)$ and a term of order depending on the smoothness of U. In particular, the term of order $O(h^2)$ in the dominating bias of $\hat{m}_{\mathrm{DK}}(x)$ coincides with that of $\tilde{m}(x)$ [see Theorem 4 in 12], which depends on the curvature of $m(x)$. The term that depends on the smoothness of U can involve naive counterparts of $m_\ell(x)$ and other functionals, which is dominated by the first term provided that $n \to \infty$ much faster than $h \to 0$, and the requirement on n to achieve this is more demanding when U is super smooth than when it is ordinary smooth, or when the order of smoothness β is larger.

Theorem 4.2. *When U is ordinary smooth of order β, under Condition O, if $nh^{1+2\beta} \to \infty$ as $n \to \infty$ and $h \to 0$, then*

(i)

$$Var \{ \hat{m}_{DK}(x) | \boldsymbol{X} \}$$
$$= \frac{f_W(x) \left\{ m_1^2(x) \xi_2^*(x) + m_2^2(x) \xi_1^*(x) - 2m_1(x) m_2(x) \psi^*(x) \right\} \eta(0, 0)}{nh^{1+2\beta} f_X^2(x) \left\{ m_1^2(x) + m_2^2(x) \right\}^2}$$
$$+ o_p \left(\frac{1}{nh^{1+2\beta}} \right);$$

(ii)

$$Var \{ \hat{m}_{OS}(x) | \boldsymbol{X} \}$$
$$= \frac{f_W(x) \eta(0, 0)}{nh^{1+2\beta} f_X^2(x) \left\{ m_1^2(x) + m_2^2(x) \right\}^2} \Big(m_1^2(x) \xi_2^*(x) + m_2^2(x) \xi_1^*(x)$$
$$- 2m_1(x) m_2(x) \psi^*(x) - \left\{ m_1(x) m_2^*(x) - m_1^*(x) m_2(x) \right\}^2$$
$$+ 4m(x) \Big[m_1(x) m_2(x) \left\{ \sigma_2^{*2}(x) - \sigma_1^{*2}(x) \right\} + \phi^*(x) \{ m_1^2(x)$$
$$- m_1^2(x) \} \Big] \Big) + o_p \left(\frac{1}{nh^{1+2\beta}} \right),$$

where $\phi^*(x) = Cov(\sin \Theta, \cos \Theta | W = x)$.

When U is super smooth of order β, if $nh^{1-2\beta_2} \exp(-2h^{-\beta}/\gamma) \to \infty$ as $n \to \infty$ and $h \to 0$, then under Condition S,

(i) $Var\{ \hat{m}_{DK}(x) | \boldsymbol{X} \} = O_p \left(\dfrac{\exp(2h^{-\beta}/\gamma)}{nh^{1-2\beta_2}} \right)$;

(ii) $Var\{ \hat{m}_{OS}(x) | \boldsymbol{X} \} = O_p \left(\dfrac{\exp(2h^{-\beta}/\gamma)}{nh^{1-2\beta_2}} \right)$;

(iii) more specifically assuming $U \sim N(0, \sigma_u^2)$, and thus $\beta = 2$, $\beta_2 = 0$, and $\gamma = 2/\sigma_u^2$, then, if B/n tends to a positive constant as $n \to \infty$,

$$Var\{\hat{m}_{CE}(x)|X\}$$
$$= 2h^2 \left\{ (2\mu_2 - 1)\frac{f_X'(x)}{f_X(x)} - \mu_2 \right\} \frac{m_1(x)m_2(x)\{m_1(x)m_2'(x) + m_1'(x)m_2(x)\}}{\{m_1^2(x) + m_2^2(x)\}^2}$$
$$+ o(h^2) + O_p \left(\frac{\exp\left(\sigma_u^2 h^{-2}\right)}{nh} \right).$$

According to Theorem 4.2, having super smooth U tends to lead to a more variable $\hat{m}.(x)$. When U is ordinary smooth, the dominating variance of $\hat{m}_{OS}(x)$ is equal to that of $\hat{m}_{DK}(x)$ plus additional terms depending on $m(x)$, $m_\ell(x)$, and other functions relating to naive inference. Having these additional terms implies that the variability of $\hat{m}_{OS}(x)$ is more affected by discrepancies between functionals of actual interest when drawing inference and their naive counterparts, such as how $m_\ell(x)$ differs from $m_\ell^*(x)$. Our asymptotic results regarding $\hat{m}_{DK}(x)$ elaborate on the impacts of measurement error when compared with findings in Di Marzio et al. [11, see Result 5 in Section 4.1.2] for a similar estimator that uses a local constant weight in place of our local linear weight. Similar to the comparison between the Nadaraya-Watson estimator and a local linear estimator of a regression function for a linear response [see Table 2.1 in 15], the dominating variance of the estimator in Di Marzio et al. [11] is of the same order as that of $\hat{m}_{DK}(x)$; and, although of the same order (of $O(h^2)$), the dominating bias of the former has an extra term given by $2m'(x)f_X'(x)/f_X(x)$ compared with the multiplier of $h^2\mu_2/2$ in the asymototic bias of $\hat{m}_{DK}(x)$. Besides the results summarized in Theorems 4.1 and 4.2, we also derived in Appendices E and F more general results for $\hat{m}_{CE}(x)$ and $\hat{m}_{OS}(x)$ when local polynomial weights of order p are used in general.

Finally, we establish asymptotic normality of $\hat{m}.(x)$ in Appendix G by exploiting established normality results in existing literature regarding the same type of estimators as $\hat{g}_{1,.}(x)$ and $\hat{g}_{2,.}(x)$ involved in our estimators.

5. Bandwidth selection

5.1. Bandwidth selection in the absence of measurement error

The performance of local polynomial estimators is sensitive to the bandwidth h. Compared to deriving a plug-in type of bandwidth, a more practically feasible and widely applicable bandwidth selection method is cross validation (CV) based a risk function for assessing estimation quality, such as k-fold or leave-one-out CV. We focus on 5-fold CV in this article for illustration purposes.

In the context of estimating a circular regression function $m(x)$ in the absence of measurement error, a sensible risk function associated with the estimator $\tilde{m}(x)$ is the mean cosine dissimilarity, $\mathbb{E}\{1 - \cos(\Theta - \tilde{m}(X))\}$. To highlight the dependence of $\tilde{m}(x)$ on h and the data used to construct the estimator, we re-express the estimator in (2) as $\tilde{m}(x; \Theta, X, h)$. A 5-fold CV entails first randomly splitting the raw data into five data sets of (approximately) equal size, then computing the following loss function formulated using cosine dissimilarity at each

candidate h,

$$D(\tilde{m}(X); \boldsymbol{X}, h) = \sum_{k=1}^{5} \frac{1}{|I_k|} \sum_{j \in I_k} \left\{ 1 - \cos(\Theta_j - \tilde{m}(X_j; \boldsymbol{\Theta}^{(-k)}, \boldsymbol{X}^{(-k)}, h)) \right\}, \qquad (13)$$

where I_k is the index set of the k-th subsample of size $|I_k|$, and $(\boldsymbol{\Theta}^{(-k)}, \boldsymbol{X}^{(-k)})$ include all data in the original sample except for those in the k-th subsample, for $k = 1, \ldots, 5$. Then $h_{\text{true}} = \arg\min_{h \in \mathcal{H}} D(\tilde{m}(X); \boldsymbol{X}, h)$ is the chosen bandwidth based on true covariate data, where $\mathcal{H}$ is the set of candidate values of h.

In the presence of covariate measurement error, an immediate hurdle in implementing the traditional CV outlined above is that, for a non-naive estimator of the regression function $\hat{m}.(x)$, the cosine dissimilarity between Θ_j and $\hat{m}.(X_j)$ cannot be computed now that X_j is unobserved but W_j is instead. Naively substituting X_j with W_j again does not solve the problem here because the cosine dissimilarity between Θ_j and $\hat{m}.(W_j)$ is a misrepresentation of the actual loss associated with the j-th observation. We present two approaches to overcome the hurdle next.

5.2. Bandwidth selection using SIMEX

The first approach is proposed by [9] as a combination of simulation-extrapolation [SIMEX, 5, Chapter 5] and CV. Ideally, one would follow the traditional CV to find

$$h_{\text{ideal}} = \arg\min_{h \in \mathcal{H}} \sum_{k=1}^{5} \frac{1}{|I_k|} \sum_{j \in I_k} \left\{ 1 - \cos(\Theta_j - \hat{m}.(X_j; \boldsymbol{\Theta}^{(-k)}, \boldsymbol{W}^{(-k)}, h)) \right\}, \qquad (14)$$

where we stress that $\hat{m}.(x)$ is a non-naive estimator of $m(x)$ constructed using error contaminated data, but we evaluate x at X_j's when computing the loss in (14). Compared with (13), a major distinction in the ideal CV in the presence of measurement error is the mismatch between the training data and the validation data exhibited in (14): while *error-contaminated* data $(\boldsymbol{\Theta}^{(-k)}, \boldsymbol{W}^{(-k)})$ are used to train the estimator $\hat{m}.(x)$, the *error-free* data $(\boldsymbol{\Theta}^{(k)}, \boldsymbol{X}^{(k)})$ are used to validate the estimator. Clearly, the "ideal" bandwidth h_{ideal} is not attainable with $\boldsymbol{X}$ now unobserved. Naive CV yields

$$h_{\text{naive}} = \arg\min_{h \in \mathcal{H}} \sum_{k=1}^{5} \frac{1}{|I_k|} \sum_{j \in I_k} \left\{ 1 - \cos(\Theta_j - \hat{m}.(W_j; \boldsymbol{\Theta}^{(-k)}, \boldsymbol{W}^{(-k)}, h)) \right\},$$

which is expected to distort h_{ideal} due to evaluating x in $\hat{m}.(x)$ at W_j's instead of X_j's. To learn how h_{ideal} takes into consideration that, even though the estimator $\hat{m}.(x)$ is constructed based on error-contaminated data, $\hat{m}.(x)$ should be evaluated at error-free data when computing the loss, we "simulate" twice the ideal CV mimicking (14) to obtain two bandwidths.

The first simulated version of ideal bandwidth is

$$h_1 = \arg\min_{h \in \mathcal{H}} \frac{1}{B} \sum_{b=1}^{B} \sum_{k=1}^{5} \frac{1}{|I_k|} \sum_{j \in I_k} \left\{ 1 - \cos(\Theta_j - \hat{m}.^*(W_j; \boldsymbol{\Theta}^{(-k)}, \boldsymbol{W}_b^{*(-k)}, h)) \right\}, \qquad (15)$$

where, for $b = 1, \ldots, B$, $\boldsymbol{W}_b^* = (W_{1,b}^*, \ldots, W_{n,b}^*)^\top = \boldsymbol{W} + \boldsymbol{U}_b^*$, in which $\boldsymbol{U}_b^* = (U_{1,b}^*, \ldots, U_{n,b}^*)^\top$ is a random sample generated from the distribution that errors in $\boldsymbol{U} = (U_1, \ldots, U_n)^\top$

follow, and $\hat{m}^*_.(x)$ is the same type of estimator as $\hat{m}_.(x)$ but for estimating $m^*(x)$. For example, when selecting a bandwidth used in the deconvoluting kernel estimator, the estimator in (14) is $\hat{m}_{\text{DK}}(x)$, which is constructed based on $(\boldsymbol{\Theta}, \boldsymbol{W})$ to estimate $m(x)$ as the circular mean of Θ given $X = x$. Accordingly, the estimator in (15) is $\hat{m}^*_{\text{DK}}(x)$ as the deconvoluting kernel estimator constructed based on $(\boldsymbol{\Theta}, \boldsymbol{W}^*_b)$ to estimate $m^*(x)$ as the mean of Θ given $W = x$. The CV leading to h_1 is parallel to that leading to h_{ideal} in the sense that, like $\boldsymbol{W}$ being a noisy surrogate of $\boldsymbol{X}$ in (14), $\boldsymbol{W}^*_b$ is also a noisy surrogate of $\boldsymbol{W}$ in (15) with the same severity of error contamination since $\text{Var}(W^*_{j,b}|W_j) = \text{Var}(W_j|X_j) = \sigma^2_u$; and, more importantly, (15) also has a mismatch between training data and validation data, with the former involving the (further) contaminated data $\boldsymbol{W}^{*(-k)}_b$, whereas the latter being the data before (further) contamination, $\boldsymbol{W}^{(k)}$. Similarly, the second simulated bandwidth results from yet another round of CV parallel to that producing h_1,

$$h_2 = \arg\min_{h \in \mathcal{H}} \frac{1}{B} \sum_{b=1}^{B} \sum_{k=1}^{5} \frac{1}{|I_k|} \sum_{j \in I_k} \left\{ 1 - \cos(\Theta_j - \hat{m}^{**}_.(W^*_{j,b}; \boldsymbol{\Theta}^{(-k)}, \boldsymbol{W}^{**(-k)}_b, h)) \right\},$$

where $\boldsymbol{W}^{**}_b = \boldsymbol{W}^*_b + \boldsymbol{U}^{**}_b$, $\boldsymbol{U}^{**}_b$ is generated in the same way as $\boldsymbol{U}^*_b$, with $\boldsymbol{U}^{**}_b$, $\boldsymbol{U}^*_b$, and $\boldsymbol{U}$ all independent of each other, for $b = 1, \ldots, B$, and $\hat{m}^{**}_.(\cdot)$ is the same type of estimator as $\hat{m}_.(\cdot)$, but for estimating the regression function when regressing Θ_j's on $W^*_{j,b}$'s.

The simulation step of SIMEX ends with outputting h_1 and h_2. At the extrapolation step of SIMEX, one attempts to recover h_{ideal} by learning from the connection between h_1 and h_2 because, with the parallel design in the two rounds of ideal CV leading to h_1 and h_2, how h_1 compares with h_2 should be similar to how h_{ideal} compares with h_1. In particular, [9] showed that $\log h_{\text{ideal}} - \log h_1 \approx \log h_1 - \log h_2$ when σ_u is small, and thus proposed to approximate h_{ideal} by $h_{\text{SIMEX}} = h_1^2/h_2$. This completes the process of bandwidth selection that we refer to as CV-SIMEX that yields h_{SIMEX} as the selected bandwidth. An obvious concern of this method is the computational burden. Even the traditional CV without the measurement error complication is computationally demanding in general; and this method requires $2 \times B$ rounds of such CV based on data noisier than the original data.

5.3. Bandwidth selection using complex error

Acknowledging that h_{ideal} in (14) is not attainable solely because the loss function there cannot be computed based on observed data $(\boldsymbol{\Theta}, \boldsymbol{W})$, we propose a second approach for bandwidth selection where we first estimate this loss function, then we choose a bandwidth by minimizing the estimated loss.

Recall that the loss function in (14) is

$$D(\hat{m}_.(X); \boldsymbol{W}, h) = \sum_{k=1}^{5} \frac{1}{|I_k|} \sum_{j \in I_k} \left\{ 1 - \cos(\Theta_j - \hat{m}_.(X_j; \boldsymbol{\Theta}^{(-k)}, \boldsymbol{W}^{(-k)}, h)) \right\}. \quad (16)$$

With observed data $(\boldsymbol{\Theta}, \boldsymbol{W})$ plugged in, along with a candidate h, (16) is unknown purely due to its dependence on X. If we view this unknown loss as a function of X, like $g(X)$ in Lemma 3.1, then we can formulate an unbiased estimator of this loss function when $U \sim N(0, \sigma^2_u)$. This

yields an estimated loss given by

$$
\begin{aligned}
&\hat{D}(\hat{m}.(W^*); \boldsymbol{W}, h) \\
&= \frac{1}{B} \sum_{b=1}^{B} \sum_{k=1}^{5} \frac{1}{|I_k|} \sum_{j \in I_k} \left\{ 1 - \cos(\Theta_j - \hat{m}.(W_{j,b}^*; \boldsymbol{\Theta}^{(-k)}, \boldsymbol{W}^{(-k)}, h)) \right\},
\end{aligned}
\tag{17}
$$

where $\{ W_{j,b}^* = W_j + i\sigma_u Z_{j,b}, \ j = 1, \ldots, n \}_{b=1}^{B}$ are generated in the same way as the complex-valued covariate data in (10). If $D(\hat{m}.(X); \boldsymbol{W}, h)$ in (16) is an entire function of X, and $\boldsymbol{W}$ results from additive normal error contamination of X, then, by Lemma 3.1, $\hat{D}(\hat{m}.(W^*); \boldsymbol{W}, h)$ in (17) is an unbiased estimator of $D(\hat{m}.(X); \boldsymbol{W}, h)$. Minimizing $\hat{D}(\hat{m}.(W^*); \boldsymbol{W}, h)$ with respect to h gives a bandwidth that is expected to improve over h_{naive} and be closer to h_{ideal}. Compared with CV-SIMEX, this bandwidth selection procedure is less cumbersome but it still demands heavy computation because, for each of the five validation data sets, one estimates $m(x)$ based on the corresponding training data at $|I_k| \times B$ complex-valued x's, i.e., at $\{ W_{j,b}^*, \ j \in I_k \}_{b=1}^{B}$. Next, we propose a revised procedure to drastically lighten the computational burden. The idea is to move the averaging in (17) "$B^{-1} \sum_{b=1}^{B}$" further inside the summand so that one can use the trick inspired by Lemma 3.1 to estimate $\hat{m}.(X_j; \boldsymbol{\Theta}^{(-k)}, \boldsymbol{W}^{(-k)}, h)$ in (16) instead of estimating the loss function as a whole.

Take the complex error estimator $\hat{m}_{\text{CE}}(x)$ as an example. In (17), the estimate of $\hat{m}_{\text{CE}}(X_j; \boldsymbol{\Theta}^{(-k)}, \boldsymbol{W}^{(-k)}, h)$ is, for each $b \in \{1, \ldots, B\}$,

$$
\begin{aligned}
&\hat{m}_{\text{CE}}(W_{j,b}^*; \boldsymbol{\Theta}^{(-k)}, \boldsymbol{W}^{(-k)}, h) \\
&= \text{atan2} \left[\hat{g}_{1,\text{CE}}(W_{j,b}^*; \boldsymbol{\Theta}^{(-k)}, \boldsymbol{W}^{(-k)}, h), \ \hat{g}_{2,\text{CE}}(W_{j,b}^*; \boldsymbol{\Theta}^{(-k)}, \boldsymbol{W}^{(-k)}, h) \right],
\end{aligned}
\tag{18}
$$

where $\hat{g}_{\ell,\text{CE}}(W_{j,b}^*; \boldsymbol{\Theta}^{(-k)}, \boldsymbol{W}^{(-k)}, h)$ is the estimate $\hat{g}_{\ell,\text{CE}}(W_{j,b}^*)$ based on data $(\boldsymbol{\Theta}^{(-k)}, \boldsymbol{W}^{(-k)})$ with the bandwidth set at h, for $\ell = 1, 2$. By moving the outer averaging in (17) towards $\hat{m}_{\text{CE}}(\cdot)$ and further passing through the atan2 function in (18), we have a different estimate of the loss function given by

$$
\begin{aligned}
&\hat{D}^*(\hat{m}_{\text{CE}}(W^*); \boldsymbol{W}, h) \\
&= \sum_{k=1}^{5} \frac{1}{|I_k|} \sum_{j \in I_k} \left\{ 1 - \cos\left(\Theta_j - \text{atan2}\left[\frac{1}{B} \sum_{b=1}^{B} \hat{g}_{1,\text{CE}}(W_{j,b}^*; \boldsymbol{\Theta}^{(-k)}, \boldsymbol{W}^{(-k)}, h), \right. \right. \right. \\
&\qquad \left. \left. \left. \frac{1}{B} \sum_{b=1}^{B} \hat{g}_{2,\text{CE}}(W_{j,b}^*; \boldsymbol{\Theta}^{(-k)}, \boldsymbol{W}^{(-k)}, h) \right] \right) \right\},
\end{aligned}
\tag{19}
$$

where, using the weight function $\mathcal{L}^*(\cdot)$ defined in (10), the first argument of atan2$[\cdot, \cdot]$ is

$$
\begin{aligned}
&\frac{1}{B} \sum_{b=1}^{B} \hat{g}_{1,\text{CE}}(W_{j,b}^*; \boldsymbol{\Theta}^{(-k)}, \boldsymbol{W}^{(-k)}, h) \\
&= \frac{1}{B} \sum_{b=1}^{B} \frac{1}{|I^{(-k)}|} \sum_{\ell \in I^{(-k)}} \sin(\Theta_\ell) \mathcal{L}^*(W_\ell - W_{j,b}^*) \\
&= \frac{1}{|I^{(-k)}|} \sum_{\ell \in I^{(-k)}} \sin(\Theta_\ell) \overline{\mathcal{L}^*}_{\ell,j},
\end{aligned}
\tag{20}
$$

in which $I^{(-k)}$ is the index set of the training data $(\boldsymbol{\Theta}^{(-k)}, \boldsymbol{W}^{(-k)})$ of size $|I^{(-k)}|$, and $\overline{\mathcal{L}^*}_{\ell,j} = B^{-1} \sum_{b=1}^{B} \mathcal{L}^*(W_\ell - W_{j,b}^*)$. In other words, (20) is an estimate of the same form as $\hat{g}_{1,\text{CE}}(W_{j,b}^*; \boldsymbol{\Theta}^{(-k)}, \boldsymbol{W}^{(-k)}, h)$ in (18), but uses a different weight, $\overline{\mathcal{L}^*}_{\ell,j}$. Similarly, the second argument of $\text{atan2}[\cdot, \cdot]$ in (19) is an estimate of the same form as $\hat{g}_{2,\text{CE}}(W_{j,b}^*; \boldsymbol{\Theta}^{(-k)}, \boldsymbol{W}^{(-k)}, h)$ in (18) but uses the weight $\overline{\mathcal{L}^*}_{\ell,j}$. Define $\hat{g}_{1,\text{CE}}(\widetilde{W}_j^*; \boldsymbol{\Theta}^{(-k)}, \boldsymbol{W}^{(-k)}, h)$ as the estimate in (20), and similarly define $\hat{g}_{2,\text{CE}}(\widetilde{W}_j^*; \boldsymbol{\Theta}^{(-k)}, \boldsymbol{W}^{(-k)}, h)$, where $\widetilde{W}_j^* = (W_{j,1}^*, \ldots, W_{j,B}^*)^\top$; and using these two estimates we have the corresponding estimate of the regression function, denoted by $\hat{m}_{\text{CE}}(\widetilde{W}_j^*; \boldsymbol{\Theta}^{(-k)}, \boldsymbol{W}^{(-k)}, h)$. Now the new estimated loss in (19) can be re-expressed as

$$
\begin{aligned}
&\hat{D}^*(\hat{m}_{\text{CE}}(W^*); \boldsymbol{W}, h) \\
&= \sum_{k=1}^{5} \frac{1}{|I_k|} \sum_{j \in I_k} \left\{ 1 - \cos(\Theta_j - \hat{m}_{\text{CE}}(\widetilde{W}_j^*; \boldsymbol{\Theta}^{(-k)}, \boldsymbol{W}^{(-k)}, h)) \right\}.
\end{aligned}
\tag{21}
$$

It is clear from (21) that $\hat{D}^*(\hat{m}_{\text{CE}}(W^*); \boldsymbol{W}, h)$ is similar to $D(\hat{m}_{\text{CE}}(X); \boldsymbol{W}, h)$ in (16) in terms of computation burden, and clearly much less burdensome than the initial estimated loss $\hat{D}(\hat{m}_{\text{CE}}(W^*); \boldsymbol{W}, h)$ in (17).

Admittedly, by passing the averaging operation through nonlinear functions like we did to modify $\hat{D}(\hat{m}_{\text{CE}}(W^*); \boldsymbol{W}, h)$, we introduce bias as an estimator of $D(\hat{m}_{\text{CE}}(X); \boldsymbol{W}, h)$. This is a small price we pay for a substantial computational gain. Regardless, by construction, $\hat{D}^*(\hat{m}_{\text{CE}}(W^*); \boldsymbol{W}, h)$ is a sensible estimator that consistently estimates $D(\hat{m}_{\text{CE}}(X); \boldsymbol{W}, h)$ given $(\boldsymbol{\Theta}, \boldsymbol{W})$ as $B \to \infty$. Consequently, $h_{\text{CE}} = \arg\min_{h \in \mathcal{H}} \hat{D}^*(\hat{m}_{\text{CE}}(W^*); \boldsymbol{W}, h)$ still tends to be closer to h_{ideal} than h_{naive} is. We call this proposed bandwidth selection method that involves complex error CV-CE, henceforth.

All bandwidth selection methods considered here require a pre-specified set of candidate values $\mathcal{H}$. A practically effective approach to specify $\mathcal{H}$ is to first find an optimal bandwidth h_0 for a simpler (to compute) estimator, such as the naive estimator. Then one creates a range of bandwidths surrounding h_0, for instance, $[0.8h_0, 1.3h_0]$. If it is always, say, the upper bound that is selected in a cross-validation for a non-naive estimator, which can happen in the presence of severe error contamination in the covariate data, we recommend adjusting the initial search window by enlarging the upper bound slightly. We find this practice of specifying $\mathcal{H}$ via trial-and-error more effective and feasible than alternative approaches, such as those based on minimizing the mean integrated squared error (MISE) of a non-naive estimator. As we show in Appendix H, even when the asymptotic MISE can be derived in some model settings, unknown functionals related to $m(x)$ that the asymptotic MISE depends on make it a practically difficult to use quantity for the search of suitable bandwidths.

6. Simulation study

6.1. Design of simulation

We are now in the position to inspect the finite sample performance of our proposed non-naive estimators for the circular regression function $m(x)$. Simulation experiments presented in this section are designed to address three issues: (i) comparisons between four non-naive estimators, $\hat{m}_{\text{DK}}(x)$, $\hat{m}_{\text{CE}}(x)$, $\hat{m}_{\text{OS}}(x)$, and the deconvoluting kernel estimator with a local

constant weight proposed by Di Marzio et al. [11], and the naive estimator $\hat{m}^*(x)$, using the ideal estimator $\tilde{m}(x)$ as a benchmark; (ii) comparisons between different bandwidth selection methods, using some optimal bandwidth h_{opt} to be defined momentarily as a benchmark; (iii) impacts of the measurement error distribution on different estimators.

For each Monte Carlo replicate, we first generate true covariate data $\{X_j\}_{j=1}^n$ of size $n \in \{50, 100, 250\}$ from a covariate distribution; then we generate circular responses $\{\Theta_j\}_{j=1}^n$ according to (1) with a regression function $m(x)$ we design and $\epsilon_j \sim$ von Mises$(0, 3)$; lastly, error-contaminated covariate data $\{W_j\}_{j=1}^n$ are generated according to (6) with U following a mean-zero normal distribution or Laplace distribution, and σ_u^2 set at a value to achieve a reliability ratio $\lambda \in \{0.8, 0.9\}$, where $\lambda = \text{Var}(X)/\text{Var}(W)$. In particular, two covariate distributions are considered, uniform$(-5, 5)$ and $N(0, 4)$; two regression functions are used, $m(x) = 2\text{atan}(x)$ and $m(x) = 2\text{atan}(1/x)$. The first regression function is simpler from the perspective of model fitting because it is smooth and monotone; the second regression function brings forth a more challenging model due to the singularity of $m(x)$ at $x = 0$ that creates a jump discontinuity.

At each simulation setting specified by $m(x)$, the covariate distribution, the value of n, and the distribution of U, we generate 100 data sets, based on each of which we implement estimation procedures determined by the choice of estimator and the bandwidth selection method. For each estimation procedure, we estimate $m(x)$ at a predetermined grid of values evenly spaced within the support of X, $x_1, \ldots, x_n$, over which $m(x)$ exhibits major features such as different degrees of curvature. The metric employed to assess the quality of estimation using $\hat{m}_{\cdot}(x)$ is the empirical mean cosine dissimilarity as an empirical risk function, $D_n(\hat{m}_{\cdot}(x); W, h) = 1 - n^{-1} \sum_{j=1}^n \cos(m(x_j) - \hat{m}_{\cdot}(x_j))$. The aforementioned optimal bandwidth h_{opt} minimizes $D_n(\hat{m}_{\cdot}(x); W, h)$.

Finally, we comment on some details imperative for the actual implementation of each estimation procedure. When computing the complex-valued weight $\mathcal{L}^*(W_j - x)$ in (10) for $\hat{m}_{\text{CE}}(x)$, we set $B^* = 250$, and we extract the real part of $\mathcal{L}^*(W_j - x)$ as the actual weight because, by construction, the imaginary part of $\mathcal{L}^*(W_j - x)$ has a mean of zero. We choose for all three proposed estimators the kernel $K(x) = 96\{x(x^2 - 15)\cos x + 3(5 - 2x^2)\sin x\}/(2\pi x^7)$, with $\phi_K(t) = (1 - t^2)^3 \mathbf{1}_{\{-1 \le t \le 1\}}$. Although this particular choice of the kernel is made to satisfy the more stringent **Condition S** to safeguard against ill-defined integral transforms in (8) and (11) when U is super smooth, computing these integrals with this $\phi_K(t)$ also tends to be more numerically stable when U is ordinary smooth. More specifically, we employ the fractional fast Fourier transform [1] to compute the integral transforms in (8) and (11) as described in detail in [22]. When implementing CV-SIMEX and CV-CE to select a bandwidth, we set $B = 30$.

6.2. Simulation results

Figure 1 provides boxplots of the empirical risk associated with each considered estimation procedure as n varies when $m(x) = 2\text{atan}(1/x)$, $X \sim N(0, 4)$, and $U \sim N(0, \sigma_u^2)$. It is evident from Figure 1 that all non-naive estimation procedures outperform naive estimation. Among the non-naive estimators when h_{opt} is used, $\hat{m}_{\text{CE}}(x)$ outperforms the rest in this more challenging model fitting exercise. This can be thanks to the (corrected) local linear weight used in $\hat{m}_{\text{CE}}(x)$ that is unbiased, in contrast to the other non-naive estimators with weights
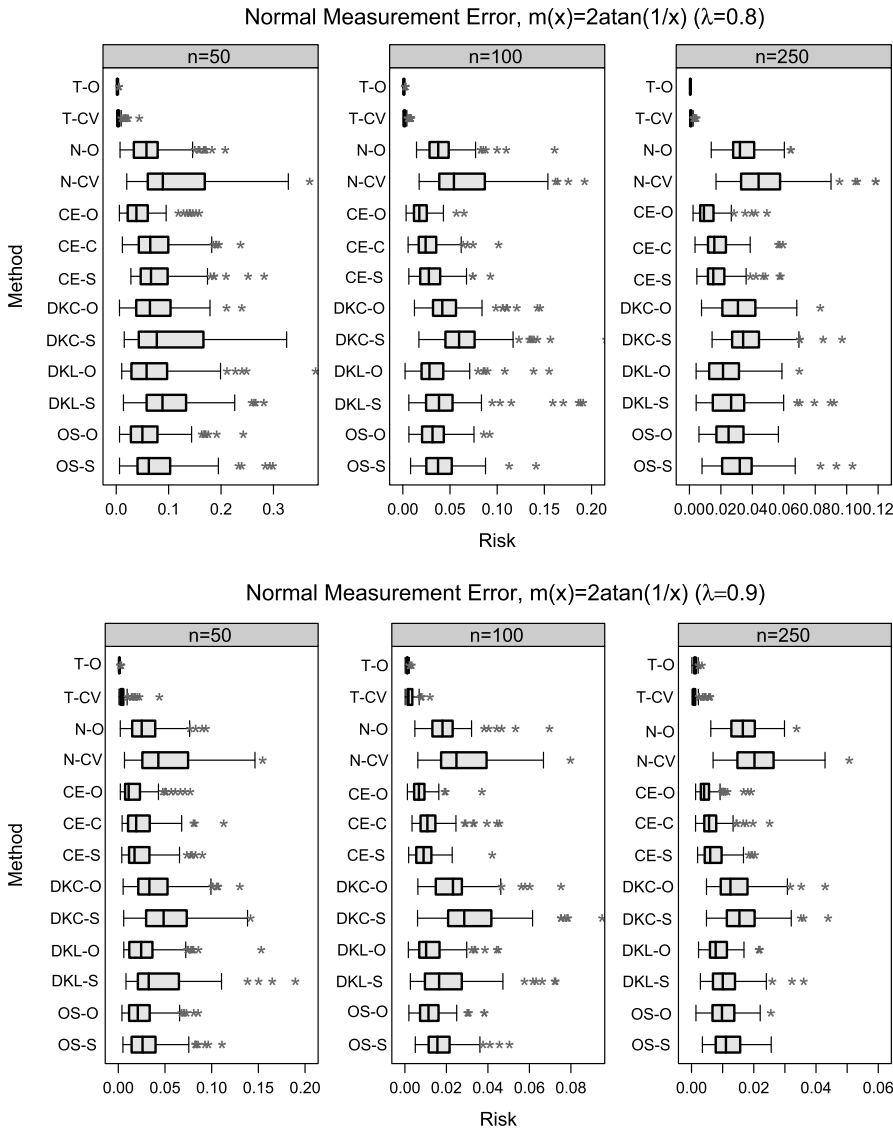
Fig 1. *Boxplots of empirical risk across 100 Monte Carlo replicates at each level of $n \in \{50, 100, 250\}$ when $X \sim N(0, 4)$, $m(x) = 2atan(1/x)$, and $U \sim N(0, \sigma_u^2)$ with $\lambda = 0.8$ (top panel) and 0.9 (bottom panel) for thirteen estimation procedures, each specified by the estimator and the bandwidth selection method that are linked by a hyphen. Labels for estimators are: T for the ideal estimator $\tilde{m}(x)$, N for the naive estimator $\hat{m}^*(x)$, CE for the complex error estimator $\hat{m}_{CE}(x)$, DKC for the deconvoluting kernel estimator with the local constant weight, DKL for our deconvoluting kernel estimator $\hat{m}_{DK}(x)$ (with the local linear weight), and OS for the one-step correction estimator $\hat{m}_{OS}(x)$. Labels for bandwidth selection methods are: O for the optimal bandwidth, CV for cross validation, C for CV-CE, and S for SIMEX.*

that are biased (although consistent) estimators for their error-free counterparts. Between the two types of deconvoluting kernel estimators, our estimator $\hat{m}_{\mathrm{DK}}(x)$ that uses the local linear weight improves over the estimator with the local constant weight. The two bandwidth selection methods, CV-SIMEX and CV-CE, for choosing a bandwidth in $\hat{m}_{\mathrm{CE}}(x)$ produce mostly comparable results. However, CV-CE substantially shortens the computation time. This

Table 1

Computation times (in seconds) different procedures take to estimate m(x) based on a sample of size
n ∈ {50, 100, 250}, including the time to choose a bandwidth h out of 50 candidates in H. The considered
procedures are naive estimation via $\hat{m}^*(x)$ paired with cross validation (N-CV), estimation via $\hat{m}_{CE}(x)$ paired
with CV-CE (CE-C), and with CV-SIMEX (CE-S), estimation via $\hat{m}_{DK}(x)$ using the local constant weight paired
with CV-SIMEX (DKC-S), via $\hat{m}_{DK}(x)$ using the local linear weight paired with CV-SIMEX (DKL-S), and
estimation via $\hat{m}_{OS}(x)$ paired CV-SIMEX (OS-S).

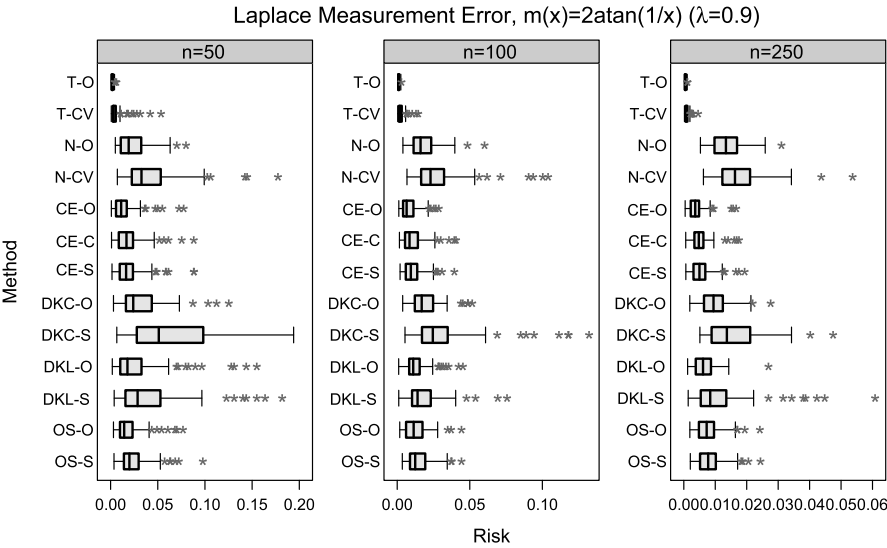| n	N-CV	CE-C	CE-S	DKC-S	DKL-S	OS-S
50	0.06	66.04	219	34.7	113	2470
100	0.16	130	625	127	371	4975
250	0.21	188	1250	192	433	12754

Fig 2. *Boxplots of empirical risk across 100 Monte Carlo replicates at each level of n ∈ {50, 100, 250} when*
X ~ N(0, 4), m(x) = 2atan(1/x), and U follows a mean-zero Laplace distribution with λ = 0.9 for thirteen
estimation procedures, each specified by the estimator and the bandwidth selection method that are linked by a
hyphen. Labels for estimators are: T for the ideal estimator $\tilde{m}(x)$, N for the naive estimator $\hat{m}^*(x)$, CE for the
complex error estimator $\hat{m}_{CE}(x)$, DKC for the deconvoluting kernel estimator with the local constant weight,
DKL for our deconvoluting kernel estimator $\hat{m}_{DK}(x)$ (with the local linear weight), and OS for the one-step
correction estimator $\hat{m}_{OS}(x)$. Labels for bandwidth selection methods are: O for the optimal bandwidth, CV for
cross validation, C for CV-CE, and S for SIMEX.

is evidenced in Table 1 that presents computation times for implementing different bandwidth selection methods, with $\mathcal{H}$ containing 50 candidate bandwidths, on a computer with an Intel Core i5 12600KF processor and a core speed of 4.5 GHz, followed by estimating $m(x)$ via various estimators based on one Monte Carlo replicate data set of size $n \in \{50, 100, 250\}$. Although time-consuming, CV-SIMEX substantially outperforms the naive CV as evidenced in a simulation study presented in Appendix I where we fitted the same regression model using $\hat{m}_{DK}(x)$ with bandwidths chosen by CV-SIMEX and the naive CV.

Figure 2 provides results under the same simulation setting as that for Figure 1, except that U follows a Laplace distribution and we focus on a reliability ratio of 0.9. Most observed patterns in Figure 1 in how different estimators compare are similarly observed here. Interestingly, even though the construction of $\hat{m}_{CE}(x)$ imposes the normality assumption on U, violating this
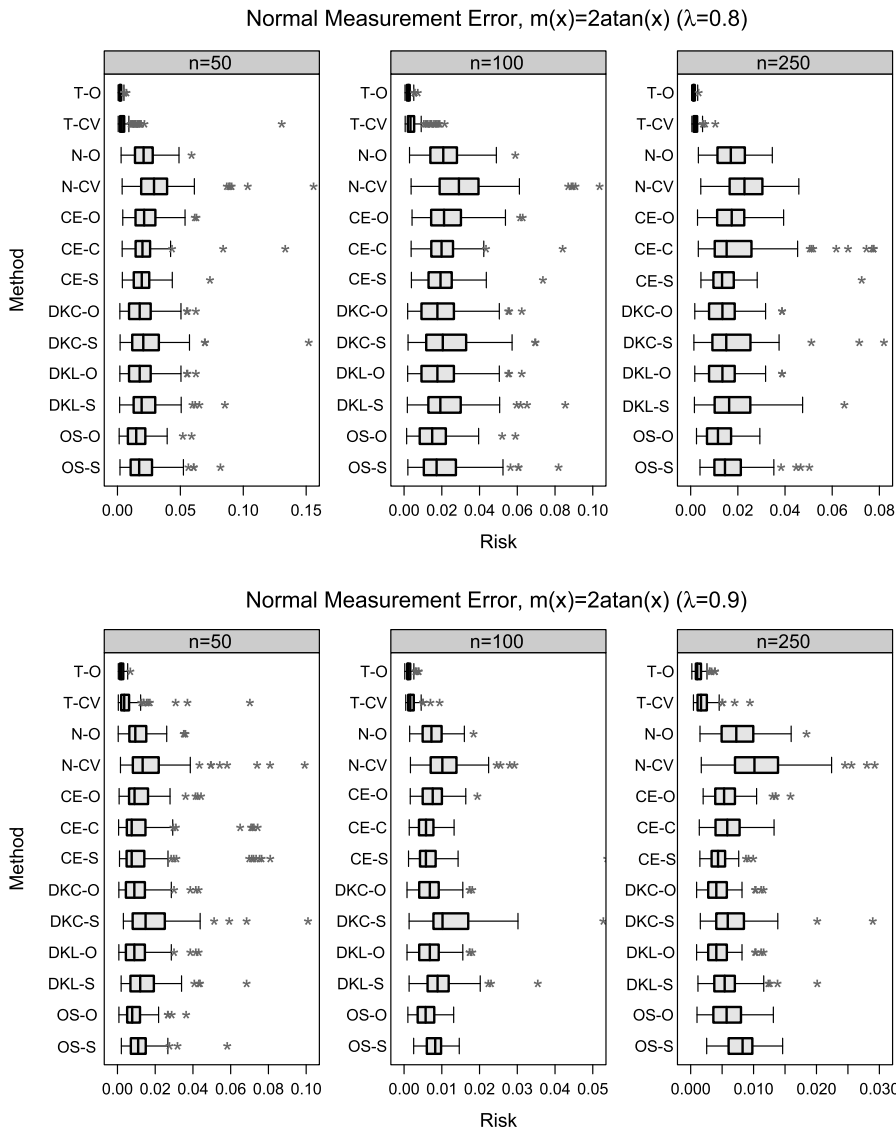
Fig 3. *Boxplots of empirical risk across 100 Monte Carlo replicates at each level of n ∈ {50, 100, 250} when X ~ Uniform(−5, 5), m(x) = 2atan(x), and U ~ N(0, 100(1 − λ)/12λ) with λ = 0.8 (top panel) and 0.9 (bottom panel) for thirteen estimation procedures, each specified by the estimator and the bandwidth selection method that are linked by a hyphen. Labels for estimators are: T for the ideal estimator m̃(x), N for the naive estimator m̂*(x), CE for the complex error estimator m̂ₒₑ(x), DKC for the deconvoluting kernel estimator with the local constant weight, DKL for our deconvoluting kernel estimator m̂₍DK₎(x) (with the local linear weight), and OS for the one-step correction estimator m̂₍OS₎(x). Labels for bandwidth selection methods are: O for the optimal bandwidth, CV for cross validation, C for CV-CE, and S for SIMEX.*

assumption does not noticeably compromise its performance, and it still tends to perform better than the other non-naive estimators.

Figure 3 repeats the simulation experiment summarized in Figure 1 using a simpler model with $m(x) = 2\arctan(x)$ and $X \sim$ uniform$(−5, 5)$. Now the non-naive estimations are more similar to each other than when fitting a more challenging model as in Figures 1 and 2. Figure 4 presents (in the top row) boxplots of fitted values for $m(x) = 2\text{atan}(x)$ based on the estimated
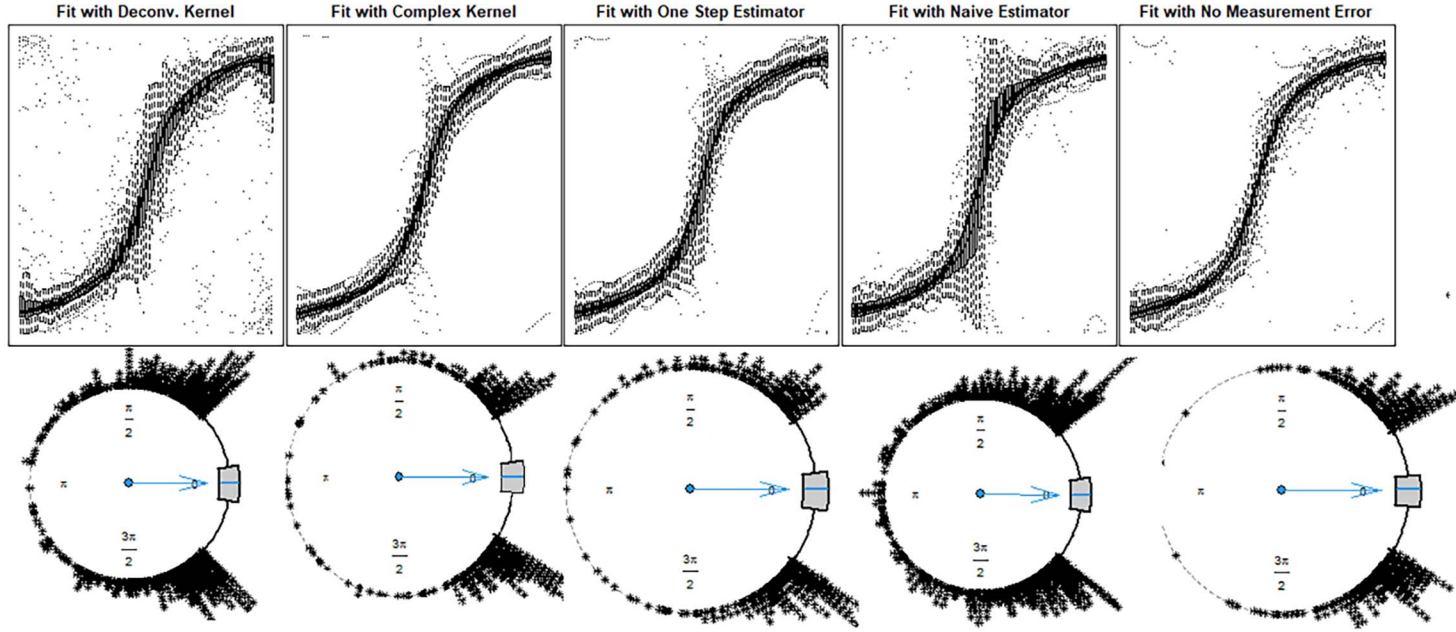
Fig 4. *Boxplots of each fitted point (top panels) and circular boxplots of errors (bottom panels) for each of four methods: $\hat{m}_{DK}(x)$, $\hat{m}_{CE}(x)$, $\hat{m}_{OS}(x)$, $\hat{m}^*(x)$, and $\tilde{m}(x)$, based on 100 Monte Carlo replicates when $X \sim uniform(-5, 5)$, $m(x) = 2atan(x)$. An error here is the fitted value minus the true value of $m(x)$.*

circular regression functions from various estimation procedures for a grid of x's, along with circular boxplots (in the bottom row) of the corresponding errors defined as the fitted value minus the truth of $m(x)$, when $n = 100$ and h_{SIMEX} is used for non-naive estimation. The former collection of boxplots shows that our proposed estimators substantially improve over $\hat{m}^*(x)$ in capturing the curvature of $m(x)$. The circular boxplots, on the other hand, reveal that many more errors from naive estimation (see the fourth panel in the bottom row of Figure 4), $\hat{m}^*(x) - m(x)$, deviate from zero and approach $\pm\pi$, whereas such drastic deviation from zero is much less observed for our proposed estimates (see the first three panels in the bottom row of Figure 4).

7. Application to wind data

7.1. Texas wind data

Now we turn to a real-life application and analyze wind direction data from the Texas Natural Resources Conservation Commission [42]. The data, available at https://doi.org/10.26023/ZQEZ-ENSF-T09, contains $n = 1756$ records of wind directions $\boldsymbol{\Theta}$ and times of day X collected by a single weather vein from May 20 to July 31 in 2003. The understanding of tendencies for wind direction relative to the time of day is crucial in meteorological studies in order to detect aberrations and predict large-scale weather events. For illustration purposes, we contaminate X with additive normal measurement errors to produce surrogate measures, W, such that an estimated reliability ratio, given by S_x^2/S_w^2, is equal to 0.9, where S_x^2 and S_w^2 are the sample variance of X and the sample variance of W, respectively.

We first fit the regression model in (1) using data $(\boldsymbol{\Theta}, X)$ to obtain the ideal estimate $\tilde{m}(x)$ with the bandwidth selected via the traditional 5-fold CV. Then we use data $(\boldsymbol{\Theta}, W)$ to carry out naive estimation to obtain $\hat{m}^*(x)$ using h_{naive}. Lastly, we compute three non-naive estimates, $\hat{m}_{\text{DK}}(x)$, $\hat{m}_{\text{CE}}(x)$, and $\hat{m}_{\text{OS}}(x)$, based on data $(\boldsymbol{\Theta}, W)$ using h_{SIMEX}. The resultant five estimated circular regression functions are plotted in Figure 5.

Contrasting the five estimated circular regression functions depicted in Figure 5, we see that the naive estimate $\hat{m}^*(x)$ has a diminished signal strength, flattening the relationship captured by the ideal estimate $\tilde{m}(x)$. Each of the proposed estimators improves on this by mimicking the pattern of $\tilde{m}(x)$ more closely, with $\hat{m}_{\text{CE}}(x)$ being the closest to $\tilde{m}(x)$ among the four estimates based on error-contaminated data. Like in the simulation study, we set $B = 100$ in the complex-valued weight in (10) when obtaining $\hat{m}_{\text{CE}}(x)$, which does display some instability, resulting in the estimated regression function less smooth than those from the other two proposed estimators. This may suggest that a larger number of Monte Carlo replicates B in the weight function for $\hat{m}_{\text{CE}}(x)$ is needed, which is consistent with our asymptotic analysis for this estimator (see Theorems 4.1 and 4.2), where it is assumed that B grows as n increases.

7.2. Santiago de Compostela wind data

The Galician Meteorological Agency of Spain collects climate data, available at the website of the Ministry for the Ecological Transition and the Demographic Challenge (https://www.miteco.gob.es/es.html). To more closely mirror our simulation settings, we downloaded $n = 100$ records of wind direction in radians from this website along with
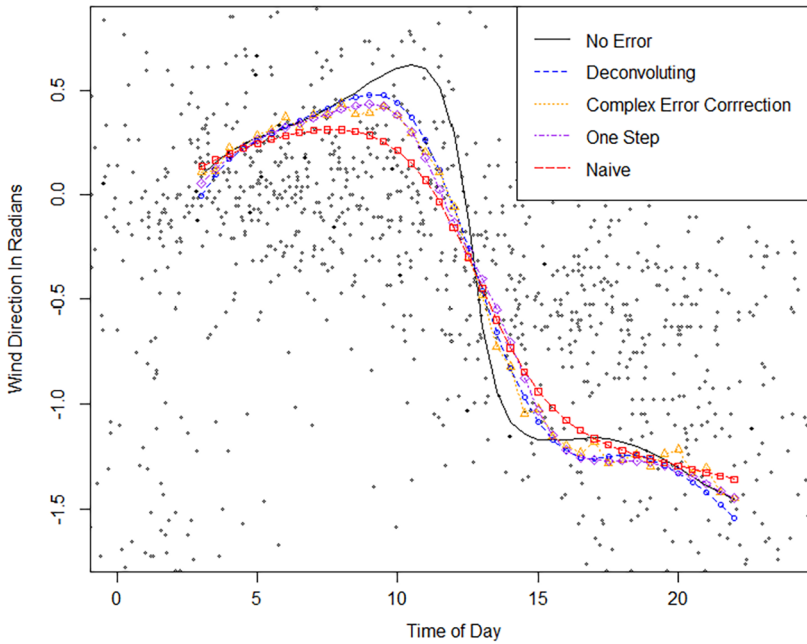
Fig 5. *Five estimated circular mean of the wind direction given time of day including the three proposed estimators,* $\hat{m}_{DK}(x)$ *(blue line running through circles* ○*),* $\hat{m}_{CE}(x)$ *(orange line running through triangles* △*), and* $\hat{m}_{OS}(x)$ *(purple line running through diamonds* ◇*), the naive estimate* $\hat{m}^*(x)$ *(red line passing through squares* □*), and the ideal estimate* $\tilde{m}(x)$ *(black line).*

measurements of gust wind speed above average in miles per hour from a weather vein in Santiago de Compostela, Spain. The quality of gust wind speed measurements depends on the response characteristics of anemometers and the effects of signal processing, besides other external factors that can introduce errors [34]. The goal of our analysis is to regress the wind direction (Θ) on the gust wind speed above average (X), while acknowledging that the measured gust wind speed (W) is contaminated by unknown measurement errors.

The proposed methods for estimating the regression function $m(x)$ remain applicable when the measurement error distribution is unknown but replicate measures of the true covariate are available. In particular, because the complex error estimator, $\hat{m}_{CE}(x)$, is formulated under the assumption of normal measurement error, it depends on the error distribution only via σ_u^2, which can be easily estimated based on replicates measurements [as in equation (4.3), 5]. The deconvoluting kernel estimator, $\hat{m}_{DK}(x)$, and the one-step correction estimator, $\hat{m}_{OS}(x)$, involve the measurement error characteristic function $\phi_U(\cdot)$. Under the assumption of a symmetric distribution, [10] provided an estimator for $\phi_U(\cdot)$ based on repeated measurements, whereas Comte and Kappus [6] proposed an estimator without assuming symmetry of U. These estimated characteristic functions can be used in place of $\phi_U(\cdot)$ in $\hat{m}_{DK}(x)$ and $\hat{m}_{OS}(x)$. The study of asymptotic properties of the resultant estimators for $m(x)$ is beyond the scope of this manuscript. We conjecture that, with an estimated $\phi_U(\cdot)$ plugged in, finite sample performance of $\hat{m}_{DK}(x)$ and $\hat{m}_{OS}(x)$ will deteriorate as reported in previous empirical study in a similar context with linear response data [22].

Without repeated measures of the gust wind speed at any given occasion when the wind direction was recorded in this application, we resort to a sensitivity analysis, by fitting the
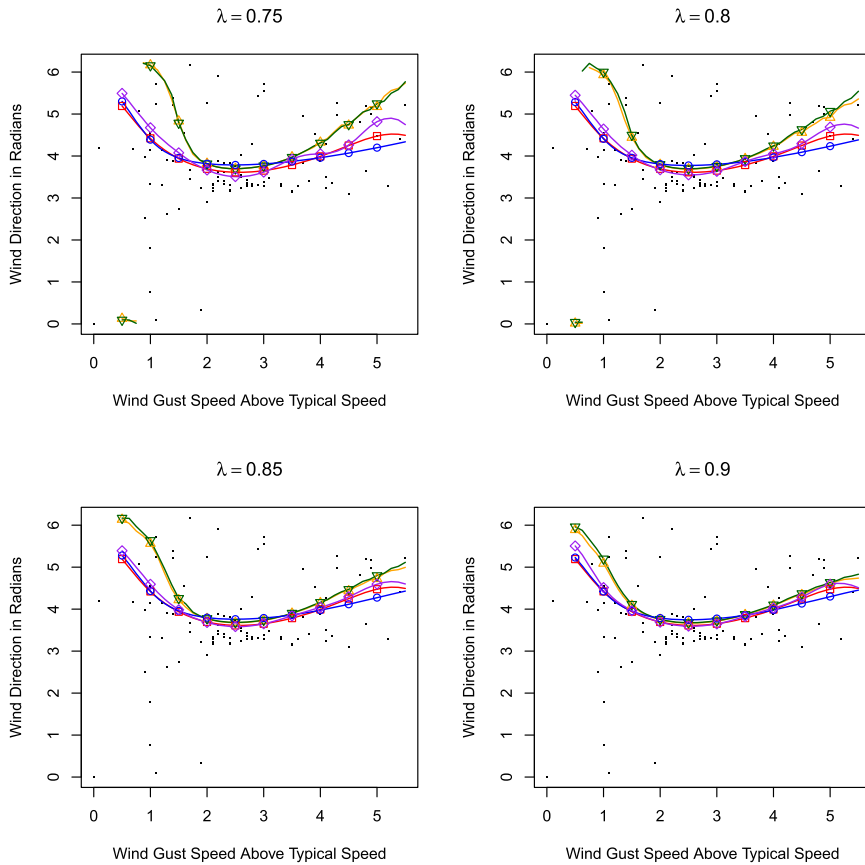
Fig 6. *Five estimated circular mean of the wind direction given the gust wind speed above average including the three proposed estimators assuming normal measurement error, $\hat{m}_{DK}(x)$ (blue line running through circles $\circ$), $\hat{m}_{CE}(x)$ using the bandwidth chosen by CV-SIMEX (orange line running through upright triangles $\triangle$), $\hat{m}_{CE}(x)$ using the bandwidth chosen by CV-CE (green line running through down facing triangles ∇) and $\hat{m}_{OS}(x)$ (purple line running through diamonds $\diamond$), and the naive estimate $\hat{m}^*(x)$ (red line passing through squares $\square$).*

nonparametric regression model under different assumed levels of error contamination corresponding to four levels of reliability ratio, $\lambda \in \{0.75, 0.8, 0.85, 0.9\}$, and two assumed error distributions: normal and Laplace errors. Figure 6 shows the estimated regression functions using our proposed methods, contrasting with the naive estimate when normal measurement errors are assumed. Besides using CV-SIMEX to select bandwidths in our estimates, we also repeated the complex error estimation using CV-CE to select the bandwidth in $\hat{m}_{CE}(x)$. The two variants of $\hat{m}_{CE}(x)$ are similar to each other but are more distinct from the other estimates, especially when a more severe error contamination is assumed. It is at the lower levels of (assumed) λ where one witnesses the non-naive estimates differ from the naive one more noticeably. The abrupt jump of $\hat{m}_{CE}(x)$ near the lower bound of the covariate when $\lambda = 0.75, 0.8$ should not be interpreted as a phenomenon of discontinuity because the response is circular. With this in mind, $\hat{m}_{CE}(x)$ appears to be more sensitive to what one assumes for the level of error contamination. The other two estimates, $\hat{m}_{DK}(x)$ and $\hat{m}_{OS}(x)$, are closer to the naive estimate $\hat{m}^*(x)$ in comparison, especially at a higher assumed value of λ, although they both exhibit slightly more curvature than $\hat{m}^*(x)$.
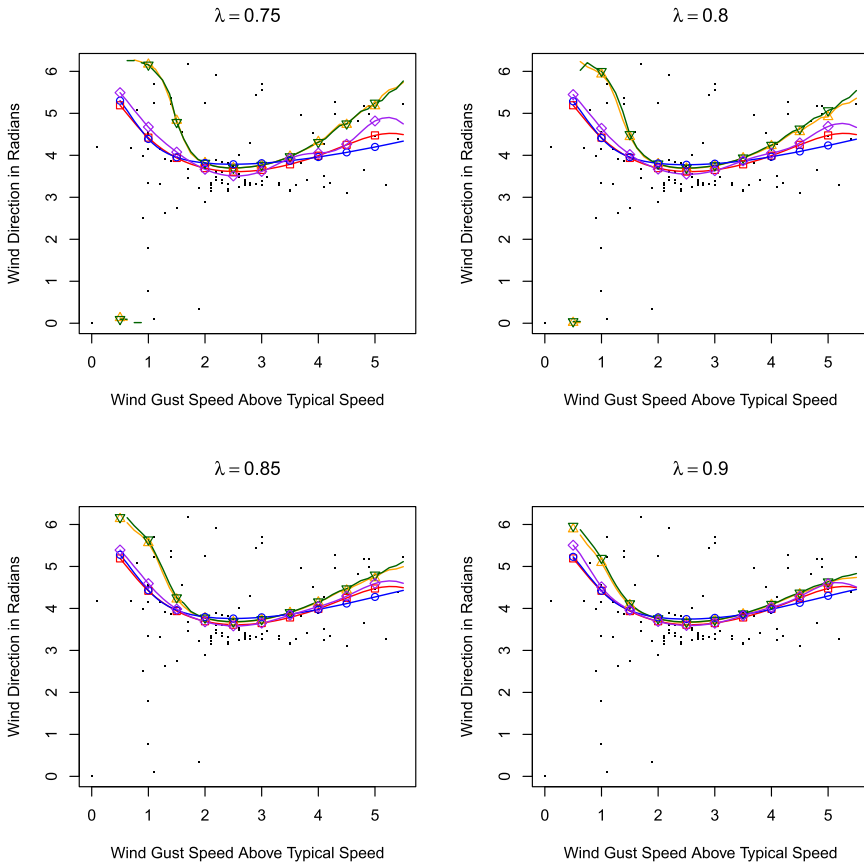
Fig 7. *Five estimated circular mean of the wind direction given the gust wind speed above average including the three proposed estimators assuming Laplace measurement error, $\hat{m}_{DK}(x)$ (blue line running through circles ○), $\hat{m}_{CE}(x)$ using the bandwidth chose by CV-SIMEX (orange line running through upright triangles △), $\hat{m}_{CE}(x)$ using the bandwidth chosen by CV-CE (green line running through down facing triangles ▽) and $\hat{m}_{OS}(x)$ (purple line running through diamonds ◇), and the naive estimate $\hat{m}^*(x)$ (red line passing through squares □).*

Figure 7 provides the counterpart estimates when assuming Laplace measurement errors, which are mostly similar to those in Figure 6. The sensitivity analysis thus suggests that the proposed estimators are more reliant on the assumed level of error contamination and relatively robust to the assumed error distribution. Although one cannot conclude which estimated regression function is closest to the (nonexistent) ideal estimate $\tilde{m}(x)$ or the (unknown) truth $m(x)$ in this application, some lessons can be learnt from this exercise. First, unless one has strong data or scientific evidence indicating non-Gaussian measurement error, we recommend using the complex error estimator $\hat{m}_{CE}(x)$ that is most computationally convenient to obtain, especially when paired with CV-CE bandwidth selection. Second, when confronted with serious concern or doubt regarding the normality assumption for measurement error, the one-step correction estimator $\hat{m}_{OS}(x)$ is the next in line as a contender that tends to be numerically more stable than the deconvoluting kernel estimator $\hat{m}_{DK}(x)$. Third, even when using $\hat{m}_{OS}(x)$ or $\hat{m}_{DK}(x)$, one may consider using the Laplace characteristic function for $\phi_U(\cdot)$ when computing the estimator for a smoother implementation of the deconvoluting operator.

8. Discussion

In this study, we develop three methods to extend local polynomial regression of a linear response to local polynomial regression of a circular response while accounting for measurement error in a linear covariate. An R package NPMEDD for implementing these methods is available to download at https://github.com/nwwoolsey/NPMEDD, where the data and the R code for reproducing results in Section 7 are also available. The proposed methods lead to novel estimators for the circular regression function that use three different strategies to achieve the common goal of correcting naive estimators for covariate measurement error. A common thread running through the novel estimators is a certain form of integral transforms relating to a deconvoluting operation. Besides adapting CV-SIMEX to the circular regression model to select bandwidths in the novel estimators, we propose the CV-CE procedure to enhance the computational efficiency and feasibility of bandwidth selection in the presence of measurement error. Moreover, we thoroughly study the asymptotic properties of the proposed estimators, and establish consistency and asymptotic normality under suitable conditions. Our theoretical investigation provides a generic recipe for studying nonparametric estimators of a circular regression function with error-in-covariate.

Generalization of the proposed methods to incorporate multiple covariates with some prone to measurement error can be realized by using the multivariate generalization for deconvoluting kernels as outlined in Yi et al. [45, Section 10.7.4]. In the real-life application in Section 7.1, the time of day can be transformed into a directional covariate as the time on the 24-hour clock, and the proposed methods can be revised by the simple adaptation of a directional kernel as opposed to a linear kernel as in [12] and [11]. Bandwidth selection in the presence of covariate measurement error remains a challenging task. Although CV-CE saves a tremendous amount of computation time compared to CV-SIMEX, we acknowledge room for improvement for CV-CE by improving the estimated loss function. Recently, in the context of local polynomial regression of a linear response with error-in-covariate, [13] proposed bandwidth selection procedures that also strive for estimating a loss function, which is the mean squared prediction error, by accounting for measurement error in the estimated regression function and also the estimated density of error-prone covariates. Our preliminary experiments on their methods suggest an increased computational intensity and instability, without noticeable gain in the final estimation quality. Lastly, we assume fully known measurement error distribution throughout the study to avoid identifiability issues for measurement error models. In practice, one would have to count on replicate data or external/validation data to estimate the error distribution, or to conduct sensitivity analysis if such data are unavailable to estimate σ_u^2 or $\phi_U(t)$.

Appendix A: Proof of Lemma 3.1

Proof. By the multivariate version of Taylor's theorem, one has a Taylor expansion of $g(t+V)$ around $V = \mathbf{0}$ given by

$$g(t + V) = \sum_{|\ell|=0}^{\infty} \frac{D^{\ell} g(t)}{\ell!} V^{\ell},$$

where $\boldsymbol{\ell}! = \ell_1! \ldots \ell_n!$ and $D^{\boldsymbol{\ell}} g(\boldsymbol{t}) = (\partial^{\ell_1} \ldots \partial^{\ell_n} / \partial t_1^{\ell_1} \ldots \partial t_n^{\ell_n}) g(\boldsymbol{t})$. It follows that

$$\mathbb{E}\{g(\boldsymbol{t} + \boldsymbol{V})\} = g(\boldsymbol{t}) + \sum_{|\boldsymbol{\ell}|>0}^{\infty} \frac{D^{\boldsymbol{\ell}} g(\boldsymbol{t})}{\boldsymbol{\ell}!} \mathbb{E}(\boldsymbol{V}^{\boldsymbol{\ell}}),$$

which reduces to $g(\boldsymbol{t})$ if $\mathbb{E}(\boldsymbol{V}^{\boldsymbol{\ell}}) = 0$, $\forall \boldsymbol{\ell}$ with $|\boldsymbol{\ell}| > 0$. This completes the proof. □

Appendix B: Proof of Lemma 3.2

The proof is similar to that given in Delaigle [7, Appendix A.2].

Proof. Define $Y = \sin \Theta$. Because $U \perp X$, we have $Y \perp W|X$, thus

$$
\begin{aligned}
E(Y|W = w) &= \int_{-1}^{1} y f_{Y|W}(y|w) \, dy \\
&= \int_{-1}^{1} y \cdot \frac{f_{Y,W}(y, w)}{f_W(w)} \, dy \\
&= \frac{1}{f_W(w)} \int_{-1}^{1} y \int_{-\infty}^{\infty} f_{Y,W|X}(y, w|x) f_X(x) \, dx \, dy \\
&= \frac{1}{f_W(w)} \int_{-\infty}^{\infty} \int_{-1}^{1} y f_{Y|X}(y|x) f_{W|X}(w|x) \, dy \, f_X(x) \, dx \\
&= \frac{1}{f_W(w)} \int_{-\infty}^{\infty} \int_{-1}^{1} y f_{Y|X}(y|x) \, dy \, f_X(x) f_U(w - x) \, dx \\
&= \frac{1}{f_W(w)} \int_{-\infty}^{\infty} m_1(x) f_X(x) f_U(w - x) \, dx.
\end{aligned}
$$

Hence,

$$m_1^*(w) f_W(w) = \int_{-\infty}^{\infty} m_1(x) f_X(x) f_U(w - x) \, dx. \tag{B.1}$$

Applying the Fourier transform on both sides of (B.1) yields

$$
\begin{aligned}
\phi_{m_1^* f_W}(t) &= \int_{-\infty}^{\infty} e^{itw} \int_{-\infty}^{\infty} m_1(x) f_X(x) f_U(w - x) \, dx \, dw \\
&= \int_{-\infty}^{\infty} m_1(x) f_X(x) e^{itx} \int_{-\infty}^{\infty} e^{it(w-x)} f_U(w - x) \, dw \, dx \\
&= \int_{-\infty}^{\infty} m_1(x) f_X(x) e^{itx} \int_{-\infty}^{\infty} e^{itu} f_U(u) \, du \, dx \\
&= \phi_U(t) \int_{-\infty}^{\infty} m_k(x) f_X(x) e^{itx} \, dx \\
&= \phi_{m_1 f_X}(t) \phi_U(t),
\end{aligned}
$$

where $\phi_{m_1^* f_W}(t)$ is the Fourier transform of $m_1^*(w) f_W(w)$, and $\phi_{m_1 f_X}(t)$ is the Fourier transform of $m_1(x) f_X(x)$. Finally, applying the inverse Fourier transform on both sides of $\phi_{m_1 f_X}(t) = \phi_{m_1^* f_W}(t) / \phi_U(t)$ gives

$$m_1(x) f_X(x) = \frac{1}{2\pi} \int e^{-itx} \frac{\phi_{m_1^* f_W}(t)}{\phi_U(t)} \, dt.$$

A similar result can be derived for $m_2^*(w) = \mathbb{E}(\cos\Theta|W = w)$. This completes the proof of Lemma 3.2. □

Even though we consider $Y = \sin\Theta$ and $Y = \cos\Theta$ in this proof, the arguments leading up to (B.1) hold for any linear random variable Y as long as all relevant integrals are well-defined. This point is important for later development, e.g, in Appendix D.4, where we consider Y to be other linear random variables. To recap, provided that all relevant integrals (including expectations) are well-defined and assuming a classical additive measurement error model relating W and X, then we have, for any linear random variable Y, that

$$\mathbb{E}(Y|W = w)f_W(w) = \int \mathbb{E}(Y|X = x)f_X(x)f_U(w - x)\,dx,$$

or, equivalently,

$$\mathbb{E}\{\mathbb{E}(Y|X)f_U(w - X)\} = \mathbb{E}(Y|W = w)f_W(w). \tag{B.2}$$

Appendix C: Notations and important results

C.1. Notations

For easy reference, we provide a list of notations next.

- For $\ell = 1, 2$, $\tau_\ell(x) = m_\ell(x)f_X(x)$, $\tau_\ell^*(x) = m_\ell^*(x)f_W(x)$, where $m_1(x) = \mathbb{E}(\sin\Theta|X = x)$, $m_2(x) = \mathbb{E}(\cos\Theta|X = x)$, $m_1^*(x) = \mathbb{E}(\sin\Theta|W = x)$, and $m_2^*(x) = \mathbb{E}(\cos\Theta|W = x)$.
- $\xi_1(x) = \mathbb{E}\{(\sin\Theta)^2|X = x\}$, $\xi_2(x) = \mathbb{E}\{(\cos\Theta)^2|X = x\}$,
- $\psi(x) = \mathbb{E}\{(\sin\Theta)(\cos\Theta)|X = x\}$.
- For $\ell \in \mathbb{N}_0$, $\mu_\ell = \int t^\ell K(t)dt$, $\nu_\ell = \int t^\ell K^2(t)dt$.
- For $\ell \in \mathbb{N}_0$, $K_{U,\ell,h}(x) = K_{U,\ell}(x/h)/h$, where

$$K_{U,\ell}(x) = x^\ell L_\ell(x) = i^{-\ell}\frac{1}{2\pi}\int e^{-itx}\frac{\phi_K^{(\ell)}(t)}{\phi_U(-t/h)}dt, \tag{C.1}$$

in which $L_\ell(x)$ is given in Equation (8).

C.2. Important lemmas

Several results that we heavily rely on in our asymptotic analyses are listed below. Some of these results are established in existing literature, while others generalize existing results.

Lemma C.1. *If a sequence of Borel functions $\{G_n(t)\}$ which satisfy the limit $\lim_{n\to\infty} G_n(t) = G(t)$ and $\sup_n |G_n(t)| \leq G^*(t)$, where $G^*(t)$ is an integrable function satisfying $\lim_{t\to\infty} |G^*(t)| = 0$, then*

$$\lim_{n\to\infty}\int G_n(t)f(x - ht)dt = f(x)\int G(t)dt,$$

where $h \to 0$ as $n \to \infty$, and x is a continuity point of function $f(\cdot)$.

This is Lemma 2.1 in [14].

Lemma C.2. *If U is ordinary smooth of order β with a non-vanishing characteristic function $\phi_U(t)$, $\|\phi_K^{(\ell)}(t)\|_\infty < \infty$, and $\int |t|^\beta \phi_K^{(\ell)}(t)dt < \infty$, then*

$$\lim_{n\to\infty} h^\beta K_{U,\ell}(x) = \frac{i^{-\ell}}{2\pi c} \int e^{-itx} |t|^\beta \phi_K^{(\ell)}(t)dt,$$

where $c = \lim_{t\to\infty} t^\beta \phi_U(t)$.

This is Lemma B.3 in [8].

Lemma C.3. *For $\boldsymbol{\ell} = (\ell_1, \ldots, \ell_Q)^\top \in \mathbb{N}_0^Q$ with $Q \geq 2$, suppose the kernel $K(\cdot)$ satisfies $\|\phi_K^{(\ell_q)}(t)\|_\infty < \infty$, $\|\phi_K^{(\ell_q+1)}(t)\|_\infty < \infty$, $\int |t|^\beta \phi_K^{(\ell_q)}(t)dt < \infty$, and $\int (|t|^\beta + |t|^{\beta-1})\{|\phi_K^{(\ell_q)}(t)| + |\phi_K^{(\ell_q+1)}(t)|\}dt < \infty$, for $q = 1, \ldots, Q$, also, U is ordinary smooth of order β with $\phi_U(t) \neq 0$, for all t, and $\|\phi'_U(t)\|_\infty < \infty$, then*

$$\lim_{n\to\infty} h^{Q\beta} \int \left\{ \prod_{q=1}^{Q} K_{U,\ell_q}(v) \right\} f(x - hv)dv = f(x)\eta(\ell_1, \ldots, \ell_Q),$$

where $h \to 0$ as $n \to \infty$, x is a continuity point of a bounded function $f(\cdot)$, and

$$\eta(\ell_1, \ldots, \ell_Q)$$
$$= \frac{i^{-|\boldsymbol{\ell}|}}{c^Q (2\pi)^{Q-1}} \int \cdots \int \left| \left(\prod_{q=1}^{Q-1} t_q \right) \left(\sum_{q=1}^{Q-1} t_q \right) \right|^\beta \left\{ \prod_{q=1}^{Q-1} \phi_K^{(\ell_q)}(t_q) \right\} \times$$
$$\phi_K^{(\ell_Q)}\left(-\sum_{q=1}^{Q-1} t_q \right) dt_1 \ldots dt_{Q-1},$$

where $c = \lim_{t\to\infty} t^\beta \phi_U(t)$.

This is a generalization of Lemma B.4 in [8] that we prove next.

Proof. Firstly, by Lemma C.2,

$$\lim_{n\to\infty} h^{Q\beta} \prod_{q=1}^{Q} K_{U,\ell_q}(v) = i^{-\sum_{q=1}^{Q} \ell_q} \frac{1}{(2\pi c)^Q} \prod_{q=1}^{Q} \int e^{-itv} |t|^\beta \phi_K^{(\ell_q)}(t)dt.$$

Secondly, by Lemma C.1,

$$\lim_{n\to\infty} h^{Q\beta} \int \left\{ \prod_{q=1}^{Q} K_{U,\ell_q}(v) \right\} f(x - hv) \, dv$$

$$= f(x) \times i^{-\sum_{q=1}^{Q} \ell_q} \frac{1}{(2\pi c)^Q} \int \prod_{q=1}^{Q} \int e^{-its} |t|^\beta \phi_K^{(\ell_q)}(t) \, dt \, ds$$

$$= i^{-\sum_{q=1}^{Q} \ell_q} \frac{f(x)}{(2\pi c)^Q} \int \int \cdots \int e^{-it_1 s} |t_1|^\beta \phi_K^{(\ell_1)}(t_1) \ldots e^{-it_Q s} |t_Q|^\beta \times$$
$$\phi_K^{(\ell_Q)}(t_Q) \, dt_1 \ldots dt_Q \, ds$$

$$= \frac{f(x)}{c^Q} \frac{i^{-\sum_{q=1}^{Q} \ell_q}}{(2\pi)^{Q-1}} \int \cdots \int \prod_{q=1}^{Q-1} \left\{ |t_q|^\beta \phi_K^{(\ell_q)}(t_q) \right\} \times$$

$$\frac{1}{2\pi} \int \int e^{-is(t_Q + \sum_{q=1}^{Q-1} t_q)} |t_Q|^\beta \phi_K^{(\ell_Q)}(t_Q) \, dt_Q \, ds \, dt_1 \ldots d_{t_{Q-1}}. \tag{C.2}$$

Thirdly, the inner double integral in (C.2) can be simplified using the Fourier integral theorem. Recall that the Fourier integral theorem [27, Section 3.2.1] states that, if a function $g(v)$ satisfies the Dirichlet conditions in every finite interval centering at zero, and $\int |g(v)| dv < \infty$, then

$$g(v) = \frac{1}{2\pi} \int \int e^{-is(t-v)} g(t) dt ds.$$

By the Fourier integral theorem,

$$\frac{1}{2\pi} \int \int e^{-is(t_Q + \sum_{q=1}^{Q-1} t_q)} |t_Q|^\beta \phi_K^{(\ell_Q)}(t_Q) \, dt_Q \, ds = \left| \sum_{q=1}^{Q-1} t_q \right|^\beta \phi_K^{(\ell_Q)} \left(-\sum_{q=1}^{Q-1} t_q \right).$$

Using this result in (C.2) gives the result stated in Lemma C.3. □

Lemma C.4. *For $(\ell_1, \ldots, \ell_Q) \in \mathbb{N}_0^Q$ with $Q \geq 2$, suppose that $\phi_K(t)$ is supported on $[-1, 1]$, $\|\phi_K^{(\ell_q)}(t)\|_\infty < \infty$, for $q = 1, \ldots, Q$, and U is super smooth of order β with $\phi_U(t) \neq 0$, for all t, then*

$$\left| \int \prod_{q=1}^{Q} K_{U,\ell_q}(v) dv \right| \leq C h^{Q\beta_2} \exp\left(Q h^{-\beta} / \gamma \right),$$

where C is some positive constant and $\beta_2 = \beta_0 I(\beta_0 < 0.5)$.

This is a generalization of Lemma B.9 in [8], and can be proved using Lemma B.8 in [8].

Appendix D: Asymptotic properties of the deconvoluting kernel estimator $\hat{m}_{DK}(x)$

D.1. The mean of the local linear weight $\mathcal{L}(W_j - x)$

Using (C.1), we re-express the new weight $\mathcal{L}(W_j - x)$ in Equation (7) as

$$\mathcal{L}(W_j - x)$$
$$= \frac{1}{n} K_{U,0,h}(W_j - x) \sum_{k=1}^{n} K_{U,2,h}(W_k - x) - \frac{1}{n} K_{U,1,h}(W_j - x) \sum_{k=1}^{n} K_{U,1,h}(W_k - x).$$

The result in Equation (9) is equivalent to

$$E\{K_{U,\ell,h}(W - x)|X\} = \left(\frac{X - x}{h}\right)^\ell K_h(X_j - x), \text{ for } \ell = 0, 1, 2. \tag{D.1}$$

Using this result and the new expression of the weight given above, we have

$$\mathbb{E}\{\mathcal{L}(W_j - x)|X\}$$
$$= \frac{1}{n} \mathbb{E} \left\{ K_{U,0,h}(W_j - x) K_{U,2,h}(W_j - x) + K_{U,0,h}(W_j - x) \sum_{k \neq j} K_{U,2,h}(W_k - x) \right.$$

$$-K_{U,1,h}^2(W_j - x) - K_{U,1,h}(W_j - x) \sum_{k \neq j} K_{U,1,h}(W_k - x) \bigg| X \bigg\}$$

$$= \frac{1}{n} \mathbb{E} \left\{ K_{U,0,h}(W_j - x) K_{U,2,h}(W_j - x) - K_{U,1,h}^2(W_j - x) \bigg| X_j \right\} +$$

$$\frac{1}{n} \left\{ K_h(X_j - x) \sum_{k \neq j} \left(\frac{X_k - x}{h} \right)^2 K_h(X_k - x) - \right.$$

$$\left. \left(\frac{X_j - x}{h} \right) K_h(X_j - x) \sum_{k \neq j} \left(\frac{X_k - x}{h} \right) K_h(X_k - x) \right\}$$

$$= \frac{1}{n} \mathbb{E} \left\{ K_{U,0,h}(W_j - x) K_{U,2,h}(W_j - x) \big| X_j \right\} - \frac{1}{n} \mathbb{E} \left\{ K_{U,1,h}^2(W_j - x) \big| X_j \right\} + \quad \text{(D.2)}$$

$$\frac{1}{n} \left\{ K_h(X_j - x) \sum_{k=1}^n \left(\frac{X_k - x}{h} \right)^2 K_h(X_k - x) - \left(\frac{X_j - x}{h} \right) K_h(X_j - x) \times \quad \text{(D.3)} \right.$$

$$\left. \sum_{k=1}^n \left(\frac{X_k - x}{h} \right) K_h(X_k - x) \right\},$$

where (D.3) is the original local linear weight $\mathcal{W}(X_j - x)$. To derive the two expectations in (D.2) and other similar expectations arising in later derivations in this appendix, we derive a more general result next.

For $\boldsymbol{\ell} = (\ell_1, \ldots, \ell_Q)^\top \in \mathbb{N}_0^Q$ with $Q \geq 2$, by Lemmas C.3 and C.4,

$$\mathbb{E} \left\{ \prod_{q=1}^Q K_{U,\ell_q,h}(W - x) \bigg| X \right\}$$

$$= \int \left\{ \prod_{q=1}^Q K_{U,\ell_q,h}(w - x) \right\} f_U(w - X) \, dw, \text{ next let } v = (w - x)/h,$$

$$= \frac{1}{h^{Q-1}} \int \left\{ \prod_{q=1}^Q K_{U,\ell_q}(v) \right\} f_U(x + hv - X) \, dv$$

$$= \frac{1}{h^{Q-1}} \int \left\{ \prod_{q=1}^Q K_{U,\ell_q}(v) \right\} f_U(X - x - hv) \, dv, \text{ assuming } f_U(-u) = f_U(u),$$

$$= \begin{cases} \dfrac{\eta(\ell_1, \ldots, \ell_Q)}{h^{Q-1+Q\beta}} f_U(X - x) + o_P \left(\dfrac{1}{h^{Q-1+Q\beta}} \right), & \text{if } U \text{ is ordinary smooth,} \\ O_P \left(\dfrac{\exp(Q h^{-\beta}/\gamma)}{h^{Q-1-Q\beta_2}} \right), & \text{if } U \text{ is super smooth.} \end{cases} \quad \text{(D.4)}$$

Using (D.4) with $\boldsymbol{\ell} = (0, 2)^\top$, we have, for the first expectation in (D.2),

$$\mathbb{E} \left\{ K_{U,0,h}(W_j - x) K_{U,2,h}(W_j - x) \big| X_j \right\}$$

$$= \begin{cases} \dfrac{\eta(0,2)}{h^{1+2\beta}} f_U(X_j - x) + o_p\left(\dfrac{1}{h^{1+2\beta}}\right), & \text{if } U \text{ is ordinary smooth of order } \beta, \\[2ex] O_p\left(\dfrac{\exp(2h^{-\beta}/\gamma)}{h^{1-2\beta_2}}\right), & \text{if } U \text{ is super smooth of order } \beta; \end{cases}$$

similarly, the second expectation in (D.2) is, with $\boldsymbol{\ell} = (1,1)^\top$ in (D.4),

$$\mathbb{E}\left\{ K_{U,1,h}^2(W_j - x) \,\middle|\, X_j \right\}$$

$$= \begin{cases} \dfrac{\eta(1,1)}{h^{1+2\beta}} f_U(X_j - x) + o_p\left(\dfrac{1}{h^{1+2\beta}}\right), & \text{if } U \text{ is ordinary smooth of order } \beta, \\[2ex] O_p\left(\dfrac{\exp(2h^{-\beta}/\gamma)}{h^{1-2\beta_2}}\right), & \text{if } U \text{ is super smooth of order } \beta. \end{cases}$$

It follows that

$$\mathbb{E}\{\mathcal{L}(W_j - x)|\boldsymbol{X}\}$$

$$= \begin{cases} \mathcal{W}(X_j - x) + \dfrac{\eta(0,2) - \eta(1,1)}{nh^{1+2\beta}} f_U(X_j - x) + \\[1ex] o_p\left(\dfrac{1}{nh^{1+2\beta}}\right), & \text{if } U \text{ is ordinary smooth,} \\[2ex] \mathcal{W}(X_j - x) + O_p\left(\dfrac{\exp(2h^{-\beta}/\gamma)}{nh^{1-2\beta_2}}\right), & \text{if } U \text{ is super smooth.} \end{cases} \tag{D.5}$$

Hence, the new weight based on $\boldsymbol{W}$, $\mathcal{L}(W_j - x)$, is a consistent estimator of the local linear weight based on $\boldsymbol{X}$, $\mathcal{W}(X_j - x)$, under suitable conditions. In particular, if U is ordinary smooth of order β, then the condition is $nh^{1+2\beta} \to \infty$, as $n \to \infty$ and $h \to 0$; if U is super smooth of order β, then the condition is $nh^{1-2\beta_2} \exp(-2h^{-\beta}/\gamma) \to \infty$, as $n \to \infty$ and $h \to 0$. We impose these conditions throughout this appendix.

D.2. The mean of $\mathcal{L}^2(W_j - x)$

To derive the asymptotic variance of $\hat{g}_{\ell,\mathrm{DK}}(x)$, for $\ell = 1, 2$, we derive the mean of $\mathcal{L}^2(W_j - x)$ given $\boldsymbol{X}$,

$$\mathbb{E}\left\{ \mathcal{L}^2(W_j - x) \,\middle|\, \boldsymbol{X} \right\} = \frac{1}{n^2}(I - 2II + III),$$

where

$$I = \left\{ K_{U,0,h}(W_j - x) \sum_{k=1}^n K_{U,2,h}(W_k - x) \right\}^2,$$

$$II = \left\{ K_{U,0,h}(W_j - x) \sum_{k=1}^n K_{U,2,h}(W_k - x) \right\} \times$$

$$\left\{ K_{U,1,h}(W_j - x) \sum_{\ell=1}^{n} K_{U,1,h}(W_\ell - x) \right\},$$

$$III = \left\{ K_{U,1,h}(W_j - x) \sum_{k=1}^{n} K_{U,1,h}(W_k - x) \right\}^2.$$

First,

$$\mathbb{E}(I|\boldsymbol{X})$$
$$= \mathbb{E} \left\{ K_{U,0,h}^2(W_j - x) K_{U,2,h}^2(W_j - x) \Big| X_j \right\}$$
$$+ \mathbb{E} \left\{ K_{U,0,h}^2(W_j - x) \Big| X_j \right\} \sum_{k \neq j} \mathbb{E} \left\{ K_{U,2,h}^2(W_k - x) \Big| X_k \right\} +$$
$$2\mathbb{E} \left\{ K_{U,0,h}^2(W_j - x) K_{U,2,h}(W_j - x) \Big| X_j \right\} \sum_{k \neq j} \mathbb{E} \left\{ K_{U,2,h}(W_k - x) \Big| X_k \right\} +$$
$$\mathbb{E} \left\{ K_{U,0,h}^2(W_j - x) \Big| X_j \right\} \sum_{\ell \neq k \neq j} \mathbb{E} \left\{ K_{U,2,h}(W_k - x) \Big| X_k \right\} \mathbb{E} \left\{ K_{U,2,h}(W_\ell - x) \Big| X_\ell \right\}.$$

Applying (D.1) and (D.4) to expectations in the above elaboration of $E(I|\boldsymbol{X})$, we have, if U is ordinary smooth of order β,

$$\mathbb{E}(I|\boldsymbol{X})$$
$$= \frac{\eta(0,0,2,2)}{h^{3+4\beta}} f_U(X_j - x) + o_p\left(\frac{1}{h^{3+4\beta}}\right) + \frac{\eta(0,0)\eta(2,2)}{h^{2+4\beta}} f_U(X_j - x) \times$$
$$\sum_{k \neq j} f_U(X_k - x) + o_p\left(\frac{n}{h^{2+4\beta}}\right) + \tag{D.6}$$
$$\left\{ \frac{\eta(0,0,2)}{h^{2+3\beta}} f_U(X_j - x) + o_p\left(\frac{1}{h^{2+3\beta}}\right) \right\} \times 2 \sum_{k \neq j} \left(\frac{X_k - x}{h}\right)^2 K_h(X_k - x) + \tag{D.7}$$
$$\left\{ \frac{\eta(0,0)}{h^{1+2\beta}} f_U(X_j - x) + o_p\left(\frac{1}{h^{1+2\beta}}\right) \right\} \sum_{\ell \neq k \neq j} \left(\frac{X_k - x}{h}\right)^2 K_h(X_k - x) \times$$
$$\left(\frac{X_\ell - x}{h}\right)^2 K_h(X_\ell - x), \tag{D.8}$$

which can be further simplified by keeping only the dominating term as we explain next. Using the approximation for a random variable A with a finite variance, $A = \mathbb{E}(A) + O_p(\sqrt{\text{Var}(A)})$, one can show that

$$\begin{cases} \left(\dfrac{X - x}{h}\right) K_h(X - x) = h f_X'(x) \mu_2 + O(h^3) + O_p(1/\sqrt{h}), \\ \left(\dfrac{X - x}{h}\right)^2 K_h(X - x) = f_X(x) \mu_2 + O(h^2) + O_p(1/\sqrt{h}), \end{cases} \tag{D.9}$$

as $n \to \infty$ and $h \to 0$, and thus

$$
\begin{cases}
\displaystyle\sum_{\ell \neq k \neq j} \left(\frac{X_k - x}{h}\right) K_h(X_k - x) \left(\frac{X_\ell - x}{h}\right) K_h(X_\ell - x) \\[2mm]
= (n-1)(n-2) \left[h^2 \left\{ f_X'(x) \right\}^2 \mu_2^2 + O_P\left(\frac{1}{h}\right) \right], \\[4mm]
\displaystyle\sum_{\ell \neq k \neq j} \left(\frac{X_k - x}{h}\right)^2 K_h(X_k - x) \left(\frac{X_\ell - x}{h}\right) K_h(X_\ell - x) \\[2mm]
= (n-1)(n-2) \left\{ h f_X(x) f_X'(x) \mu_2^2 + O_P\left(\frac{1}{h}\right) \right\}, \\[4mm]
\displaystyle\sum_{\ell \neq k \neq j} \left(\frac{X_k - x}{h}\right)^2 K_h(X_k - x) \left(\frac{X_\ell - x}{h}\right)^2 K_h(X_\ell - x) \\[2mm]
= (n-1)(n-2) \left\{ f_X^2(x) \mu_2^2 + O_P\left(\frac{1}{h}\right) \right\},
\end{cases}
\tag{D.10}
$$

which are of order n^2/h, and hence dominate $\sum_{k \neq j} \{(X_k - x)/h\} K_h(X-k-x)$ and $\sum_{k \neq j} \{(X_k - x)/h\}^2 K_h(X-k-x)$ that are of order $n/\sqrt{h}$. Now one can see that, with $nh^{1+2\beta} \to \infty$, (D.6) is of order $n/h^{2+4\beta}$, which is dominated by (D.8) that is of order $n^2/h^{2+2\beta}$; and, if $\beta > 0.5$ as we assume henceforth, (D.6) dominates (D.7) that is of order $n/h^{2.5+3\beta}$. We thus conclude that

$$
\begin{aligned}
&\mathbb{E}(I|\mathbf{X}) \\
&= \left\{ \frac{\eta(0,0)}{h^{1+2\beta}} f_U(X_j - x) + o_P\left(\frac{1}{h^{1+2\beta}}\right) \right\} \times \\
&\quad \sum_{\ell \neq k \neq j} \left(\frac{X_k - x}{h}\right)^2 K_h(X_k - x) \left(\frac{X_\ell - x}{h}\right)^2 K_h(X_\ell - x).
\end{aligned}
\tag{D.11}
$$

If U is super smooth of order β, assuming $nh^{1-2\beta_2} \exp(-2h^{-\beta}/\gamma) \to \infty$, we have

$$
\begin{aligned}
&\mathbb{E}(I|\mathbf{X}) \tag{D.12} \\
&= O_P\left(\frac{\exp(4h^{-\beta}/\gamma)}{h^{3-4\beta_2}}\right) + O_P\left(\frac{n\exp(4h^{-\beta}/\gamma)}{h^{2-4\beta_2}}\right) + O_P\left(\frac{\exp(3h^{-\beta}/\gamma)}{h^{2-3\beta_2}}\right) \times \\
&\quad 2\sum_{k \neq j} \left(\frac{X_k - x}{h}\right)^2 K_h(X_k - x) + O_P\left(\frac{\exp(2h^{-\beta}/\gamma)}{h^{1-2\beta_2}}\right) \times \\
&\quad \sum_{\ell \neq k \neq j} \left(\frac{X_k - x}{h}\right)^2 K_h(X_k - x) \left(\frac{X_\ell - x}{h}\right)^2 K_h(X_\ell - x) \\
&= O_P\left(\frac{\exp(2h^{-\beta}/\gamma)}{h^{1-2\beta_2}}\right) \sum_{\ell \neq k \neq j} \left(\frac{X_k - x}{h}\right)^2 K_h(X_k - x) \left(\frac{X_\ell - x}{h}\right)^2 K_h(X_\ell - x). \tag{D.13}
\end{aligned}
$$

Second,

$$
\mathbb{E}(II|\mathbf{X})
$$

$$
= \mathbb{E}\left\{ K_{U,0,h}(W_j - x)K_{U,1,h}^2(W_j - x)K_{U,2,h}(W_j - x)\Big| X_j \right\} +
$$

$$
\mathbb{E}\left\{ K_{U,0,h}(W_j - x)K_{U,1,h}(W_j - x)\Big| X_j \right\} \times
$$

$$
\sum_{k \neq j} \mathbb{E}\left\{ K_{U,1,h}(W_k - x)K_{U,2,h}(W_k - x)\Big| X_k \right\} +
$$

$$
\mathbb{E}\left\{ K_{U,0,h}(W_j - x)K_{U,1,h}^2(W_j - x)\Big| X_j \right\} \sum_{k \neq j} \mathbb{E}\left\{ K_{U,2,h}(W_k - x)\Big| X_k \right\} +
$$

$$
\mathbb{E}\left\{ K_{U,0,h}(W_j - x)K_{U,1,h}(W_j - x)K_{U,2,h}(W_j - x)\Big| X_j \right\} \times
$$

$$
\sum_{\ell \neq j} \mathbb{E}\left\{ K_{U,1,h}(W_\ell - x)\Big| X_\ell \right\} + \mathbb{E}\left\{ K_{U,0,h}(W_j - x)K_{U,1,h}(W_j - x)\Big| X_j \right\} \times
$$

$$
\sum_{\ell \neq k \neq j} \mathbb{E}\left\{ K_{U,1,h}(W_\ell - x)\Big| X_\ell \right\} \mathbb{E}\left\{ K_{U,2,h}(W_k - x)\Big| X_k \right\}.
$$

By (D.1) and (D.4), if U is ordinary smooth of order β, then

$$
\mathbb{E}(II|X)
$$

$$
= \frac{\eta(0,1,1,2)}{h^{3+4\beta}} f_U(X_j - x) + o_p\left(\frac{1}{h^{3+4\beta}} \right) + \frac{\eta(0,1)\eta(1,2)}{h^{2+4\beta}} f_U(X_j - x) \times
$$

$$
\sum_{k \neq j} f_U(X_k - x) + o_p\left(\frac{n}{h^{2+4\beta}} \right) + \left\{ \frac{\eta(0,1,1)}{h^{2+3\beta}} f_U(X_j - x) + o_p\left(\frac{1}{h^{2+3\beta}} \right) \right\} \times
$$

$$
\sum_{k \neq j} \left(\frac{X_k - x}{h} \right)^2 K_h(X_k - x) + \left\{ \frac{\eta(0,1,2)}{h^{2+3\beta}} f_U(X_j - x) + o_p\left(\frac{1}{h^{2+3\beta}} \right) \right\} \times
$$

$$
\sum_{\ell \neq j} \left(\frac{X_\ell - x}{h} \right) K_h(X_\ell - x) + \left\{ \frac{\eta(0,1)}{h^{1+2\beta}} f_U(X_j - x) + o_p\left(\frac{1}{h^{1+2\beta}} \right) \right\} \times
$$

$$
\sum_{\ell \neq k \neq j} \left(\frac{X_\ell - x}{h} \right) K_h(X_\ell - x) \left(\frac{X_k - x}{h} \right)^2 K_h(X_k - x)
$$

$$
= \left\{ \frac{\eta(0,1)}{h^{1+2\beta}} f_U(X_j - x) + o_p\left(\frac{1}{h^{1+2\beta}} \right) \right\} \times
$$

$$
\sum_{\ell \neq k \neq j} \left(\frac{X_\ell - x}{h} \right) K_h(X_\ell - x) \left(\frac{X_k - x}{h} \right)^2 K_h(X_k - x), \tag{D.14}
$$

which is of the same order as that of (D.11). If U is super smooth of order β,

$$
\mathbb{E}(II|X)
$$

$$
= O_p\left(\frac{\exp(4h^{-\beta}/\gamma)}{h^{3-4\beta_2}} \right) + O_p\left(\frac{n\exp(4h^{-\beta}/\gamma)}{h^{2-4\beta_2}} \right)
$$

$$
O_p\left(\frac{\exp(3h^{-\beta}/\gamma)}{h^{2-3\beta_2}} \right) \sum_{k \neq j} \left(\frac{X_k - x}{h} \right)^2 K_h(X_k - x) +
$$

$$
O_p\left(\frac{\exp(3h^{-\beta}/\gamma)}{h^{2-3\beta_2}} \right) \sum_{\ell \neq j} \left(\frac{X_\ell - x}{h} \right) K_h(X_\ell - x) +
$$

$$O_p\left(\frac{\exp(2h^{-\beta}/\gamma)}{h^{1-2\beta_2}}\right) \sum_{\ell\neq k\neq j} \left(\frac{X_\ell - x}{h}\right) K_h(X_\ell - x)\left(\frac{X_k - x}{h}\right)^2 K_h(X_k - x)$$

$$= O_p\left(\frac{\exp(2h^{-\beta}/\gamma)}{h^{1-2\beta_2}}\right) \sum_{\ell\neq k\neq j} \left(\frac{X_\ell - x}{h}\right) K_h(X_\ell - x)\left(\frac{X_k - x}{h}\right)^2 K_h(X_k - x), \qquad \text{(D.15)}$$

which is of the same order as that of (D.13). Third,

$$\mathbb{E}(III|X)$$
$$= \mathbb{E}\left\{K_{U,1,h}^4(W_j - x)\Big| X_j\right\} +$$
$$2\mathbb{E}\left\{K_{U,1,h}^3(W_j - x)\Big| X_j\right\} \sum_{k\neq j}\mathbb{E}\left\{K_{U,1,h}(W_k - x)\Big| X_k\right\} +$$
$$\mathbb{E}\left\{K_{U,1,h}^2(W_j - x)\Big| X_j\right\} \sum_{k\neq j}\mathbb{E}\left\{K_{U,1,h}^2(W_k - x)\Big| X_k\right\} +$$
$$\mathbb{E}\left\{K_{U,1,h}^2(W_j - x)\Big| X_j\right\} \sum_{\ell\neq k\neq j}\mathbb{E}\{K_{U,1,h}(W_\ell - x)|X_\ell\}\mathbb{E}\{K_{U,1,h}(W_k - x)|X_k\}.$$

By (D.1) and (D.4), if U is ordinary smooth of order β, then

$$E(III|X)$$
$$= \frac{\eta(1,1,1,1)}{h^{3+4\beta}}f_U(X_j - x) + o_p\left(\frac{1}{h^{3+4\beta}}\right) + \left\{\frac{\eta(1,1,1)}{h^{2+3\beta}}f_U(X_j - x) + \right.$$
$$\left. o_p\left(\frac{1}{h^{2+3\beta}}\right)\right\} \times 2\sum_{k\neq j}\left(\frac{X_k - x}{h}\right)K_h(X_k - x) + \left\{\frac{\eta(1,1)}{h^{1+2\beta}}f_U(X_j - x) + \right.$$
$$\left. o_p\left(\frac{1}{h^{1+2\beta}}\right)\right\} \sum_{k\neq j}\left\{\frac{\eta(1,1)}{h^{1+2\beta}}f_U(X_k - x) + o_p\left(\frac{1}{h^{1+2\beta}}\right)\right\} +$$
$$\left\{\frac{\eta(1,1)}{h^{1+2\beta}}f_U(X_j - x) + o_p\left(\frac{1}{h^{1+2\beta}}\right)\right\} \sum_{\ell\neq k\neq j}\left(\frac{X_\ell - x}{h}\right)K_h(X_\ell - x)\times$$
$$\left(\frac{X_k - x}{h}\right)K_h(X_k - x)$$
$$= \left\{\frac{\eta(1,1)}{h^{1+2\beta}}f_U(X_j - x) + o_p\left(\frac{1}{h^{1+2\beta}}\right)\right\} \sum_{\ell\neq k\neq j}\left(\frac{X_\ell - x}{h}\right)K_h(X_\ell - x)\times$$
$$\left(\frac{X_k - x}{h}\right)K_h(X_k - x), \qquad \text{(D.16)}$$

which is of the same order as that of (D.11). If U is super smooth of order β, then

$$E(III|X)$$
$$= O_p\left(\frac{\exp(4h^{-\beta}/\gamma)}{h^{3-4\beta_2}}\right) + O_p\left(\frac{\exp(3h^{-\beta}/\gamma)}{h^{2-3\beta_2}}\right) \times 2\sum_{k\neq j}\left(\frac{X_k - x}{h}\right)K_h(X_k - x) +$$
$$O_p\left(\frac{\exp(2h^{-\beta}/\gamma)}{h^{1-2\beta_2}}\right) \times \sum_{\ell\neq k\neq j}\left(\frac{X_\ell - x}{h}\right)K_h(X_\ell - x)\left(\frac{X_k - x}{h}\right)K_h(X_k - x) +$$

$$O_p\left(\frac{n\exp(4h^{-\beta}/\gamma)}{h^{2-4\beta_2}}\right)$$

$$= O_p\left(\frac{\exp(2h^{-\beta}/\gamma)}{h^{1-2\beta_2}}\right) \sum_{\ell\neq k\neq j}\left(\frac{X_\ell-x}{h}\right)K_h(X_\ell-x)\left(\frac{X_k-x}{h}\right)K_h(X_k-x), \qquad \text{(D.17)}$$

which is of the same order as that of (D.13). Combining (D.11), (D.14), and (D.16) for ordinary smooth U, we have

$$\mathbb{E}\left\{\mathcal{L}^2(W_j-x)\big|X\right\}$$

$$= \frac{f_U(X_j-x)}{n^2h^{1+2\beta}}\left\{\eta(0,0)\sum_{\ell\neq k\neq j}\left(\frac{X_k-x}{h}\right)^2 K_h(X_k-x)\left(\frac{X_\ell-x}{h}\right)^2 K_h(X_\ell-x)\right.$$

$$-2\eta(0,1)\sum_{\ell\neq k\neq j}\left(\frac{X_k-x}{h}\right)^2 K_h(X_k-x)\left(\frac{X_\ell-x}{h}\right)K_h(X_\ell-x) \qquad \text{(D.18)}$$

$$\left.+\eta(1,1)\sum_{\ell\neq k\neq j}\left(\frac{X_k-x}{h}\right)K_h(X_k-x)\left(\frac{X_\ell-x}{h}\right)K_h(X_\ell-x)\right\}$$

$$+o_p\left(\frac{1}{h^{1+2\beta}}\right),$$

and, using (D.13), (D.15), and (D.17) for super smooth U gives

$$\mathbb{E}\left\{\mathcal{L}^2(W_j-x)\big|X\right\} = O_p\left(\frac{\exp(2h^{-\beta}/\gamma)}{h^{1-2\beta_2}}\right). \qquad \text{(D.19)}$$

D.3. Expectations of $\hat{g}_{1,DK}(x)$ and $\hat{g}_{2,DK}(x)$

For $\hat{g}_{1,\mathrm{DK}}(x) = n^{-1}\sum_{j=1}^n \sin\Theta_j\mathcal{L}(W_j-x)$ and $\hat{g}_{2,\mathrm{DK}}(x) = n^{-1}\sum_{j=1}^n \cos\Theta_j\ \mathcal{L}(W_j-x)$, by (D.5),

$$\mathbb{E}\left\{\hat{g}_{\ell,\mathrm{DK}}(x)\big|X\right\}$$

$$= \frac{1}{n}\sum_{j=1}^n m_\ell(X_j)\mathbb{E}\left\{\mathcal{L}(W_j-x)\big|X\right\}$$

$$= \frac{1}{n}\sum_{j=1}^n m_\ell(X_j)\mathcal{W}(X_j-x)+$$

$$\begin{cases} \dfrac{\eta(0,2)-\eta(1,1)}{n^2h^{1+2\beta}}\displaystyle\sum_{j=1}^n m_\ell(X_j)f_U(X_j-x)+ & \\[2mm] o_p\left(\dfrac{1}{nh^{1+2\beta}}\right), & \text{if } U \text{ is ordinary smooth,} \\[3mm] O_p\left(\dfrac{\exp(2h^{-\beta}/\gamma)}{nh^{1-2\beta_2}}\right), & \text{if } U \text{ is super smooth.} \end{cases} \qquad \text{(D.20)}$$

Using existing results (e.g., those in (D.9)), one can show that

$$\mathbb{E}\left\{m_\ell(X)\mathcal{W}(X-x)\right\} = \tau_\ell(x)f_X(x)\mu_2 + \frac{h^2}{2}\{\tau_\ell^{(2)}(x)f_X(x)\mu_2^2+$$

$$\tau_\ell(x) f_X^{(2)}(x)\mu_4 - 2\tau_\ell'(x) f_X'(x)\mu_2^2\} + O(h^4),$$

where $\tau_\ell(x) = m_\ell(x) f_X(x)$. In addition, by (B.1), $\mathbb{E}\{m_\ell(X) f_U(X - x)\} = \tau_\ell^*(x)$, where $\tau_\ell^*(x) = m_\ell^*(x) f_W(x)$. It follows that

$$\mathbb{E}\left\{\hat{g}_{\ell,\mathrm{DK}}(x)|\mathbf{X}\right\}$$

$$= \tau_\ell(x) f_X(x)\mu_2 + \frac{h^2}{2}\left\{\tau_\ell^{(2)}(x) f_X(x)\mu_2^2 + \tau_\ell(x) f_X^{(2)}(x)\mu_4 - \right.$$

$$\left. 2\tau_\ell'(x) f_X'(x)\mu_2^2\right\} + o(h^2)+ \tag{D.21}$$

$$\begin{cases} \dfrac{\eta(0,2) - \eta(1,1)}{nh^{1+2\beta}}\tau_\ell^*(x) + o_P\left(\dfrac{1}{nh^{1+2\beta}}\right), & \text{if } U \text{ is ordinary smooth,} \\ O_P\left(\dfrac{\exp(2h^{-\beta}/\gamma)}{nh^{1-2\beta_2}}\right), & \text{if } U \text{ is super smooth.} \end{cases}$$

Hence, $\hat{g}_{\ell,\mathrm{DK}}(x)$ is an asymptotically unbiased estimator of $m_\ell(x) f_X^2(x)\mu_2$ if $nh^{1+2\beta} \to \infty$ for ordinary smooth U, or $nh^{1-2\beta_2}\exp(-2h^{-\beta}/\gamma) \to \infty$ for super smooth U. If we further assume that $nh^{3+2\beta} \to \infty$ for ordinary smooth U, or $nh^{3-2\beta_2}\exp(-2h^{-\beta}/\gamma) \to \infty$ for super smooth U, then the dominating bias of $\hat{g}_{\ell,\mathrm{DK}}(x)$ (as an estimator for $m_\ell(x) f_X^2(x)\mu_2$) is equal to that of $\tilde{g}_\ell(x)$, i.e., the counterpart estimator based on error-free data.

D.4. Variances and covariance of $\hat{g}_{1,DK}(x)$ and $\hat{g}_{2,DK}(x)$

Because $\mathrm{Var}\{\hat{g}_{\ell,\mathrm{DK}}(x)|\mathbf{X}\}$ is dominated by $\mathbb{E}\{\hat{g}_{\ell,\mathrm{DK}}^2(x)|\mathbf{X}\}$, we focus on this conditional expectation next.

Define $\xi_1(x) = \mathbb{E}\{(\sin\Theta)^2|X = x\}$ and $\xi_2(x) = \mathbb{E}\{(\cos\Theta)^2|X = x\}$. Results in Sections D.1 and D.2 regarding moments of $\mathcal{L}(W_j - x)$ are now helpful in deriving the following expectation,

$$\mathbb{E}\left\{\hat{g}_{1,\mathrm{DK}}^2(x)\big|\mathbf{X}\right\} = \frac{1}{n^2}\mathbb{E}\left[\left\{\sum_{j=1}^n \sin\Theta_j\mathcal{L}(W_j - x)\right\}^2\bigg|\mathbf{X}\right]$$

$$= \frac{1}{n^2}\mathbb{E}\left\{\sum_{j=1}^n (\sin\Theta_j)^2\mathcal{L}^2(W_j - x)\bigg|\mathbf{X}\right\}$$

$$+ \frac{1}{n^2}\mathbb{E}\left\{\sum_{j\neq k}\sin\Theta_j\mathcal{L}(W_j - x)\sin\Theta_k\mathcal{L}(W_k - x)\bigg|\mathbf{X}\right\}$$

$$= \frac{1}{n^2}\sum_{j=1}^n \xi_1(X_j)\mathbb{E}\left\{\mathcal{L}^2(W_j - x)\big|\mathbf{X}\right\}+$$

$$\frac{1}{n^2}\sum_{j\neq k}m_1(X_j)m_1(X_k)\mathbb{E}\{\mathcal{L}(W_j - x)\mathcal{L}(W_k - x)|\mathbf{X}\},$$

where the second summation is dominated by the first summation and thus is absorbed in the non-dominating (lower-order) terms of the first summation that is elaborated next.

When U is ordinary smooth, by (D.18), we have

$$\mathbb{E}\left\{\hat{g}_{1,\mathrm{DK}}^2(x)\big|\mathbf{X}\right\}$$

$$= \frac{1}{n^4 h^{1+2\beta}} \sum_{j=1}^{n} \xi_1(X_j) f_U(X_j - x) \left\{ \eta(0,0) \sum_{\ell \neq k \neq j} \left(\frac{X_k - x}{h} \right)^2 K_h(X_k - x) \times \right.$$

$$\left(\frac{X_\ell - x}{h} \right)^2 K_h(X_\ell - x) - 2\eta(0,1) \sum_{\ell \neq k \neq j} \left(\frac{X_k - x}{h} \right)^2 K_h(X_k - x) \left(\frac{X_\ell - x}{h} \right)$$

$$K_h(X_\ell - x) + \eta(1,1) \sum_{\ell \neq k \neq j} \left(\frac{X_k - x}{h} \right) K_h(X_k - x) \left(\frac{X_\ell - x}{h} \right) K_h(X_\ell - x) \right\}$$

$$+ o_p \left(\frac{1}{nh^{2+2\beta}} \right).$$

Finally, using (B.2) and (D.10), one can show that

$$\mathbb{E}\{\hat{g}_{1,\text{DK}}^2(x)|\boldsymbol{X}\} = \frac{\eta(0,0)}{nh^{1+2\beta}} \xi_1^*(x) f_W(x) f_X^2(x) \mu_2^2 + o_p \left(\frac{1}{nh^{1+2\beta}} \right),$$

where $\xi_1^*(x) = \mathbb{E}\{(\sin\Theta)^2 | W = x\}$. Similarly,

$$\mathbb{E}\{\hat{g}_{2,\text{DK}}^2(x)|\boldsymbol{X}\} = \frac{\eta(0,0)}{nh^{1+2\beta}} \xi_2^*(x) f_W(x) f_X^2(x) \mu_2^2 + o_p \left(\frac{1}{nh^{1+2\beta}} \right),$$

where $\xi_2^*(x) = \mathbb{E}\{(\cos\Theta)^2 | W = x\}$. Hence, when U is ordinary smooth,

$$\text{Var}\{\hat{g}_{\ell,\text{DK}}(x)|\boldsymbol{X}\} = \frac{\eta(0,0)}{nh^{1+2\beta}} \xi_\ell^*(x) f_W(x) f_X^2(x) \mu_2^2 + o_p \left(\frac{1}{nh^{1+2\beta}} \right), \text{ for } \ell = 1, 2. \quad \text{(D.22)}$$

When U is super smooth, by (D.19), one has

$$\text{Var}\{\hat{g}_{\ell,\text{DK}}(x)|\boldsymbol{X}\} = O_p \left(\frac{\exp(2h^{-\beta}/\gamma)}{nh^{1-2\beta_2}} \right), \text{ for } \ell = 1, 2. \quad \text{(D.23)}$$

Because $\text{Cov}(\hat{g}_{1,\text{DK}}(x), \hat{g}_{2,\text{DK}}(x)|\boldsymbol{X})$ is dominated by $\mathbb{E}\{\hat{g}_{1,\text{DK}}(x)\hat{g}_{2,\text{DK}}(x)|\boldsymbol{X}\}$, we focus on deriving this conditional expectation next.

Define $\psi(x) = \mathbb{E}(\sin\Theta\cos\Theta | X = x)$, then we have

$$\mathbb{E}\left\{ \hat{g}_{1,\text{DK}}(x)\hat{g}_{2,\text{DK}}(x) \middle| \boldsymbol{X} \right\}$$

$$= \frac{1}{n^2} \sum_{j=1}^{n} \mathbb{E}\left\{ \sin\Theta_j \cos\Theta_j \mathcal{L}^2(W_j - x) \middle| \boldsymbol{X} \right\} +$$

$$\frac{1}{n^2} \sum_{j \neq k} \mathbb{E}\left\{ \sin\Theta_j \mathcal{L}(W_j - x) \cos\Theta_k \mathcal{L}(W_k - x) \middle| \boldsymbol{X} \right\}$$

$$= \frac{1}{n^2} \sum_{j=1}^{n} \psi(X_j) \mathbb{E}\left\{ \mathcal{L}^2(W_j - x) \middle| \boldsymbol{X} \right\} +$$

$$\frac{1}{n^2} \sum_{j \neq k} m_1(X_j) m_2(X_k) \mathbb{E}\left\{ \mathcal{L}(W_j - x) \mathcal{L}(W_k - x) \middle| \boldsymbol{X} \right\}.$$

Using earlier findings for the above two conditional expectations, we obtain that, when U is ordinary smooth,

$$\text{Cov}\{\hat{g}_{1,\text{DK}}(x), \hat{g}_{2,\text{DK}}(x)|\boldsymbol{X}\} = \frac{\eta(0,0)}{nh^{1+2\beta}} \psi^*(x) f_W(x) f_X^2(x) \mu_2^2 + o_p \left(\frac{1}{nh^{1+2\beta}} \right), \quad \text{(D.24)}$$

where $\psi^*(x) = \mathbb{E}(\sin\Theta\cos\Theta|W = x)$; and, when U is super smooth,

$$\text{Cov}\{\hat{g}_{1,\text{DK}}(x), \hat{g}_{2,\text{DK}}(x)|X\} = O_p\left(\frac{\exp(2h^{-\beta}/\gamma)}{nh^{1-2\beta_2}}\right). \tag{D.25}$$

D.5. Asymptotic bias of $\hat{m}_{DK}(x)$

With the first and second asymptotic moments of $\hat{g}_{\ell,\text{DK}}(x)$ derived, we are ready to derive the asymptotic bias of $\hat{m}_{\text{DK}}(x)$. Suppressing the dependence on x for simpler notations, we view $\hat{m}.$ as a bivariate function with arguments $(\hat{g}_{1,.}, \hat{g}_{2,.})$, i.e., $\hat{m}. = \text{atan2}[\hat{g}_{1,.}, \hat{g}_{2,.}]$. A Taylor expansion of $\hat{m}.$ around $(\hat{g}_{1,.}, \hat{g}_{2,.}) = (g_1, g_2)$ is given by

$$\begin{aligned}
&\hat{m}. - m \\
&= \frac{1}{g_1^2 + g_2^2}\left\{g_2\left(\hat{g}_{1,.} - g_1\right) - g_1\left(\hat{g}_{2,.} - g_2\right)\right\} - \frac{1}{(g_1^2 + g_2^2)^2}\times \\
&\left[g_1 g_2\left\{\left(\hat{g}_{1,.} - g_1\right)^2 - \left(\hat{g}_{2,.} - g_2\right)^2\right\} + \left(g_2^2 - g_1^2\right)\left(\hat{g}_{1,.} - g_1\right)\left(\hat{g}_{2,.} - g_2\right)\right] + \\
&O_p\left\{\left(\hat{g}_{1,.} - g_1\right)^3\right\} + O_p\left\{\left(\hat{g}_{2,.} - g_2\right)^3\right\}.
\end{aligned} \tag{D.26}$$

Taking the expectation of both sides of (D.26) gives

$$\begin{aligned}
&\text{Bias}\,(\hat{m}.) \\
&= \frac{1}{g_1^2 + g_2^2}\left\{g_2\mathbb{E}(\hat{g}_{1,.} - g_1) - g_1\mathbb{E}(\hat{g}_{2,.} - g_2)\right\} - \frac{1}{(g_1^2 + g_2^2)^2}\left(g_1 g_2\left[\mathbb{E}\left\{(\hat{g}_{1,.} - g_1)^2\right\}\right.\right. \\
&\left.\left. -\mathbb{E}\left\{(\hat{g}_{2,.} - g_2)^2\right\}\right] + \left(g_2^2 - g_1^2\right)\mathbb{E}\left\{(\hat{g}_{1,.} - g_1)(\hat{g}_{2,.} - g_2)\right\}\right) \\
&+ O\{(\hat{g}_{1,.} - g_1)^3\} + O\{(\hat{g}_{2,.} - g_2)^3\},
\end{aligned} \tag{D.27}$$

We consider the asymptotic results conditioning on X in this appendix, and (D.27) still holds when all the moments are conditional moments given X.

For the deconvoluting kernel estimator $\hat{m}_{\text{DK}}(x)$, $g_\ell(x) = m_\ell(x)f_X^2(x)\mu_2$, for $\ell = 1, 2$, in (D.27); $\mathbb{E}(\hat{g}_{\ell,\text{DK}} - g_\ell|X)$ as the bias of $\hat{g}_{\ell.\text{DK}}$ is revealed by (D.21); $\mathbb{E}\{(\hat{g}_{\ell,\text{DK}} - g_\ell)^2|X\} = \text{Var}(\hat{g}_{\ell,\text{DK}}|X) + \{\text{Bias}(\hat{g}_{\ell,\text{DK}}|X)\}^2$ is dominated by $\text{Var}(\hat{g}_{\ell,\text{DK}}|X)$ given (D.22) and (D.23), with $\{\text{Bias}(\hat{g}_{\ell,\text{DK}}|X)\}^2$ absorbed in lower-order terms in $\text{Bias}(\hat{g}_{\ell,\text{DK}}|X)$; lastly, $\mathbb{E}\{(\hat{g}_1 - g_1)(\hat{g}_2 - g_2)|X\}$ is equal to $\text{Cov}(\hat{g}_{1,\text{DK}}, \hat{g}_{2,\text{DK}}|X) + \text{Bias}(\hat{g}_{1,\text{DK}}|X)\text{Bias}(\hat{g}_{2,\text{DK}}|X)$, which is dominated by the first tem $\text{Cov}(\hat{g}_{1,\text{DK}}, \hat{g}_{2,\text{DK}}|X)$ given in (D.24) and (D.25), with the second term absorbed in lower order terms in $\text{Bias}(\hat{g}_{\ell,\text{DK}}|X)$. Therefore, putting (D.21), (D.22), and (D.24) in (D.27), we have, for ordinary smooth U,

$$\begin{aligned}
&\text{Bias}\,\{\hat{m}_{\text{DK}}(x)|X\} \\
&= \frac{h^2\mu_2}{2}\left\{\frac{m_1^{(2)}(x)m_2(x) - m_1(x)m_2^{(2)}(x)}{m_1^2(x) + m_2^2(x)}\right\} + o(h^2) + \\
&\frac{f_W(x)}{nh^{1+2\beta}f_X^2(x)\mu_2\left\{m_1^2(x) + m_2^2(x)\right\}}\times \\
&\left[\{\eta(0,2) - \eta(1,1)\}\{m_1^*(x)m_2(x) - m_1(x)m_2^*(x)\} - \right.
\end{aligned} \tag{D.28}$$

$$\eta(0,0)\mu_2 \times \frac{m_1(x)m_2(x)\left\{\xi_1^*(x) - \xi_2^*(x)\right\} + \psi^*(x)\left\{m_2^2(x) - m_1^2(x)\right\}}{m_1^2(x) + m_2^2(x)} \Bigg] +$$
$$o_P\left(\frac{1}{nh^{1+2\beta}}\right). \tag{D.29}$$

Viewing $m(x) = \text{atan2}[m_1(x), m_2(x)]$ as a univariate function with argument x, one has

$$m'(x) = \frac{m_1'(x)m_2(x) - m_1(x)m_2'(x)}{m_1^2(x) + m_2^2(x)},$$

following which one can show that

$$\frac{m_1^{(2)}(x)m_2(x) - m_1(x)m_2^{(2)}(x)}{m_1^2(x) + m_2^2(x)} = m^{(2)}(x) + 2m'(x)\frac{C'(x)}{C(x)}$$
$$= m^{(2)}(x) + 2m'(x)\frac{m_1(x)m_1'(x) + m_2(x)m_2'(x)}{m_1^2(x) + m_2^2(x)},$$

where $C(x) = \sqrt{m_1^2(x) + m_2^2(x)}$. This reveals that the dominating term in (D.28) coincides with the dominating bias of the ideal estimator $\tilde{m}(x)$ stated in [12, see Theorem 4]. For the second part of the dominating bias that is of order $1/(nh^{1+2\beta})$ following (D.28) above, one can see that parts of it reduce to zero if $W = X$, i.e., in the absence of measurement error so that, for instance, $m_\ell^*(x) = m_\ell(x)$. This second part of the dominating bias reflects the effect of measurement error on the bias of $\hat{m}_{\text{DK}}(x)$. Certainly, if $nh^{3+2\beta} \to \infty$ as $n \to \infty$ and $h \to 0$, then this second term is absorbed in $o(h^2)$ in (D.28), suggesting that a larger sample size is required if we want $\hat{m}_{\text{DK}}(x)$ to achieve a comparable level of bias as that of $\tilde{m}(x)$.

Similarly, if U is super smooth, incorporating (D.21), (D.23), and (D.25) in (D.27), we have

$$\text{Bias}\left\{\hat{m}_{\text{DK}}(x)|X\right\} = \frac{h^2\mu_2}{2} \times \left\{\frac{m_1^{(2)}(x)m_2(x) - m_1(x)m_2^{(2)}(x)}{m_1^2(x) + m_2^2(x)}\right\}$$
$$+ o(h^2) + O_P\left(\frac{\exp(2h^{-\beta}/\gamma)}{nh^{1-2\beta_2}}\right), \tag{D.30}$$

where the effect of measurement error is (relatively more) vaguely reflected in the last term that depends on the order of smoothness β. However it is easy to see that if $nh^{4-2\beta_2}\exp(-2h^{-\beta}/\gamma) \to \infty$ as $n \to \infty$ and $h \to 0$, then the dominating bias of $\hat{m}_{\text{DK}}(x)$ again coincides with that of $\tilde{m}(x)$. In other words, in the presence of super smooth measurement error, an even larger sample size than what is needed when U is ordinary smooth would be required for the bias of $m_{\text{DK}}(x)$ to be comparable with that of $\tilde{m}(x)$. These give the results in Theorem 4.1 in the main article regarding the asymptotic bias of $m_{\text{DK}}(x)$.

D.6. Asymptotic variance of $\hat{m}_{DK}(x)$

Now viewing $\hat{m}_\cdot^2$ as a bivariate function with arguments $(\hat{g}_{1,\cdot}, \hat{g}_{2,\cdot})$, we have a Taylor expansion of $\hat{m}_\cdot^2$ around $(\hat{g}_{1,\cdot}, \hat{g}_{2,\cdot}) = (g_1, g_2)$ given by

$$\hat{m}_\cdot^2$$

$$= m^2 + \frac{2m}{g_1^2 + g_2^2} \left\{ g_2 \left(\hat{g}_{1,\cdot} - g_1 \right) - g_1 \left(\hat{g}_{2,\cdot} - g_2 \right) \right\} + \frac{2m}{\left(g_1^2 + g_2^2 \right)^2} \times$$

$$\left[g_1 g_2 \left\{ \left(\hat{g}_{2,\cdot} - g_2 \right)^2 - \left(\hat{g}_{1,\cdot} - g_1 \right)^2 \right\} + \left(g_1^2 - g_2^2 \right) \left(\hat{g}_{1,\cdot} - g_1 \right) \left(\hat{g}_{2,\cdot} - g_2 \right) \right] +$$

$$\frac{1}{\left(g_1^2 + g_2^2 \right)^2} \left\{ g_2^2 \left(\hat{g}_{1,\cdot} - g_1 \right)^2 + g_1^2 \left(\hat{g}_{2,\cdot} - g_2 \right)^2 - 2 g_1 g_2 \left(\hat{g}_{1,\cdot} - g_1 \right) \left(\hat{g}_{2,\cdot} - g_2 \right) \right\} +$$

$$O_p \left\{ \left(\hat{g}_{1,\cdot} - g_1 \right)^3 \right\} + O_p \left\{ \left(\hat{g}_{2,\cdot} - g_2 \right)^3 \right\}.$$

Taking conditional expectations of both sides of the above equation and dropping terms that can be absorbed in lower-order terms of $\mathbb{E}(\hat{g}_{\ell,\cdot} - g_\ell | X)$ in (D.21), we have

$$\mathbb{E} \left(\hat{m}_{\cdot}^2 \middle| X \right)$$

$$= m^2 + \frac{2m}{g_1^2 + g_2^2} \left\{ g_2 \text{Bias} \left(\hat{g}_{1,\cdot} | X \right) - g_1 \text{Bias} \left(\hat{g}_{2,\cdot} | X \right) \right\} +$$

$$\frac{1}{\left(g_1^2 + g_2^2 \right)^2} \left\{ g_1 (2mg_2 + g_1) \text{Var} \left(\hat{g}_{2,\cdot} | X \right) - g_2 (2mg_1 - g_2) \text{Var} \left(\hat{g}_{1,\cdot} | X \right) \right\} + \quad \text{(D.31)}$$

$$\frac{2}{\left(g_1^2 + g_2^2 \right)^2} \left\{ m \left(g_1^2 - g_2^2 \right) - g_1 g_2 \right\} \text{Cov} \left(\hat{g}_{1,\cdot}, \hat{g}_{2,\cdot} | X \right) + \text{lower-order terms}.$$

Using results in Sections D.3 and D.4, we have, for ordinary smooth U,

$$\mathbb{E} \left\{ \hat{m}_{\text{DK}}^2 (x) | X \right\}$$

$$= m^2(x) + h^2 m(x) \mu_2 \left\{ m^{(2)}(x) + 2m'(x) \frac{C'(x)}{C(x)} \right\} + o_p(h^2) +$$

$$\frac{2 f_W(x)}{n h^{1+2\beta} f_X^2(x) \left\{ m_1^2(x) + m_2^2(x) \right\}} \left\{ m(x) \{ \eta(0,2) - \eta(1,1) \} \left\{ m_1^*(x) m_2(x) - \right. \right.$$

$$m_1(x) m_2^*(x) \right\} / \mu_2 + \frac{\eta(0,0)}{m_1^2(x) + m_2^2(x)} \left(m(x) m_1(x) m_2(x) \left\{ \xi_2^*(x) - \xi_1^*(x) \right\} + \quad \text{(D.32)}$$

$$\frac{m_1^2(x) \xi_2^*(x) + m_2^2(x) \xi_1^*(x)}{2} + \psi^*(x) \left[m(x) \left\{ m_1^2(x) - m_2^2(x) \right\} \right.$$

$$\left. \left. - m_1(x) m_2(x) \right] \right) \right\} + o_p \left(\frac{1}{n h^{1+2\beta}} \right),$$

and, for super smooth U,

$$\mathbb{E} \left\{ \hat{m}_{\text{DK}}^2 (x) | X \right\}$$

$$= m^2(x) + h^2 m(x) \mu_2 \left\{ m^{(2)}(x) + 2m'(x) \frac{C'(x)}{C(x)} \right\} + o_p(h^2) + O_p \left(\frac{\exp(2h^{-\beta}/\gamma)}{n h^{1-2\beta_2}} \right).$$

Focusing on the term of order $O(h^2)$ and the term of order $O(1/(nh^{1+2\beta}))$ for ordinary smooth U, we have

$$\mathbb{E}\{\hat{m}_{\text{DK}}^2(x) | X\} = m^2(x) + h^2 \times A + \frac{1}{n h^{1+2\beta}} \times B + o(h^2) + o_p \left(\frac{1}{n h^{1+2\beta}} \right) \quad \text{(D.33)}$$

and, with $\mathrm{Bias}\{\hat{m}_{\mathrm{DK}}(x)|X\} = h^2 \times C + \{1/(nh^{1+2\beta})\} \times D + o_p(1/(nh^{1+2\beta}))$,

$$
\begin{aligned}
&[\mathbb{E}\{\hat{m}_{\mathrm{DK}}(x)|X\}]^2 \\
&= [\mathrm{Bias}\{\hat{m}_{\mathrm{DK}}(x)|X\} + m(x)]^2 \\
&= \mathrm{Bias}^2\{\hat{m}_{\mathrm{DK}}(x)|X\} + m^2(x) + 2m(x)\mathrm{Bias}\{\hat{m}_{\mathrm{DK}}(x)|X\} \\
&= m^2(x) + 2m(x)\left(h^2 \times C + \frac{1}{nh^{1+2\beta}} \times D\right) + o(h^2) + o_p\left(\frac{1}{nh^{1+2\beta}}\right),
\end{aligned}
$$

where A and B are indicated by (D.32), C and D are implied by (D.28) and (D.29). In fact, $A - 2m(x)C = 0$, and thus, for ordinary smooth U,

$$
\begin{aligned}
&\mathrm{Var}\{\hat{m}_{\mathrm{DK}}(x)|X\} \\
&= h^2\{A - 2m(x)C\} + \frac{1}{nh^{1+2\beta}}\{B - 2m(x)D\} + o(h^2) + o_p\left(\frac{1}{nh^{1+2\beta}}\right) \\
&= \frac{f_W(x)\eta(0,0)}{nh^{1+2\beta}f_X^2(x)\{m_1^2(x) + m_2^2(x)\}^2}\{m_1^2(x)\xi_2^*(x) + m_2^2(x)\xi_1^*(x) \\
&\quad - 2m_1(x)m_2(x)\psi^*(x)\} + o_p\left(\frac{1}{nh^{1+2\beta}}\right);
\end{aligned}
$$

and, for super smooth U,

$$
\mathrm{Var}\{\hat{m}_{\mathrm{DK}}(x)|X\} = O_p\left(\frac{\exp(2h^{-\beta}/\gamma)}{nh^{1-2\beta_2}}\right).
$$

Appendix E: Asymptotic properties of the complex error estimator $\hat{m}_{\mathrm{CE}}(x)$

We consider the p-th order local polynomial weight used in $\hat{m}_{\mathrm{CE}}(x)$ in this appendix, for $p \geq 0$. Throughout this appendix, we assume normal measurement error and conditions imposed in Lemma 3.1 in the main article are satisfied, such as relevant functions being entire functions.

E.1. Expectations of $\hat{g}_{1,CE}(x)$ and $\hat{g}_{2,CE}(x)$

Recall that

$$
\begin{aligned}
\hat{g}_{1,\mathrm{CE}}(x) &= \frac{1}{n}\sum_{j=1}^{n}\sin\Theta_j \mathcal{L}^*(W_j - x) \\
&= \frac{1}{n}\sum_{j=1}^{n}\sin\Theta_j \times \frac{1}{B}\sum_{b=1}^{B}\mathcal{W}^*(W_{j,b}^* - x; W_b^*) \\
&= \frac{1}{B}\sum_{b=1}^{B}\frac{1}{n}\sum_{j=1}^{n}\sin\Theta_j \mathcal{W}^*(W_{j,b}^* - x; W_b^*) \\
&= \frac{1}{B}\sum_{b=1}^{B}\hat{g}_{1,b}^*(x),
\end{aligned}
$$

where $\hat{g}^*_{1,b}(x) = n^{-1} \sum_{j=1}^n \sin \Theta_j \mathcal{W}^*(W^*_{j,b} - x; W^*_b)$ is the same type of estimator as the ideal local linear estimator $\tilde{g}_1(x)$ for $g_1(x) = m_1(x) = \mathbb{E}(\sin \Theta | X = x)$, with the complex-valued weight $\mathcal{W}^*(W^*_{j,b} - x; W^*_b)$ in place of the normalized local linear weight $\mathcal{W}^*(X_j - x; X)$ in (5) in the main article, or, more generically, in place of the naive local polynomial weight of order p. By Lemma 3.1, $\mathbb{E}\{\hat{g}^*_{1,b}(x) | X, \Theta\} = \tilde{g}_1(x)$, and thus $\mathbb{E}\{\hat{g}_{1,\mathrm{CE}}(x) | X, \Theta\} = \tilde{g}_1(x)$. It follows that, if $\sigma_1^2(\cdot) = \mathrm{Var}(\sin \Theta | X = \cdot) = \xi_1(\cdot) - m_1^2(\cdot)$ is continuous in a neighborhood of x, and that $nh \to \infty$ as $n \to \infty$ and $h \to 0$, then

$$
\begin{aligned}
&\mathbb{E}\{\hat{g}_{1,\mathrm{CE}}(x) | X\} \\
&= \mathbb{E}\left[\mathbb{E}\left\{\hat{g}_{1,\mathrm{CE}}(x) | X, \Theta\right\} \big| X\right] \\
&= \mathbb{E}\left\{\tilde{g}_1(x) | X\right\}, \text{ next use Theorem 3.1 in [15]}, \\
&= \begin{cases} m_1(x) + e_1^\top S^{-1} \tilde{c}_p \frac{1}{(p+2)!} \left\{ m_1^{(p+2)}(x) + \right. \\ \left. (p+2) m_1^{(p+1)}(x) \frac{f'_X(x)}{f_X(x)} \right\} h^{p+2} + o_p(h^{p+2}), & \text{if } p \text{ is even}, \\ m_1(x) + e_1^\top S^{-1} c_p \frac{1}{(p+1)!} m_1^{(p+1)}(x) h^{p+1} + o_p(h^{p+1}), & \text{if } p \text{ is odd}, \end{cases}
\end{aligned} \qquad \text{(E.1)}
$$

where $S = (\mu_{r_1+r_2})_{0 \le r_1, r_2 \le p}$, $c_p = (\mu_{p+1}, \ldots, \mu_{2p+1})^\top$, $\tilde{c}_p = (\mu_{p+2}, \ldots, \mu_{2p+2})^\top$, and e_1 is the $(p+1) \times 1$ unit-norm vector with the first entry equal to 1. Similarly, if $\sigma_2^2(\cdot) = \mathrm{Var}(\cos \Theta | X = \cdot) = \xi_2(\cdot) - m_2^2(\cdot)$ is continuous in a neighborhood of x, and $nh \to \infty$ as $n \to \infty$ and $h \to 0$, then

$$
\begin{aligned}
&\mathbb{E}\{\hat{g}_{2,\mathrm{CE}}(x) | X\} \\
&= \begin{cases} m_2(x) + e_1^\top S^{-1} \tilde{c}_p \frac{1}{(p+2)!} \left\{ m_2^{(p+2)}(x) + \right. \\ \left. (p+2) m_2^{(p+1)}(x) \frac{f'_X(x)}{f_X(x)} \right\} h^{p+2} + o_p(h^{p+2}), & \text{if } p \text{ is even}, \\ m_2(x) + e_1^\top S^{-1} c_p \frac{1}{(p+1)!} m_2^{(p+1)} h^{p+1} + o_p(h^{p+1}), & \text{if } p \text{ is odd}, \end{cases}
\end{aligned} \qquad \text{(E.2)}
$$

In the special case considered in the main article with $p = 1$, we have

$$
\mathbb{E}\{\hat{g}_{\ell,\mathrm{CE}}(x) | X\} = m_\ell(x) + \frac{h^2}{2} \mu_2 m_\ell^{(2)}(x) + O_p(h^4), \text{ for } \ell = 1, 2.
$$

E.2. *Variances of* $\hat{g}_{1,CE}(x)$ *and* $\hat{g}_{2,CE}(x)$

By Lemma 3.1, $\mathbb{E}\{\hat{g}^*_{1,b}(x) | X, \Theta\} = \tilde{g}_1(x)$, and $\mathbb{E}[\{\hat{g}^*_{1,b}(x)\}^2 | X, \Theta] = \tilde{g}_1^2(x)$, thus

$$
\begin{aligned}
\mathrm{Var}\left\{\hat{g}^*_{1,b}(x) \big| X\right\} &= \mathbb{E}\left[\left\{\hat{g}^*_{1,b}(x)\right\}^2 \big| X\right] - \left[\mathbb{E}\left\{\hat{g}^*_{1,b}(x) \big| X\right\}\right]^2 \\
&= \mathbb{E}\left\{\tilde{g}_1^2(x) \big| X\right\} - \left[\mathbb{E}\left\{\tilde{g}_1(x) | X\right\}\right]^2 \\
&= \mathrm{Var}\{\tilde{g}_1(x) | X\} \\
&= e_1^\top S^{-1} v^* S^{-1} e_1 \frac{\sigma_1^2(x)}{nh f_X(x)} + o_p\left(\frac{1}{nh}\right),
\end{aligned} \qquad \text{(E.3)}
$$

where $v^* = (v_{r_1+r_2})_{0 \le r_1, r_2 \le p}$, and the last equation comes from the asymptotic variance of a local polynomial estimator of $m_1(x)$ in Theorem 3.1 in [15] under the assumption that $\sigma_1^2(\cdot)$ is continuous in a neighborhood of x, and that $nh \to \infty$ as $n \to \infty$ and $h \to 0$. It follows that

$$
\mathrm{Var}\left\{\hat{g}_{1,\mathrm{CE}}(x) | X\right\}
$$

$$
= \operatorname{Var} \left\{ \frac{1}{B} \sum_{b=1}^{B} \hat{g}_{1,b}^{*}(x) \,\middle|\, X \right\},
$$

$$
= \frac{1}{B^2} \sum_{b=1}^{B} \operatorname{Var} \left\{ \hat{g}_{1,b}^{*}(x) \,\middle|\, X \right\} + \frac{1}{B^2} \sum_{a \neq b} \operatorname{Cov} \left\{ \hat{g}_{1,a}^{*}(x), \hat{g}_{1,b}^{*}(x) \,\middle|\, X \right\}
$$

$$
= \frac{1}{B} \operatorname{Var} \left\{ \tilde{g}_1(x) \,\middle|\, X \right\} + \frac{B-1}{B} \left[\mathbb{E} \left\{ \hat{g}_{1,a}^{*}(x) \hat{g}_{1,b}^{*}(x) \,\middle|\, X \right\} - \right.
$$

$$
\left. \mathbb{E} \left\{ \hat{g}_{1,a}^{*}(x) \,\middle|\, X \right\} \mathbb{E} \left\{ \hat{g}_{1,b}^{*}(x) \,\middle|\, X \right\} \right], \text{ where } a \neq b,
$$

$$
= e_1^{\top} S^{-1} v^{*} S^{-1} e_1 \frac{\sigma_1^2(x)}{Bnh f_X(x)} + o_p \left(\frac{1}{Bnh} \right) + \frac{B-1}{B} \mathbb{E} \left\{ \hat{g}_{1,a}^{*}(x) \hat{g}_{1,b}^{*}(x) \,\middle|\, X \right\} -
$$

$$
\frac{B-1}{B} \left[\mathbb{E} \left\{ \tilde{g}_1(x) \,\middle|\, X \right\} \right]^2, \text{ by (E.3)}. \tag{E.4}
$$

Assume that $nh^{p+3} \to \infty$, if p is even, or, if p is odd, assume $nh^{p+2} \to \infty$, then the first two terms (of order $1/(nh)$ at a fixed B) are included in $o_p(h^{p+1+I(p \text{ is even})})$ that is the order of the last term. In what follows, we look into the only expectation left unsolved above, $\mathbb{E}\{\hat{g}_{1,a}^{*}(x)\hat{g}_{1,b}^{*}(x)|X\}$, which requires new tactics to solve due to its dependence on two sets of complex-valued error-contaminated covariate values, $W_a^{*} = (W_{1,a}^{*}, \ldots, W_{n,a}^{*})^{\top}$ and $W_b^{*} = (W_{1,b}^{*}, \ldots, W_{n,b}^{*})^{\top}$.

Adopting notations for weighted least squares estimators as in [Chapter 3, 15], we have

$$
\hat{g}_{1,a}^{*}(x)\hat{g}_{1,b}^{*}(x) = e_1^{\top} (G_a^{\top} \Omega_a G_a)^{-1} G_a^{\top} \Omega_a Y Y^{\top} \Omega_b G_b (G_b^{\top} \Omega_b G_b)^{-1} e_1,
$$

where $Y = (\sin \Theta_1, \ldots, \sin \Theta_n)^{\top}$, $\Omega_a = \operatorname{diag}(K_h(W_{1,a}^{*} - x), \ldots, K_h(W_{n,a}^{*} - x))$,

$$
G_a = \begin{bmatrix} 1 & W_{1,a}^{*} - x & \cdots & (W_{1,a}^{*} - x)^p \\ \vdots & \vdots & \ddots & \vdots \\ 1 & W_{n,a}^{*} - x & \cdots & (W_{n,a}^{*} - x)^p \end{bmatrix},
$$

and Ω_b and G_b are similarly defined. Recall that, for $j = 1, \ldots, n$, $W_{j,a}^{*} = W_j + i\sigma_u Z_{j,a}$ and $W_{j,b}^{*} = W_j + i\sigma_u Z_{j,b}$, with $Z_{j,a} \perp Z_{j,b}$ from $N(0,1)$ for $a \neq b$. Utilizing the law of iterated expectation, we first look into the expectation conditioning on W_a^{*}, W_b^{*}, and X,

$$
\mathbb{E} \left\{ \hat{g}_{1,a}^{*}(x) \hat{g}_{1,b}^{*}(x) \,\middle|\, W_a^{*}, W_b^{*}, X \right\}
$$

$$
= e_1^{\top} (G_a^{\top} \Omega_a G_a)^{-1} G_a^{\top} \Omega_a \mathbb{E}(Y Y^{\top} | X) \Omega_b G_b (G_b^{\top} \Omega_b G_b)^{-1} e_1
$$

$$
= e_1^{\top} (n^{-1} G_a^{\top} \Omega_a G_a)^{-1} (n^{-2} G_a^{\top} \Omega_a \Sigma \Omega_b G_b) (n^{-1} G_b^{\top} \Omega_b G_b)^{-1} e_1, \tag{E.5}
$$

where

$$
\Sigma = \mathbb{E}\left(Y Y^{\top} \,\middle|\, X \right) = \begin{bmatrix} \xi_1(X_1) & m_1(X_1)m_1(X_2) & \cdots & m_1(X_1)m_1(X_n) \\ m_1(X_2)m_1(X_1) & \xi_1(X_2) & \cdots & m_1(X_2)m_1(X_n) \\ \vdots & \vdots & \ddots & \vdots \\ m_1(X_n)m_1(X_1) & m_1(X_n)m_1(X_2) & \cdots & \xi_1(X_n) \end{bmatrix}.
$$

Because

$$
n^{-1} G_a^{\top} \Omega_a G_a = \left[\frac{1}{n} \sum_{j=1}^{n} (W_{j,a}^{*} - x)^{\ell+q-2} K_h(W_{j,a}^{*} - x) \right]_{1 \leq \ell, q \leq (p+1)},
$$

with the expectation of the (ℓ, q) entry conditioning on X equal to, by Lemma 3.1, for $\ell, q = 1, \ldots, p + 1$,

$$\frac{1}{n} \sum_{j=1}^{n} \mathbb{E} \left\{ (W_{j,a}^* - x)^{\ell+q-2} K_h (W_{j,a}^* - x) \Big| X \right\}$$

$$= \frac{1}{n} \sum_{j=1}^{n} (X_j - x)^{\ell+q-2} K_h (X_j - x)$$

$$= h^{\ell+q-2} \mu_{\ell+q-2} f_X(x) + h^{\ell+q-1} \mu_{\ell+q-1} f_X'(x) + h^{\ell+q} \mu_{\ell+q} f_X^{(2)}(x)/2$$

$$+ O_p \left(\frac{h^{\ell+q-2}}{\sqrt{nh}} \right),$$

therefore,

$$\mathbb{E} \left(n^{-1} G_a^\top \Omega_a G_a \Big| X \right)$$
$$= H \left\{ f_X(x) S + h f_X'(x) S_1 + h^2 f_X^{(2)}(x) S_2/2 + O_p \left(\frac{1}{\sqrt{nh}} \right) \right\} H, \tag{E.6}$$

where $H = \mathrm{diag}(1, h, \ldots, h^p)$, $S_\ell = (\mu_{r_1+r_2+\ell})_{0 \le r_1, r_2 \le p}$, for $\ell = 1, 2$. This also gives $\mathbb{E}\{n^{-1} G_b^\top \Omega_b G_b | X\}$.

Next we study the middle portion of (E.5), $n^{-2} G_a^\top \Omega_a \Sigma \Omega_b G_b$. The (ℓ, q)- entry of this matrix is

$$(n^{-2} G_a^\top \Omega_a \Sigma \Omega_b G_b)_{\ell, q}$$
$$= \frac{1}{n^2} \sum_{j=1}^{n} \sum_{k=1}^{n} K_h (W_{j,b}^* - x) K_h (W_{k,a}^* - x) \times \tag{E.7}$$
$$\mathbb{E}(Y_j Y_k | X)(W_{j,b}^* - x)^{q-1}(W_{k,a}^* - x)^{\ell-1}, \text{ for } \ell, q = 1, \ldots, p + 1.$$

For the summand of the above double sum, we have, if $j \ne k$, the expectation of the summand given X is

$$m_1(X_j)(X_j - x)^{q-1} K_h (X_j - x) m_1(X_k)(X_k - x)^{\ell-1} K_h (X_k - x)$$
$$= h^{q+\ell-2} \mu_{q-1} \mu_{\ell-1} \tau_1^2(x) + h^{q+\ell-1} (\mu_{q-1} \mu_\ell + \mu_q \mu_{\ell-1}) \tau_1(x) \tau_1'(x) +$$
$$\frac{h^{q+\ell}}{2} \left[\mu_{q-1} \mu_{\ell+1} \tau_1(x) \tau_1^{(2)}(x) + \mu_{q+1} \mu_{\ell-1} \tau_1(x) \tau_1^{(2)}(x) + 2\mu_q \mu_\ell \{\tau_1'(x)\}^2 \right] +$$
$$O_p \left(h^{q+\ell-3} \right).$$

If $j = k$, the expectation of the summand given X is

$$\mathbb{E} \left\{ K_h (W_{j,b}^* - x) K_h (W_{j,a}^* - x)(W_{j,b}^* - x)^{q-1}(W_{j,a}^* - x)^{\ell-1} \Big| X_j \right\} \xi_1(X_j)$$
$$= \mathbb{E} \left[\mathbb{E} \left\{ (W_{j,a}^* - x)^{\ell-1} K_h (W_{j,a}^* - x) \Big| W_j, X_j \right\} \times$$
$$\mathbb{E} \left\{ (W_{j,b}^* - x)^{q-1} K_h (W_{j,b}^* - x) \Big| W_j, X_j \right\} \Big| X_j \right] \xi_1(X_j)$$
$$= \mathbb{E} \left[\mathbb{E} \left\{ (W_{j,a}^* - x)^{\ell-1} K_h (W_{j,a}^* - x) \Big| W_j \right\} \times$$

$$\mathbb{E}\left\{(W_{j,b}^* - x)^{q-1} K_h(W_{j,b}^* - x)\Big| W_j\right\}\Big| X_j\right] \xi_1(X_j). \tag{E.8}$$

To find the inner expectation in (E.8), we consider the following Fourier transform,

$$\begin{aligned}
\phi_{(W^*-x)^k K_h(W^*-x)}(t) &= \int (W^* - x)^k K_h(W^* - x) e^{itx} \, dx \\
&= \exp(itW^*) h^k \int u^k K(u) e^{-ithu} \, du \\
&= \exp(itW^*) h^k i^{-k} \left. \frac{\partial^k \phi_K(s)}{\partial s^k}\right|_{s=-ht} \\
&= \exp(itW^*) h^k i^{-k} \phi_K^{(k)}(-ht).
\end{aligned} \tag{E.9}$$

Using the inverse Fourier transform, we elaborate the inner expectation in (E.8) as follows,

$$\begin{aligned}
&\mathbb{E}\left\{(W_{j,a}^* - x)^{\ell-1} K_h(W_{j,a}^* - x)\Big| W_j\right\} \\
&= \int (W_j + i\sigma_U z - x)^{\ell-1} K_h(W_j + i\sigma_U z - x) f_Z(z) \, dz \\
&= \int \frac{1}{2\pi} \int e^{-itx} \phi_{(W_j+i\sigma_U z-x)^{\ell-1} K_h(W_j+i\sigma_U z-x)}(t) \, dt \times \frac{1}{\sqrt{2\pi}} e^{-z^2/2} \, dz \\
&= \frac{1}{2\pi} h^{\ell-1} i^{-(\ell-1)} \int \exp\{it(W_j - x)\} \phi_K^{(\ell-1)}(-ht) \times \\
&\quad \frac{1}{\sqrt{2\pi}} \int \exp(-t\sigma_U z - z^2/2) \, dz dt, \text{ by (E.9)}, \\
&= \frac{1}{2\pi} h^{\ell-1} i^{-(\ell-1)} \int \exp\{it(W_j - x)\} \phi_K^{(\ell-1)}(-ht) \exp(t^2 \sigma_U^2/2) \, dt \\
&= h^{\ell-1} i^{-(\ell-1)} \frac{1}{2\pi} \int \exp\{it(W_j - x)\} \frac{\phi_K^{(\ell-1)}(-ht)}{\phi_U(t)} \, dt \\
&= h^{\ell-2} i^{-(\ell-1)} \frac{1}{2\pi} \int \exp\left\{-is\left(\frac{W_j - x}{h}\right)\right\} \frac{\phi_K^{(\ell-1)}(s)}{\phi_U(-s/h)} \, ds \\
&= h^{\ell-1} K_{U,\ell-1,h}(W_j - x), \text{ by (C.1)}.
\end{aligned}$$

Similarly, $\mathbb{E}\{(W_{j,b}^* - x)^{q-1} K_h(W_{j,b}^* - x)|W_j\} = h^{q-1} K_{U,q-1,h}(W_j - x)$. Hence, (E.8) reduces to, by (D.4),

$$\begin{aligned}
&h^{\ell+q-2} \mathbb{E}\left\{K_{U,\ell-1,h}(W_j - x) K_{U,q-1,h}(W_j - x)|X_j\right\} \xi_1(X_j) \\
&= \xi_1(X_j) h^{\ell+q-2} \times O_p\left(\frac{\exp(2h^{-\beta}/\gamma)}{h^{1-2\beta_2}}\right).
\end{aligned}$$

Finally, we have the expectation of (E.7) given X and assuming $\xi_1(X) = O_p(1)$,

$$\begin{aligned}
&\mathbb{E}\left\{\left(n^{-2} \boldsymbol{G}_a^\top \boldsymbol{\Omega}_a \boldsymbol{\Sigma} \boldsymbol{\Omega}_b \boldsymbol{G}_b\right)_{\ell,q}\Big| X\right\} \\
&= h^{q+\ell-2} \mu_{q-1} \mu_{\ell-1} \tau_1^2(x) + h^{q+\ell-1}(\mu_{q-1}\mu_\ell + \mu_q\mu_{\ell-1})\tau_1(x)\tau_1'(x) + \\
&\quad \frac{h^{q+\ell}}{2}\left[\mu_{q-1}\mu_{\ell+1}\tau_1(x)\tau_1^{(2)}(x) + \mu_{q+1}\mu_{\ell-1}\tau_1(x)\tau_1^{(2)}(x) + 2\mu_q\mu_\ell\{\tau_1'(x)\}^2\right] + \\
&\quad O_p\left(h^{q+\ell-3}\right) + O_p\left(\frac{\exp(2h^{-\beta}/\gamma)}{nh^{3-q-\ell-2\beta_2}}\right).
\end{aligned}$$

Putting these entries in the matrix, we have

$$
\begin{aligned}
&\mathbb{E}\left(n^{-2}\boldsymbol{G}_a^{\top}\boldsymbol{\Omega}_a\boldsymbol{\Sigma}\boldsymbol{\Omega}_b\boldsymbol{G}_b\big|\,\boldsymbol{X}\right)\\
&=\boldsymbol{H}\bigg(\tau_1^2(x)\boldsymbol{c}_0\boldsymbol{c}_0^{\top}+h\tau_1(x)\tau_1'(x)\left(\boldsymbol{c}_0\boldsymbol{c}_1^{\top}+\boldsymbol{c}_1\boldsymbol{c}_0^{\top}\right)\\
&\quad+\frac{h^2}{2}\left[\tau_1(x)\tau_1^{(2)}(x)\left(\boldsymbol{c}_0\boldsymbol{c}_2^{\top}+\boldsymbol{c}_2\boldsymbol{c}_0^{\top}\right)+2\left\{\tau_1'(x)\right\}^2\boldsymbol{c}_1\boldsymbol{c}_1^{\top}\right]\\
&\quad+O_p(1/h)+O_p\left(\frac{\exp(2h^{-\beta/\gamma})}{nh^{1-2\beta_2}}\right)\bigg)\boldsymbol{H},
\end{aligned}
\tag{E.10}
$$

where $\boldsymbol{c}_0=(\mu_0,\mu_1,\ldots,\mu_p)^{\top}$, $\boldsymbol{c}_1=(\mu_1,\mu_2,\ldots,\mu_{p+1})^{\top}$, $\boldsymbol{c}_2=(\mu_2,\mu_3,\ldots,\mu_{p+1})^{\top}$.

Using (E.6), (E.10), and that fact that $\boldsymbol{e}_1^{\top}\boldsymbol{H}^{-1}=\boldsymbol{e}_1^{\top}$, we have

$$
\begin{aligned}
&\mathbb{E}[\hat{g}_{1,a}^*(x)\hat{g}_{1,b}^*(x)|\boldsymbol{X}]\\
&=m_1^2(x)+\boldsymbol{e}_1^{\top}\boldsymbol{S}^{-1}\bigg[\frac{hm_1(x)}{f_X(x)}\left\{-m_1(x)f_X'(x)\left(\boldsymbol{c}_0\boldsymbol{c}_0^{\top}\boldsymbol{S}^{-1}\boldsymbol{S}_1+\boldsymbol{S}_1\boldsymbol{S}^{-1}\boldsymbol{c}_0\boldsymbol{c}_0^{\top}\right)+\right.\\
&\quad\tau_1'(x)(\boldsymbol{c}_0\boldsymbol{c}_1^{\top}+\boldsymbol{c}_1\boldsymbol{c}_0^{\top})\Big\}+\frac{h^2}{2f_X(x)}\bigg(\frac{1}{f_X(x)}\left[\tau_1(x)\tau_1^{(2)}(x)\left(\boldsymbol{c}_0\boldsymbol{c}_2^{\top}+\boldsymbol{c}_2\boldsymbol{c}_0^{\top}\right)+\right.\\
&\quad2\left\{\tau_1'(x)\right\}^2\boldsymbol{c}_1\boldsymbol{c}_1^{\top}\bigg]-\frac{2m_1(x)}{f_X(x)}f_x'(x)\tau_1'(x)(\boldsymbol{c}_0\boldsymbol{c}_1^{\top}+\boldsymbol{c}_1\boldsymbol{c}_0^{\top})\left(\boldsymbol{S}^{-1}\boldsymbol{S}_1+\boldsymbol{S}_1\boldsymbol{S}^{-1}\right)+\\
&\quad\frac{2m_1^2(x)}{f_X(x)}\left\{f_X'(x)\right\}^2\boldsymbol{S}_1\boldsymbol{S}^{-1}\boldsymbol{c}_0\boldsymbol{c}_0^{\top}\boldsymbol{S}^{-1}\boldsymbol{S}_1-m_1^2(x)f_X^{(2)}(x)\left(\boldsymbol{c}_0\boldsymbol{c}_0^{\top}\boldsymbol{S}^{-1}\boldsymbol{S}_2+\right.\\
&\quad\boldsymbol{S}_2\boldsymbol{S}^{-1}\boldsymbol{c}_0\boldsymbol{c}_0^{\top}\Big)\bigg)+o(h^2)+O_p\left(\frac{\exp(2h^{-\beta}/\gamma)}{nh^{1-2\beta_2}}\right)\bigg]\boldsymbol{S}^{-1}\boldsymbol{e}_1\\
&\triangleq m_1^2(x)+\boldsymbol{e}_1^{\top}\boldsymbol{S}^{-1}\mathbb{C}(h,m_1,p)\boldsymbol{S}^{-1}\boldsymbol{e}_1,
\end{aligned}
$$

where we use $\mathbb{C}(h,m_1,p)$ to denote the expression between $\boldsymbol{e}_1^{\top}\boldsymbol{S}^{-1}$ and $\boldsymbol{S}^{-1}\boldsymbol{e}_1$, with its dependence on h, p, and the function $m_1(\cdot)$ highlighted. Using this result in (E.4) yields, if p is odd,

$$
\begin{aligned}
&\mathrm{Var}\{\hat{g}_{1,\mathrm{CE}}(x)|\boldsymbol{X}\}\\
&=\left(1-\frac{1}{B}\right)\boldsymbol{e}_1^{\top}\boldsymbol{S}^{-1}\left\{\mathbb{C}(h,p,m_1)\boldsymbol{S}^{-1}\boldsymbol{e}_1-2m_1(x)m_1^{(p+1)}(x)\frac{h^{p+1}}{(p+1)!}\boldsymbol{c}_p\right\};
\end{aligned}
\tag{E.11}
$$

and, if p is even,

$$
\begin{aligned}
&\mathrm{Var}\left\{\hat{g}_{1,\mathrm{CE}}(x)|\boldsymbol{X}\right\}\\
&=\left(1-\frac{1}{B}\right)\boldsymbol{e}_1^{\top}\boldsymbol{S}^{-1}\bigg[\mathbb{C}(h,p,m_1)\boldsymbol{S}^{-1}\boldsymbol{e}_1-2m_1(x)\Big\{m_1^{(p+2)}(x)+\\
&\quad(p+2)m_1^{(p+1)}(x)\frac{f_x'(x)}{f_X(x)}\Big\}\frac{h^{p+2}}{(p+2)!}\tilde{\boldsymbol{c}}_p\bigg].
\end{aligned}
$$

Results for $\mathrm{Var}\{\hat{g}_{2,\mathrm{CE}}(x)|\boldsymbol{X}\}$ are given by the same expressions but with $m_2(x)$ and $m_2'(x)$ replacing $m_1(x)$ and $m_1'(x)$, respectively. Because U is assumed to be normal when formulating $\hat{g}_{1,\mathrm{CE}}(x)$ and $\hat{g}_{2,\mathrm{CE}}(x)$, we have $\beta=2$, $\beta_2=0$ (because $\beta_0=0$), and $\gamma=2/\sigma_u^2$ in these asymptotic variance results.

In the special case considered in the main article with $p = 1$, (E.11) reduces to

$$
\begin{aligned}
&\mathrm{Var}\{\hat{g}_{1,\mathrm{CE}}(x)|\mathbf{X}\} \\
&= \left(1 - \frac{1}{B}\right) \frac{h^2 m_1(x)}{f_X(x)} \left[(\mu_2 + \mu_4/\mu_2) f_X^{(2)}(x) m_1(x) - 2 f_X'(x) m_1'(x) \times \right. \\
&\quad \left. \left\{1 + \mu_2 \frac{f_X'(x) m_1(x)}{f_X(x) m_1'(x)} - \mu_2\right\}\right] + O_p(h^4) + O_p\left(\frac{\exp(2h^{-\beta}/\gamma)}{nh^{1-2\beta_2}}\right).
\end{aligned}
$$

Similarly,

$$
\begin{aligned}
&\mathrm{Var}\{\hat{g}_{2,\mathrm{CE}}(x)|\mathbf{X}\} \\
&= \left(1 - \frac{1}{B}\right) \frac{h^2 m_2(x)}{f_X(x)} \left[(\mu_2 + \mu_4/\mu_2) f_X^{(2)}(x) m_2(x) - 2 f_X'(x) m_2'(x) \times \right. \\
&\quad \left. \left\{1 + \mu_2 \frac{f_X'(x) m_2(x)}{f_X(x) m_2'(x)} - \mu_2\right\}\right] + O_p(h^4) + O_p\left(\frac{\exp(\sigma_u^2 h^{-2})}{nh}\right).
\end{aligned}
$$

E.3. *Covariance of $\hat{g}_{1,CE}(x)$ and $\hat{g}_{2,CE}(x)$*

Using similar tactics adopted in the previous two subsections, one can derive expectations involved in the covariance of $\hat{g}_{1,\mathrm{CE}}(x)$ and $\hat{g}_{2,\mathrm{CE}}(x)$. In particular,

$$
\begin{aligned}
&\mathrm{Cov}\left\{\hat{g}_{1,\mathrm{CE}}(x),\ \hat{g}_{2,\mathrm{CE}}(x)\big|\, X\right\} \\
&= \mathbb{E}\left\{\hat{g}_{1,\mathrm{CE}}(x)\hat{g}_{2,\mathrm{CE}}(x)\big|\, X\right\} - \mathbb{E}\left\{\hat{g}_{1,\mathrm{CE}}(x)\big|\, X\right\} \mathbb{E}\left\{\hat{g}_{2,\mathrm{CE}}(x)\big|\, X\right\} \\
&= \frac{1}{B}\mathbb{E}\left\{\hat{g}_{1,b}^*(x)\hat{g}_{2,b}^*(x)\Big|\, X\right\} + \left(1 - \frac{1}{B}\right)\mathbb{E}\left\{\hat{g}_{1,a}^*(x)\hat{g}_{2,b}^*(x)\Big|\, X\right\} - \\
&\quad \mathbb{E}\left\{\tilde{g}_1(x)|\, X\right\}\mathbb{E}\left\{\tilde{g}_2(x)|\, X\right\},\ \text{where } a \neq b, \\
&= \frac{1}{B}\mathbb{E}\left\{\tilde{g}_1(x)\tilde{g}_2(x)|\, X\right\} + \left(1 - \frac{1}{B}\right)\mathbb{E}\left\{\hat{g}_{1,a}^*(x)\hat{g}_{2,b}^*(x)\Big|\, X\right\} - \\
&\quad \mathbb{E}\left\{\tilde{g}_1(x)|\, X\right\}\mathbb{E}\left\{\tilde{g}_2(x)|\, X\right\}. \tag{E.12}
\end{aligned}
$$

The term in (E.12) as the products of two expectations can be elaborated using (E.1). The first expectation in (E.12) only concerns estimators in the absence of measurement error and can be derived using a similar strategy employed to derive (E.5) without involving error-contaminated covariates, which gives

$$
\begin{aligned}
&\mathbb{E}\left\{\tilde{g}_1(x)\tilde{g}_2(x)|\, X\right\} \\
&= m_1(x)m_2(x) + \frac{e_1^\top S^{-1}}{f_X^2(x)}\left(h f_X(x)\left\{-m_1(x)m_2(x)f_X'(x)\left(c_0 c_0^\top S^{-1} S_1 + \right.\right.\right. \\
&\quad \left. S_1 S^{-1} c_0 c_0^\top\right) + m_1(x)\tau_2'(x)c_1 c_0^\top + \tau_1'(x)m_2(x)c_0 c_1^\top\right\} + \frac{h^2}{2}\left[\tau_1(x)\tau_2^{(2)}(x)\times\right. \\
&\quad c_2 c_0^\top + \tau_1^{(2)}(x)\tau_2(x)c_0 c_2^\top + 2\tau_1'(x)\tau_2'(x)c_1 c_1^\top - 2 f_X'(x)\left\{m_1(x)\tau_2'(x)\times\right. \\
&\quad \left(S_1 S^{-1} c_1 c_0^\top + c_1 c_0^\top S^{-1} S_1\right) + \tau_1'(x)m_2(x)\left(S_1 S^{-1} c_0 c_1^\top + c_0 c_1^\top S^{-1} S_1\right)\right\} + \\
&\quad 2 m_1(x)m_2(x)\left\{f_X'(x)\right\}^2 S_1 S^{-1} c_0 c_0^\top S^{-1} S_1 - m_1(x)m_2(x)f_X(x)f_X^{(2)}(x)\times
\end{aligned}
$$

$$\left(c_0 c_0^\top S^{-1} S_2 + S_2 S^{-1} c_0 c_0^\top\right)\right] + o(h^2) + O_p\left(\frac{1}{h}\right)\right)S^{-1}e_1$$

$$\triangleq m_1(x)m_2(x) + e_1^\top S^{-1}\left\{\tilde{\mathbb{C}}(h,p) + O_p\left(\frac{1}{h}\right)\right\}S^{-1}e_1.$$

The second expectation in (E.12) is also similar to (E.5) and involves error-contaminated covariates in W_a^* and W_b^*, and thus has the same dominating terms as those in $\mathbb{E}\{\tilde{g}_1(x)\tilde{g}_2(x)|X\}$ above but with the last term being $O_p(\exp(2h^{-\beta}/\gamma)/(nh^{1-2\beta_2}))$ in place of $O_p(1/h)$. Putting the four expectations in (E.12) together reveals that, if p is odd,

$$\text{Cov}\left\{\hat{g}_{1,\text{CE}}(x),\,\hat{g}_{2,\text{CE}}(x)\big|X\right\}$$
$$= e_1^\top S^{-1}\left[\left\{\tilde{\mathbb{C}}(h,p) + O_p\left(\frac{1}{Bh}\right) + O_p\left(\frac{\exp(2h^{-\beta}/\gamma)}{nh^{1-2\beta_2}}\right)\right\}S^{-1}e_1 - \right. \tag{E.13}$$
$$\left.\left\{m_1^{(p+1)}(x)m_2(x) + m_1(x)m_2^{(p+1)}(x)\right\}\frac{h^{p+1}}{(p+1)!}c_p\right];$$

and, if p is even,

$$\text{Cov}\left\{\hat{g}_{1,\text{CE}}(x),\,\hat{g}_{2,\text{CE}}(x)\big|X\right\}$$
$$= e_1^\top S^{-1}\left(\left\{\tilde{\mathbb{C}}(h,p) + O_p\left(\frac{1}{Bh}\right) + O_p\left(\frac{\exp(2h^{-\beta}/\gamma)}{nh^{1-2\beta_2}}\right)\right\}S^{-1}e_1 - \right.$$
$$\left[m_1(x)\left\{m_2^{(p+2)}(x) + (p+2)m_2^{(p+1)}(x)\frac{f_X'(x)}{f_X(x)}\right\} + \right. \tag{E.14}$$
$$\left.\left. m_2(x)\left\{m_1^{(p+2)}(x) + (p+2)m_1^{(p+1)}(x)\frac{f_X'(x)}{f_X(x)}\right\}\right]\frac{h^{p+2}}{(p+2)!}\tilde{c}_p\right).$$

Let B/n tend to a positive constant as $n\to\infty$ so that the term $O_p(1/(Bh))$ in (E.13) and (E.14) is included in $O_p(\exp(2h^{-\beta}/\gamma)/(nh^{1-2\beta_2}))$. This also allows dropping the factor $(1-1/B)$ in the dominating terms in (E.11).

In the special case with $p=1$, following algebraic simplifications, we have

$$\text{Cov}\left\{\hat{g}_{1,\text{CE}}(x),\,\hat{g}_{2,\text{CE}}(x)\big|X\right\}$$
$$= \frac{h^2}{f_X^2(x)}\left(m_1(x)m_2(x)\left[(\mu_2 + \mu_4/\mu_2)f_X(x)f_X^{(2)}(x) - 2\mu_2\left\{f_X'(x)\right\}^2\right]\right.$$
$$\left. + \mu_2 f_X(x)\left\{f_X(x) - f_X'(x)\right\}\left\{m_1(x)m_2'(x) + m_1'(x)m_2(x)\right\}\right)$$
$$+ O(h^4) + O_p\left(\frac{1}{Bh}\right) + O_p\left(\frac{\exp(\sigma_u^2 h^{-2})}{nh}\right).$$

E.4. Asymptotic bias and variance of $\hat{m}_{CE}(x)$

Using the biases of $\hat{g}_{1,\text{CE}}(x)$ and $\hat{g}_{2,\text{CE}}(x)$ in Appendix E.1, their variances in Appendix E.2, and the covariance of them in Appendix E.3 in (D.27), we have, if p is odd,

$$\text{Bias}\left\{\hat{m}_{CE}(x)|X\right\}$$
$$= e_1^\top S^{-1}\left(\frac{h^{p+1}}{(p+1)!\left\{m_1^2(x) + m_2^2(x)\right\}}\left(m_2(x)m_1^{(p+1)}(x) - m_1(x)m_2^{(p+1)}(x) + \right.\right.$$

$$\frac{1}{m_1^2(x) + m_2^2(x)} \left[2m_1(x)m_2(x) \left\{ m_1(x)m_1^{(p+1)}(x) - m_2(x)m_2^{(p+1)}(x) \right\} + \right.$$

$$\left. \left\{ m_2^2(x) - m_1^2(x) \right\} \left\{ m_1^{(p+1)}(x)m_2(x) + m_1(x)m_2^{(p+1)}(x) \right\} \right] \right) c_p + o_p \left(h^{p+1} \right) -$$

$$\frac{1}{\left\{ m_1^2(x) + m_2^2(x) \right\}^2} \left[m_1(x)m_2(x) \left\{ \mathbb{C}(h, p, m_1) - \mathbb{C}(h, p, m_2) \right\} + \right.$$

$$\left. \left\{ m_2^2(x) - m_1^2(x) \right\} \left\{ \tilde{\mathbb{C}}(h, p) + O \left(\frac{\exp(2h^{-\beta}/\gamma)}{nh^{1-2\beta_2}} \right) \right\} \right] S^{-1} e_1 \right);$$

and, if p is even,

$$\text{Bias} \left\{ \hat{m}_{\text{CE}}(x) | X \right\}$$

$$= e_1^\top S^{-1} \left(\frac{h^{p+2}}{(p+2)! \left\{ m_1^2(x) + m_2^2(x) \right\}} \left(m_2(x) \left\{ m_1^{(p+2)}(x) + (p+2)m_1^{(p+1)}(x) \times \right. \right. \right.$$

$$\left. \frac{f_X'(x)}{f_X(x)} \right\} - m_1(x) \left\{ m_2^{(p+2)}(x) + (p+2)m_2^{(p+1)}(x) \frac{f_X'(x)}{f_X(x)} \right\} +$$

$$\frac{2 + m_1^2(x) - m_2^2(x)}{m_1^2(x) + m_2^2(x)} \left[m_1(x) \left\{ m_1^{(p+2)}(x) + (p+2)m_1^{(p+1)}(x) \frac{f_X'(x)}{f_X(x)} \right\} + \right.$$

$$\left. \left. m_2(x) \left\{ m_2^{(p+2)}(x) + (p+2)m_2^{(p+1)}(x) \frac{f_X'(x)}{f_X(x)} \right\} \right] \right) \tilde{c}_p + o_p \left(h^{p+2} \right) -$$

$$\frac{1}{\left\{ m_1^2(x) + m_2^2(x) \right\}^2} \left[m_1(x)m_2(x) \left\{ \mathbb{C}(h, p, m_1) - \mathbb{C}(h, p, m_2) \right\} + \right.$$

$$\left. \left\{ m_2^2(x) - m_1^2(x) \right\} \left\{ \tilde{\mathbb{C}}(h, p) + O_p \left(\frac{\exp(2h^{-\beta}/\gamma)}{nh^{1-2\beta_2}} \right) \right\} \right] S^{-1} e_1 \right).$$

In the special case with $p = 1$,

$$\text{Bias} \left\{ \hat{m}_{\text{CE}}(x) | X \right\}$$

$$= \frac{h^2 \mu_2}{2} \left\{ m^{(2)}(x) + 2m'(x) \frac{m_1(x)m_1'(x) + m_2(x)m_2'(x)}{m_1^2(x) + m_2^2(x)} \right\} +$$

$$\frac{h^2}{f_X(x) \left\{ m_1^2(x) + m_2^2(x) \right\}^2} \left[2(1 - \mu_2)f_X'(x)m_1(x)m_2(x)\{m_1(x)m_1'(x) \right.$$

$$- m_2(x)m_2'(x)\} + \mu_2\{f_X(x) - f_X'(x)\}\{m_1^2(x) - m_2^2(x)\}\{m_1(x)m_2'(x)$$

$$\left. + m_1'(x)m_2(x)\} \right] + O_p(h^4) + O_p \left(\frac{\exp(\sigma_u^2 h^{-2})}{nh} \right). \tag{E.15}$$

Using these bias results, along with the biases of $\hat{g}_{1,\text{CE}}(x)$ and $\hat{g}_{2,\text{CE}}(x)$ in Appendix E.1, their variances in Appendix E.2, and the covariance of them in Appendix E.3, in (D.31) reveals $\text{Var}\{\hat{m}_{\text{CE}}(x) | X\}$. In the special case with $p = 1$,

$$\text{Var}\{\hat{m}_{\text{CE}}(x) | X\}$$

$$= 2h^2 \left\{ (2\mu_2 - 1)f_X'(x)/f_X(x) - \mu_2 \right\} \frac{m_1(x)m_2(x) \left\{ m_1(x)m_2'(x) + m_1'(x)m_2(x) \right\}}{\left\{ m_1^2(x) + m_2^2(x) \right\}^2}$$

$$+ O_p(h^4) + O_p\left(\frac{\exp\left(\sigma_u^2 h^{-2}\right)}{nh}\right). \tag{E.16}$$

E.5. *Asymptotic normality of $\hat{g}_{1,CE}$ and $\hat{g}_{2,CE}$*

Our complex-error corrected estimator falls under a larger umbrella of Monte-Carlo corrected score estimators. To this end, [31] proved than these Monte-Carlo corrected score estimators are asymptotically identical to SIMEX estimators under the same model. Further, [38] showed that the SIMEX estimator applied to the nonparametric regression model with normal errors, and utilizing a local linear kernel, is asymptotically normally distributed. With this in mind, it can be shown that our method is asymptotically equivalent to a Monte-Carlo corrected score, specifically one that arises from the objective function

$$\mathcal{F} = \sum_{i=1}^{n} \left[Y_i - \alpha - \beta(X_i - x)\right]^2 K_h(X_i - x).$$

Thus the complex-error estimators, $\hat{g}_{1,\text{CE}}$ and $\hat{g}_{2,\text{CE}}$, are asymptotically equivalent to the SIMEX estimator, which is asymptotically normal.

Appendix F: Asymptotic properties of the one-step correction estimator $\hat{m}_{\text{OS}}(x)$

We consider the p-th order local polynomial weight used in $\hat{m}_{\text{OS}}(x)$ in this appendix, for $p \geq 0$. For this generalization, one needs to change "$\ell = 0, 1, 2, 3$" in **Condition O** stated in Section 4 to "$\ell = 0, 1, \ldots, 2p + 1$"; similarly, one needs to change "$\ell = 0, 1, 2$" in **Condition S** to "$\ell = 0, 1, \ldots, 2p$".

F.1. *The first two moments of $\hat{g}_{1,OS}(x)$ and $\hat{g}_{2,OS}(x)$*

Because the one-step correction estimator $\hat{g}_{\ell,\text{OS}}(x)$ for $g_\ell(x) = m_\ell(x) f_X(x)$ is precisely the estimator for, in notations used in Huang and Zhou [22], $B(x) = \mathbb{E}(Y|X = x) f_X(x)$ with a linear response Y, the bias and variance of $\hat{g}_{\ell,\text{OS}}(x)$ are readily available in Huang and Zhou [22]. In particular, recall that $S = (\mu_{r_1+r_2})_{0 \leq r_1, r_2 \leq p}$, $\boldsymbol{c}_p = (\mu_{p+1}, \ldots, \mu_{2p+1})^\top$, and $\tilde{\boldsymbol{c}}_p = (\mu_{p+2}, \ldots, \mu_{2p+2})^\top$, by Equation (A.5) in [22], we have for $\ell = 1, 2$,

$$\begin{aligned}
&\mathbb{E}\left\{\hat{g}_{\ell,\text{OS}}(x)\big| \boldsymbol{W}\right\} \\
&= \begin{cases} g_\ell(x) + (N_{\ell,p} * D)(x) + (M_{\ell,p} * D)(x)\dfrac{h^{p+1}}{(p+1)!} + o_p(h^{p+1}), & \text{if } p \text{ is odd,} \\ g_\ell(x) + (N_{\ell,p} * D)(x) + (M_{\ell,p} * D)(x)\dfrac{h^{p+2}}{(p+2)!} + o_p(h^{p+2}), & \text{if } p \text{ is even,} \end{cases}
\end{aligned} \tag{F.1}$$

where $N_{\ell,p}(w) = m_\ell^*(w) \sum_{r=1}^{p} f_W^{(r)}(w) \mu_r h^r / r!$, and, if p is odd,

$$M_{\ell,p}(w) = m_\ell^*(w) f_W^{(p+1)}(w) \mu_{p+1} + m_\ell^{*(p+1)}(w) f_W(w) \boldsymbol{e}_1^\top \boldsymbol{S}^{-1} \boldsymbol{c}_p, \tag{F.2}$$

and, if p is even,

$$\begin{aligned}
&M_{\ell,p}(w) \\
&= m_\ell^*(w) f_W^{(p+2)}(w) \mu_{p+2} + \left\{m_\ell^{*(p+2)}(w) f_W(w) + (p+2) m_\ell^{*(p+1)} f_W'(w)\right\} \boldsymbol{e}_1^\top \boldsymbol{S}^{-1} \tilde{\boldsymbol{c}}_p,
\end{aligned} \tag{F.3}$$

with the Fourier transform of $N_{\ell,p}(w)$ and $M_{\ell,p}(w)$ compactly supported on I_t, and $D(x) = (2\pi)^{-1} \int_{I_t} e^{-itx}/\phi_U(t)dt$. Because the Fourier transform of $M_{\ell,p}(w)$, $\phi_{M_{\ell,p}}(t)$, is compactly supported on I_t, the convolution of $M_{\ell,p}(\cdot)$ and $D(\cdot)$ is

$$
\begin{aligned}
(M_{\ell,p} * D)(x) &= \int M_{\ell,p}(w)D(x-w)dw \\
&= \frac{1}{2\pi} \int M_{\ell,p}(w) \int_{I_t} e^{-it(x-w)}/\phi_U(t)\, dt dw \\
&= \frac{1}{2\pi} \int_{I_t} e^{-itx} \frac{1}{\phi_U(t)} \int e^{itw} M_{\ell,p}(w)\, dw dt \\
&= \frac{1}{2\pi} \int_{I_t} e^{-itx} \frac{\phi_{M_{\ell,p}}(t)}{\phi_U(t)}\, dt.
\end{aligned}
$$

In other words, $(M_{\ell,p} * D)(x)$ is the same integral transform of $M_{\ell,p}(\cdot)$ appearing in Lemma 3.2 in the main article but expressed in the form of convolution, that is, $(M_{\ell,p} * D)(x) = \mathcal{T}_U(M_{\ell,p})(x)$ by adopting the integral transform $\mathcal{T}_U$ defined in Section 3.3 in the main article. Similarly, $(N_{\ell,p} * D)(x) = \mathcal{T}_U(N_{\ell,p})(x)$.

Define $S^* = (\eta(r_1, r_2))_{0 \le r_1, r_2 \le p}$. Then, according to Appendix B in [22], we have, for $\ell = 1, 2$,

$$
\text{Var}\left\{ \hat{g}_{\ell,\text{OS}}(x) \middle| W \right\} = e_1^\top S^{-1} S^* S^{-1} e_1 \frac{\sigma_\ell^{*2}(x) f_W(x)}{nh^{1+2\beta}} + o_p\left(\frac{1}{nh^{1+2\beta}} \right), \tag{F.4}
$$

if U is ordinary smooth, and $\text{Var}\left\{ \hat{g}_{\ell,\text{OS}}(x) \middle| W \right\}$ is bounded from above by

$$
e_1^\top S^{-1} S^{-1} e_1 \frac{C \sigma_\ell^{*2}(x) f_W(x) \exp(2h^{-\beta}/\gamma)}{f_X^2(x) nh^{1-2\beta_2}} + o_p\left(\frac{\exp(2h^{-\beta}/\gamma)}{nh^{1-2\beta_2}} \right), \tag{F.5}
$$

if U is super smooth, where $\beta_2 = \beta_0 I(\beta_0 < 0.5)$, $\sigma_1^{*2}(x) = \text{Var}(\sin\Theta|W = x) = \xi_1^*(x) - m_1^{*2}(x)$ and $\sigma_2^{*2}(x) = \text{Var}(\cos\Theta|W = x) = \xi_2^*(x) - m_2^{*2}(x)$.

Lastly, adopting the same strategy elaborated in Appendix B in [22], one can show that

$$
\text{Cov}\left\{ \hat{g}_{1,\text{OS}}(x), \hat{g}_{2,\text{OS}}(x) \middle| W \right\} = e_1^\top S^{-1} S^* S^{-1} e_1 \frac{\phi^*(x) f_W(x)}{nh^{1+2\beta}} + o_p\left(\frac{1}{nh^{1+2\beta}} \right), \tag{F.6}
$$

if U is ordinary smooth, and $\text{Cov}\left\{ \hat{g}_{1,\text{OS}}(x), \hat{g}_{2,\text{OS}}(x) \middle| W \right\}$ is bounded from above by

$$
e_1^\top S^{-1} S^{-1} e_1 \frac{C \phi^*(x) f_W(x) \exp(2h^{-\beta}/\gamma)}{f_X^2(x) nh^{1-2\beta_2}} + o_p\left(\frac{\exp(2h^{-\beta}/\gamma)}{nh^{1-2\beta_2}} \right), \tag{F.7}
$$

if U is super smooth, where $\phi^*(x) = \text{Cov}(\sin\Theta, \cos\Theta|W = x) = \psi^*(x) - m_1^*(x)m_2^*(x)$.

F.2. Asymptotic bias and variance of $\hat{m}_{OS}(x)$

By (D.27), with $g_\ell(x) = m_\ell(x) f_X(x)$, for $\ell = 1, 2$,

$$
\begin{aligned}
&\text{Bias}\left\{ \hat{m}_{\text{OS}}(x) \middle| W \right\} \\
&= \frac{1}{g_1^2(x) + g_2^2(x)} \left[g_2(x)\text{Bias}\left\{ \hat{g}_{1,\text{OS}}(x) \middle| W \right\} - g_1(x)\text{Bias}\left\{ \hat{g}_{2,\text{OS}}(x) \middle| W \right\} \right] -
\end{aligned}
$$

$$\frac{1}{\left\{g_1^2(x) + g_2^2(x)\right\}^2} \left(g_1(x)g_2(x) \left[\mathrm{Var}\left\{\hat{g}_{1,\mathrm{OS}}(x)\middle|\boldsymbol{W}\right\} - \mathrm{Var}\left\{\hat{g}_{2,\mathrm{OS}}(x)\middle|\boldsymbol{W}\right\}\right] +$$

$$\left\{g_2^2(x) - g_1^2(x)\right\} \mathrm{Cov}\left\{\hat{g}_{1,\mathrm{OS}}(x), \hat{g}_{2,\mathrm{OS}}(x)\middle|\boldsymbol{W}\right\}\right) + \text{lower-order terms},$$

where the lower-order terms refer to the non-dominating terms of $\mathbb{E}\{\hat{g}_{\ell,\mathrm{OS}}(x) - g_\ell(x)|\boldsymbol{W}\}$ indicated in (F.1). By (F.1), (F.4), and (F.6), if U is ordinary smooth,

$$\mathrm{Bias}\left\{\hat{m}_{\mathrm{OS}}(x)|\boldsymbol{W}\right\}$$

$$= \frac{1}{f_X(x)\left\{m_1^2(x) + m_2^2(x)\right\}}\left[m_2(x)\mathcal{T}_U(N_{1,p})(x) - m_1(x)\mathcal{T}_U(N_{2,p})(x)\right.$$

$$+ \left\{m_2(x)\mathcal{T}_U(M_{1,p})(x) - m_1(x)\mathcal{T}_U(M_{2,p})(x)\right\} \frac{h^{p+1+I(p \text{ is even})}}{(p+1+I(p \text{ is even}))!}\right]$$

$$+ o_p\left(h^{p+1+I(p \text{ is even})}\right) + \frac{\boldsymbol{e}_1^\top \boldsymbol{S}^{-1}\boldsymbol{S}^*\boldsymbol{S}^{-1}\boldsymbol{e}_1 f_W(x)}{nh^{1+2\beta}f_X^2(x)\left\{m_1^2(x) + m_2^2(x)\right\}^2} \times$$

$$\left[m_1(x)m_2(x)\left\{\sigma_1^{*2}(x) - \sigma_2^{*2}(x)\right\} + \phi^*(x)\left\{m_2^2(x) - m_1^2(x)\right\}\right] + o_p\left(\frac{1}{nh^{1+2\beta}}\right);$$

and, if U is super smooth, using (F.1), (F.5), and (F.7), we have

$$\mathrm{Bias}\left\{\hat{m}_{\mathrm{OS}}(x)|\boldsymbol{W}\right\}$$

$$= \frac{1}{f_X(x)\left\{m_1^2(x) + m_2^2(x)\right\}}\left[m_2(x)\mathcal{T}_U(N_{1,p})(x) - m_1(x)\mathcal{T}_U(N_{2,p})(x)\right.$$

$$+ \left\{m_2(x)\mathcal{T}_U(M_{1,p})(x) - m_1(x)\mathcal{T}_U(M_{2,p})(x)\right\} \frac{h^{p+1+I(p \text{ is even})}}{(p+1+I(p \text{ is even}))!}\right]$$

$$+ o_p\left(h^{p+1+I(p \text{ is even})}\right) + O_p\left(\frac{\exp(2h^{-\beta}/\gamma)}{nh^{1-2\beta_2}}\right).$$

In the special case we focus on in our study with a local linear weight, i.e., with $p = 1$, because $N_{\ell,1}(w) = m_\ell^*(w)f_W'(w)\mu_1 h = 0$, for $\ell = 1, 2$, we have, if U is ordinary smooth,

$$\mathrm{Bias}\left\{\hat{m}_{\mathrm{OS}}(x)|\boldsymbol{W}\right\}$$

$$= \frac{h^2\mu_2}{2f_X(x)\left\{m_1^2(x) + m_2^2(x)\right\}}\left\{m_2(x)\mathcal{T}_U(M_{1,1})(x) - m_1(x)\mathcal{T}_U(M_{2,1})(x)\right\} + o_p\left(h^2\right)$$

$$+ \frac{f_W(x)\eta(0,0)\left[m_1(x)m_2(x)\left\{\sigma_1^{*2}(x) - \sigma_2^{*2}(x)\right\} + \phi^*(x)\left\{m_2^2(x) - m_1^2(x)\right\}\right]}{nh^{1+2\beta}f_X^2(x)\left\{m_1^2(x) + m_2^2(x)\right\}^2}$$

$$+ o_p\left(\frac{1}{nh^{1+2\beta}}\right);$$

and, if U is super smooth,

$$\mathrm{Bias}\left\{\hat{m}_{\mathrm{OS}}(x)|\boldsymbol{W}\right\} = \frac{h^2\mu_2}{2f_X(x)\left\{m_1^2(x) + m_2^2(x)\right\}}\{m_2(x)\mathcal{T}_U(M_{1,1})(x)$$

$$- m_1(x)\mathcal{T}_U(M_{2,p})(x)\} + o_p\left(h^2\right) + O_p\left(\frac{\exp(2h^{-\beta}/\gamma)}{nh^{1-2\beta_2}}\right). \tag{F.8}$$

Since the dominating terms in these asymptotic bias do not depend on $\boldsymbol{W}$, they can be interpreted as the unconditional asymptotic dominating bias.

Using (F.1), (F.4), and (F.6) in (D.31) gives that, for ordinary smooth U,

$$
\begin{aligned}
&\mathbb{E}\left\{\hat{m}_{\text{OS}}^2(x)\middle|\boldsymbol{W}\right\} \\
&= m^2(x) + \frac{2m(x)}{f_X(x)\left\{m_1^2(x)+m_2^2(x)\right\}}\Bigg[m_2(x)\mathcal{T}_U(N_{1,p})(x) - m_1(x)\mathcal{T}_U(N_{2,p})(x) \\
&\quad + \left\{m_2(x)\mathcal{T}_U(M_{1,p})(x) - m_1(x)\mathcal{T}_U(M_{2,p})(x)\right\}\frac{h^{p+1+I(p\text{ is even})}}{(p+1+I(p\text{ is even}))!}\Bigg] \\
&\quad + o_p\left(h^{p+1+I(p\text{ is even})}\right) + \frac{2\boldsymbol{e}_1^\top \boldsymbol{S}^{-1}\boldsymbol{S}^*\boldsymbol{S}^{-1}\boldsymbol{e}_1 f_W(x)}{nh^{1+2\beta}f_X^2(x)\left\{m_1^2(x)+m_2^2(x)\right\}^2}\times \\
&\quad \left(m(x)m_1(x)m_2(x)\left\{\sigma_2^{*2}(x) - \sigma_1^{2*}(x)\right\} + \frac{1}{2}\left\{m_1^2(x)\sigma_2^{*2}(x) + m_2^2(x)\sigma_1^{*2}(x)\right\} \right. \\
&\quad \left. + \phi^*(x)\left[m(x)\left\{m_1^2(x)-m_2^2(x)\right\} - m_1(x)m_2(x)\right]\right) + o_p\left(\frac{1}{nh^{1+2\beta}}\right);
\end{aligned}
$$

and, for super smooth U,

$$
\begin{aligned}
&\mathbb{E}\left\{\hat{m}_{\text{OS}}^2(x)\middle|\boldsymbol{W}\right\} \\
&= m^2(x) + \frac{2m(x)}{f_X(x)\left\{m_1^2(x)+m_2^2(x)\right\}}\Bigg[m_2(x)\mathcal{T}_U(N_{1,p})(x) - m_1(x)\mathcal{T}_U(N_{2,p})(x) \\
&\quad + \left\{m_2(x)\mathcal{T}_U(M_{1,p})(x) - m_1(x)\mathcal{T}_U(M_{2,p})(x)\right\}\frac{h^{p+1+I(p\text{ is even})}}{(p+1+I(p\text{ is even}))!}\Bigg] \\
&\quad + o_p\left(h^{p+1+I(p\text{ is even})}\right) + O_p\left(\frac{\exp(2h^{-\beta}/\gamma)}{nh^{1-2\beta_2}}\right).
\end{aligned}
$$

It follows that, for ordinary smooth U,

$$
\begin{aligned}
&\text{Var}\left\{\hat{m}_{\text{OS}}(x)|\boldsymbol{W}\right\} \\
&= \frac{\boldsymbol{e}_1^\top \boldsymbol{S}^{-1}\boldsymbol{S}^*\boldsymbol{S}^{-1}\boldsymbol{e}_1 f_W(x)}{nh^{1+2\beta}f_X^2(x)\left\{m_1^2(x)+m_2^2(x)\right\}^2}\left(4m(x)m_1(x)m_2(x)\left\{\sigma_2^{*2}(x) - \sigma_1^{*2}(x)\right\}\right. \\
&\quad + 2\phi^*(x)\left[2m(x)\left\{m_1^2(x)-m_2^2(x)\right\} - m_1(x)m_2(x)\right] + m_1^2(x)\sigma_2^{*2}(x) \\
&\quad + m_2^2(x)\sigma_1^{*2}(x) - \frac{2}{f_X(x)\left\{m_1^2(x)+m_2^2(x)\right\}}\Big[m_1(x)m_2(x)\left\{\sigma_1^{*2}(x) - \sigma_2^{*2}(x)\right\} \\
&\quad \left. + \phi^*(x)\left\{m_2^2(x)-m_1^2(x)\right\}\Big]\left\{m_2(x)\mathcal{T}_U(N_{1,p})(x) - m_1(x)\mathcal{T}_U(N_{2,p})(x)\right\}\right) \\
&\quad + o_p\left(h^{p+1+I(p\text{ is even})}\right) + o_p\left(\frac{1}{nh^{1+2\beta}}\right),
\end{aligned}
$$

which reduces to, when $p = 1$,

$$
\begin{aligned}
&\mathrm{Var}\,\{\hat{m}_{\mathrm{OS}}(x)|\,\boldsymbol{W}\} \\
&= \frac{f_W(x)\eta(0,0)}{nh^{1+2\beta}f_X^2(x)\{m_1^2(x)+m_2^2(x)\}^2}\Big(m_1^2(x)\xi_2^*(x)+m_2^2(x)\xi_1^*(x) \\
&\quad 2m_1(x)m_2(x)\psi^*(x) - \big\{m_1(x)m_2^*(x)-m_1^*(x)m_2(x)\big\}^2 \\
&\quad + 4m(x)\Big[m_1(x)m_2(x)\big\{\sigma_2^{*2}(x)-\sigma_1^{*2}(x)\big\}+\phi^*(x)\{m_1^2(x) \\
&\quad - m_2^2(x)\}\Big]\Big) + o_p\left(\frac{1}{nh^{1+2\beta}}\right).
\end{aligned}
\tag{F.9}
$$

For super smooth U,

$$
\begin{aligned}
&\mathrm{Var}\,\{\hat{m}_{\mathrm{OS}}(x)|\boldsymbol{W}\} \\
&= O_p\left(\frac{\exp(2h^{-\beta}/\gamma)}{nh^{1-2\beta_2}}\right) - \frac{m_2(x)\mathcal{T}_U(N_{1,p})(x)-m_1(x)\mathcal{T}(N_{2,p})(x)}{f_X^2(x)\left\{m_1^2(x)+m_2^2(x)\right\}^2}\times \\
&\quad \Big[m_2(x)\mathcal{T}_U(N_{1,p})(x)-m_1(x)\mathcal{T}(N_{2,p})(x)+2\{m_2(x)\mathcal{T}_U(M_{1,p})(x) \\
&\quad - m_1(x)\mathcal{T}(M_{2,p})(x)\}\frac{h^{p+1+I(p\text{ is even})}}{(p+1+I(p\text{ is even}))!}\Big] + o_p\left(h^{p+1+I(p\text{ is even})}\right),
\end{aligned}
$$

which reduces to, when $p = 1$,

$$
\mathrm{Var}\,\{\hat{m}_{\mathrm{OS}}(x)|\,\boldsymbol{W}\} = O_p\left(\frac{\exp(2h^{-\beta}/\gamma)}{nh^{1-2\beta_2}}\right).
$$

Appendix G: Asymptotic normality of $\hat{m}.(x)$

Let "$\xrightarrow{\mathscr{L}}$" denote "converges in law to," i.e., "converges in distribution to," and let "$\xrightarrow{\mathscr{P}}$" denote "converges in probability to." A lemma is established next in preparation for proving the asymptotic normality of our proposed estimators for $m(x)$.

Lemma G.1. *Consider three sequences of random variables indexed by n, A_n, B_n, and $C_n = atan2[A_n, B_n]$. If $d_n(A_n - \mu_A, B_n - \mu_B)^\top \xrightarrow{\mathscr{L}} N(\boldsymbol{0}_2, \boldsymbol{\Sigma})$ as $n \to \infty$, where $\mu_A^2 + \mu_B^2 \neq 0$, $\boldsymbol{0}_2$ is a 2×1 vector of zeros, $\boldsymbol{\Sigma}$ is a 2×2 variance-covariance matrix, and d_n is a positive non-random sequence that diverges to ∞ as $n \to \infty$, then $d_n(C_n - atan2[\mu_A, \mu_B]) \xrightarrow{\mathscr{L}} N(0, (\mu_A^2\boldsymbol{\Sigma}[2,2]+\mu_B^2\boldsymbol{\Sigma}[1,1]-2\mu_A\mu_B\boldsymbol{\Sigma}[1,2])/(\mu_A^2+\mu_B^2)^2)$ as $n \to \infty$.*

Proof. Because $d_n(A_n - \mu_A) \xrightarrow{\mathscr{L}} N(0, \boldsymbol{\Sigma}[1,1])$, where $d_n \to \infty$ as $n \to \infty$, by Theorem 2.3.4 in [25], $A_n \xrightarrow{\mathscr{P}} \mu_A$. Similarly, $B_n \xrightarrow{\mathscr{P}} \mu_B$. By the continuous mapping theorem [25, Theorem 2.1.4], $(A_n - \mu_A)^\ell \xrightarrow{\mathscr{P}} 0$ and $(B_n - \mu_B)^\ell \xrightarrow{\mathscr{P}} 0$, for $\ell \in \mathbb{N}$, as $n \to \infty$.

Using a Taylor expansion of C_n around $(A_n, B_n) = (\mu_A, \mu_B)$, we have

$$
d_n(C_n - \mathrm{atan2}[\mu_A, \mu_B]) \tag{G.1}
$$

$$
= \frac{1}{\mu_A^2 + \mu_B^2}\{\mu_B d_n(A_n - \mu_A) - \mu_A d_n(B_n - \mu_B)\} \tag{G.2}
$$

$$-\frac{1}{(\mu_A^2+\mu_B^2)^2}\left[\mu_A\mu_B\left\{d_n(A_n-\mu_A)^2-d_n(B_n-\mu_B)^2\right\}+(\mu_B^2-\mu_A^2)d_n(A_n-\mu_A)(B_n-\mu_B)\right] \tag{G.3}$$

$$+O_p\{d_n(A_n-\mu_A)^3\}+O_p\{d_n(B_n-\mu_B)^3\}.$$

By Slutsky's theorem, (G.2), which is equal to $(\mu_B,-\mu_A)d_n(A_n-\mu_A,B_n-\mu_B)^\top/(\mu_A^2+\mu_B^2)$, converges in distribution to $N(0,(\mu_B,-\mu_A)\Sigma(\mu_B,-\mu_A)^\top/(\mu_A^2+\mu_B^2)^2)$. In (G.3), since $d_n(A_n-\mu_A)^2=d_n(A_n-\mu_A)\times(A_n-\mu_A)$, with $d_n(A_n-\mu_A)\xrightarrow{\mathscr{L}}N(0,\Sigma[1,1])$ and $A_n-\mu_A\xrightarrow{\mathscr{P}}0$, we have $d_n(A_n-\mu_A)^2\xrightarrow{\mathscr{P}}0$ by Slutsky's theorem. Similarly, $d_n(B_n-\mu_B)^2\xrightarrow{\mathscr{P}}0$ and $d_n(A_n-\mu_A)(B_n-\mu_B)\xrightarrow{\mathscr{P}}0$. Hence, (G.3) and the remainder terms following (G.3) all converge in probability to zero. Putting these together and using Slutsky's theorem again, we establish the limiting distribution of $d_n(C_n-\text{atan2}[\mu_A,\mu_B])$. □

Now return to our context of estimating the circular mean function $m(x)$. By letting $A_n=\hat{g}_{1,\cdot}(x)$ and $B_n=\hat{g}_{2,\cdot}(x)$ in Lemma G.1, we have $C_n=\hat{m}_\cdot(x)$, where "$\cdot$" refers to one of the acronyms ("DK," "CE," and "OS") relating to our proposed estimators for $m(x)$. According to Lemma G.1, the asymptotic normality of our proposed estimators follows directly from the established asymptotic normality of $\hat{g}_{1,\cdot}(x)$ and $\hat{g}_{2,\cdot}(x)$. For example, by the asymptotic normality of the one-step correction estimators established in [22] for $\hat{g}_{1,\text{OS}}(x)$ and $\hat{g}_{2,\text{OS}}(x)$, we can immediately conclude that $d_n\{\hat{m}_{\text{OS}}(x)-\text{atan2}[m_1(x)f_X(x),m_2(x)f_X(x)]\}=d_n\{\hat{m}_{\text{OS}}(x)-m(x)\}$ converges in distribution to a mean-zero normal distribution as $n\to\infty$, where $d_n=\sqrt{n}h^{1/2+\beta}$ if U is ordinary smooth of order β, and $d_n=\sqrt{n}h^{1-\beta_2}\exp(-h^{-\beta}/\gamma)$ if U is super smooth of order β. Similar conclusions for $\hat{m}_{\text{DK}}(x)$ and $\hat{m}_{\text{CE}}(x)$ can be established by the asymptotic normality of $\hat{g}_{\ell,\text{DK}}(x)$ according to [8] and that of $\hat{g}_{\ell,\text{CE}}(x)$ shown in Section E.5.

Appendix H: Mean integrated square error and optimal bandwidths

A commonly employed strategy for choosing the bandwidth in local polynomial estimation is through the minimization of the mean integrated square error (MISE),

$$\text{MISE}=\int_{\mathbb{R}}\left([\text{Bias}\{\hat{m}_\cdot(x)|\boldsymbol{X}\}]^2+\text{Var}\{\hat{m}_\cdot(x)|\boldsymbol{X}\}\right)w(x)\,dx, \tag{H.1}$$

following the definition in Fan and Gijbels [Section 2.1, 15], where $w(x)\geq0$ is a weight function that we set to one in our study, and $\hat{m}_\cdot(x)$ generically refers to one of our proposed estimator. With the asymptotic bias and variance of our proposed estimators in the presence of ordinary smooth measurement error derived, the asymptotic MISE (AMISE) of $\hat{m}_{\text{DK}}(x)$ and $\hat{m}_{\text{OS}}(x)$ in this case can be used to find an asymptotically optimal bandwidth. In the presence of super smooth measurement error, we only have the order for some upper bound of the asymptotic variance of $\hat{m}_{\text{DK}}(x)$ and $\hat{m}_{\text{OS}}(x)$, which is less useful for optimal bandwidth derivation based on AMISE. Even though the dominating bias and dominating variance of $\hat{m}_{\text{CE}}(x)$ presented in Theorems 4.1 and 4.2 include elaborated terms of order $O(h^2)$, they also implicitly depend on terms of order $O_p(\exp(\sigma_u^2 h^{-2})/(nh))$ that prohibit one from recovering a useful AMISE for this estimator. In what follows, we present the AMISEs of $\hat{m}_{\text{DK}}(x)$ and $\hat{m}_{\text{OS}}(x)$, and the corresponding optimal bandwidths when local linear weights are used while assuming ordinary smooth measurement errors.

By Theorem 4.1, the dominating bias of $\hat{m}_{\mathrm{DK}}(x)$ and $\hat{m}_{\mathrm{OS}}(x)$ are of the same form given by $\mathbb{DB}(x)h^2 + \mathbb{DB}^*(x)/(nh^{1+2\beta})$, where $\mathbb{DB}(x)$ is the dominating terms in $\hat{m}_{\mathrm{DK}}(x)$ or $\hat{m}_{\mathrm{OS}}(x)$ that multiply h^2, which depends on functionals of $m(x)$ and μ_2, and $\mathbb{DB}^*(x)$ is the dominating terms in these estimators that multiply $1/(nh^{1+2\beta})$, which depends on naive counterparts of functionals of $m(x)$ besides these functionals. By Theorem 4.2, the dominating variance of $\hat{m}_{\mathrm{DK}}(x)$ and $\hat{m}_{\mathrm{OS}}(x)$ also share a common form given by $\mathbb{DV}(x)/(nh^{1+2\beta})$, with $\mathbb{DV}(x)$ involving many of the aforementioned functionals. These dominating terms, $\mathbb{DB}(x)$, $\mathbb{DB}^*(x)$, and $\mathbb{DV}(x)$, are free of n and h.

Using these generic forms of the dominating bias and dominating variance of an estimator in (H.1), we have

$$\mathrm{AMISE} = \int_{\mathbb{R}} \left[\left\{ \mathbb{DB}(x)h^2 + \frac{\mathbb{DB}^*(x)}{nh^{1+2\beta}} \right\}^2 + \frac{\mathbb{DV}(x)}{nh^{1+2\beta}} \right] \, dx.$$

Using the dominating terms in the above AMISE, we find an asymptotic optimal h given by

$$h_{\mathrm{Optimal}} = \left[\frac{(1+2\beta)}{4n} \frac{\int_{\mathbb{R}} \mathbb{DV}(x) \, dx}{\int_{\mathbb{R}} \{\mathbb{DB}(x)\}^2 \, dx} \right]^{1/(5+2\beta)}. \tag{H.2}$$

Using this asymptotic optimal bandwidth in the bias results in Theorems 4.1 and 4.2 gives the bias of $\hat{m}_{\mathrm{DK}}(x)$ and $\hat{m}_{\mathrm{OS}}(x)$ both in the order of $O(n^{-2/(5+2\beta)})$, whereas the variance of them are in the order of $O_p(n^{-4/(5+2\beta)})$ when U is ordinary smooth.

Appendix I: Using naive bandwidths in a non-naive estimator

Focusing on the deconvoluting kernel estimator, $\hat{m}_{\mathrm{DK}}(x)$, we conducted a simulation study to compare the naive 5-fold cross-validation bandwidth selection, which ignores measurement errors in the naive loss function, and our proposed CV-SIMEX, with the optimal bandwidth h_{opt} defined in Section 6.1 serving as a reference. In this experiment, we have $m(x) = 2\mathrm{atan}(1/x)$ as the regression function, with $X \sim N(0, 4)$ and $U \sim N(0, 1)$, yielding $\lambda = 0.8$. Figure I.1 presents the empirical risk of $\hat{m}_{\mathrm{DK}}(x)$ when different bandwidths are used based on 100 Monte Carlo replicates at each considered sample size. The empirical risk is computed based on fitted values of $m(x)$ at a sequence of x ranging from -3 to 3 at increments of 0.1. The improvement in the estimation when using a bandwidth chosen by CV-SIMEX is evident compared to when one implements a naive CV. As n increases, CV-SIMEX becomes closer in performance to the optimal choice of bandwidth.
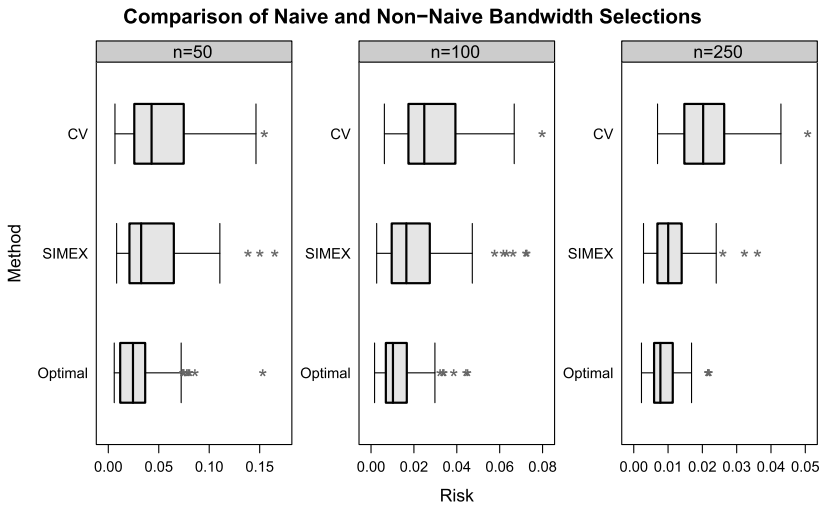
Fig I.1. *Boxplots of empirical risk across 100 Monte Carlo replicates at each level of* $n \in \{50, 100, 250\}$ *for* $\hat{m}_{DK}(x)$ *when three bandwidth selection methods are used: a naive cross-validation (CV), the proposed CV-SIMEX (SIMEX), and the optimal bandwidth (Optimal).*

References

[1] Bailey, D. H. and Swarztrauber, P. B. (1994). A fast method for the numerical evaluation of continuous fourier and laplace transforms. *Siam Journal on Scientific Computing*, 15(5):1105–1110. MR1289155

[2] Batschelet, E. (1981). Circular statistics in biology. *ACADEMIC PRESS, 111 FIFTH AVE., NEW YORK, NY 10003, 1981, 388*. MR0659065

[3] Birkett, P. J., Vanak, A. T., Muggeo, V. M., Ferreira, S. M., and Slotow, R. (2012). Animal perception of seasonal thresholds: changes in elephant movement in relation to rainfall patterns. *PloS one*, 7(6):e38363.

[4] Buonaccorsi, J. P. (2010). *Measurement error: models, methods, and applications*. CRC press. MR2682774

[5] Carroll, R. J., Ruppert, D., Stefanski, L. A., and Crainiceanu, C. M. (2006). *Measurement Error in Nonlinear Models: A Modern Perspective*. CRC Press. MR2243417

[6] Comte, F. and Kappus, J. (2015). Density deconvolution from repeated measurements without symmetry assumption on the errors. *Journal of Multivariate Analysis*, 140:31–46. MR3372551

[7] Delaigle, A. (2014). Nonparametric kernel methods with errors-in-variables: constructing estimators, computing them, and avoiding common mistakes. *Australian & New Zealand Journal of Statistics*, 56(2):105–124. MR3226432

[8] Delaigle, A., Fan, J., and Carroll, R. (2009). A design-adaptive local polynomial estimator for the errors-in-variables problem. *Journal of the American Statistical Association*, 104(485):348–359. MR2504382

[9] Delaigle, A. and Hall, P. (2008). Using SIMEX for smoothing-parameter choice in errors-in-variables problems. *Journal of the American Statistical Association*, 103(481):280–287. MR2394636

[10] Delaigle, A., Hall, P., and Meister, A. (2008). On deconvolution with repeated measurements. *The Annals of Statistics*, 36(2):665. MR2396811

[11] Di Marzio, M., Fensore, S., and Taylor, C. C. (2023). Kernel regression for errors-in-variables problems in the circular domain. *Statistical Methods & Applications*, pages 1–21. MR4662900

[12] Di Marzio, M., Panzera, A., and Taylor, C. C. (2013). Non-parametric regression for circular responses. *Scandinavian Journal of Statistics*, 40(2):238–255. MR3066413

[13] Dong, H., Otsu, T., and Taylor, L. (2023). Bandwidth selection for nonparametric regression with errors-in-variables. *Econometric Reviews*, 42(4):393–419. MR4593838

[14] Fan, J. (1991). Asymptotic normality for deconvolution kernel density estimators. *Sankhyā: The Indian Journal of Statistics, Series A*, 53(1):97–110. MR1177770

[15] Fan, J. and Gijbels, I. (1996). *Local Polyonomial Modelling and Its Applications*. Chaman And Hall, 1st edition. MR1383587

[16] Fisher, N. I. and Lee, J. (1992). Regression models for an angular response. *Biometrics*, 48(48):665–677. MR1187598

[17] Fuller, W. A. (2009). *Measurement error models*. John Wiley & Sons. MR2301581

[18] García-Portugués, E., Crujeiras, R. M., and González-Manteiga, W. (2013). Exploring wind direction and SO_2 concentration by circular–linear density estimation. *Stochastic Environmental Research and Risk Assessment*, 27:1055–1067.

[19] Gill, J. and Hangartner, D. (2010). Circular data in political science and how to handle it. *Political Analysis*, 18(3):316–336.

[20] Gould, A. L. (1969). A regression technique for angular variates. *Biometrics*, 25(4):683–700.

[21] Hodel, F. H. and Fieberg, J. R. (2022). Circular–linear copulae for animal movement data. *Methods in Ecology and Evolution*, 13(5):1001–1013.

[22] Huang, X. and Zhou, H. (2017). An alternative local polynomial estimator for the error-in-variables problem. *Journal of Nonparametric Statistics*, 29:301–325. MR3635015

[23] Johnson, R. A. and Wehrly, T. (1978). Some angular-linear distributions and related regression models. *Journal of the American Statistical Association*, 73:602–606. MR0514163

[24] Kamisan, N., Hussin, A., Zubairi, Y., and Hassan, S. (2011). Distribution of wind direction recorded at maximum wind speed: A case study of Malaysian wind data for. *International Journal of Physical Sciences*, 6(7):1840–1850.

[25] Lehmann, E. L. (1999). *Elements of Large-Sample Theory*. Springer New York, 1st edition. MR1663158

[26] Levin, B. Y. (1996). *Lectures on entire functions*, volume 150. American Mathematical Soc. MR1400006

[27] Lewis, B., Onder, N., and Prudil, A. (2021). *Advanced Mathematics for Engineering Students: The Essential Toolbox*. Butterworth-Heinemann.

[28] Mardia, K. V. and Jupp, P. E. (2000). *Directional statistics*, volume 2. Wiley Online Library. MR1828667

[29] Meilán-Vila, A., Francisco-Fernández, M., Crujeiras, R. M., and Panzera, A. (2020). Nonparametric multivariate regression estimation for circular responses. *arXiv preprint arXiv:2001.10317*. MR4297272

[30] Nadaraya, E. (1964). On estimating regression. *Theory of Probability and Its Applications*, 9(1):141–142.

[31] Novick, S. (2000). *Parametric Modeling in the Presence of Measurement Error: Monte Carlo Corrected Scores*. PhD thesis, University of North Carolina. MR2702759

[32] Novick, S. J. and Stefanski, L. A. (2002). Corrected score estimation via complex variable simulation extrapolation. *Journal of the American Statistical Association*, 97(458):472–481. MR1941464

[33] Paine, P. J., Preston, S., Tsagris, M., and Wood, A. T. (2020). Spherical regression models with general covariates and anisotropic errors. *Statistics and Computing*, 30:153–165. MR4057477

[34] Pirooz, A. A. S., Flay, R. G., Minola, L., Azorin-Molina, C., and Chen, D. (2020). Effects of sensor response and moving average filter duration on maximum wind gust measurements. *Journal of Wind Engineering and Industrial Aerodynamics*, 206:104354.

[35] Rivest, L.-P., Duchesne, T., Nicosia, A., and Fortin, D. (2016). A general angular regression model for the analysis of data on animal movement in ecology. *Journal of the Royal Statistical Society: Series C: Applied Statistics*, 65(3):445–463. MR3470586

[36] Scealy, J. and Wood, A. T. (2019). Scaled von Mises–Fisher distributions and regression models for paleomagnetic directional data. *Journal of the American Statistical Association*. MR4047280

[37] Scott, D. W. (2015). *Multivariate density estimation: theory, practice, and visualization*. John Wiley & Sons. MR3329609

[38] Song, W., Ayub, K., and Shi, J. (2023). Extrapolation estimation for nonparametric regression with measurement error. *Scandinavian Journal of Statistics*, pages 1–31. MR4705259

[39] Stefanski, L., Novick, S., and Devanarayan, V. (2005). Estimating a nonlinear function of a normal mean. *Biometrics*, 92:732–736. MR2202659

[40] Stefanski, L. A. (1989). Unbiased estimation of a nonlinear function a normal mean with application to measurement error models. *Communications in Statistics-Theory and Methods*, 18(12):4335–4358. MR1046712

[41] Stefanski, L. A. and Carroll, R. J. (1990). Deconvolving kernel density estimators. *Statistics*, 21(2):169–184. MR1054861

[42] TNRCC (2008). Texas Natural Resources Conservation Commission. BAMEX: Texas Natural Resources Conservation Commission. Version 1.0. UCAR/NCAR - Earth Observing Laboratory. https://doi.org/10.26023/ZQEZ-ENSF-T09 Accessed 03 Oct 2023.

[43] Wang, F. and Gelfand, A. E. (2013). Directional data analysis under the general projected normal distribution. *Statistical methodology*, 10(1):113–127. MR2974815

[44] Woolsey, N. and Huang, X. (2024). Debiased inference in error-in-variable problems with non-Gaussian measurement error. *arXiv preprint arXiv:2505.02754*.

[45] Yi, G., Gustafson, P., and Delaigle, A. (2021). *Handbook of Measurement Error Models*. Champan and Hall/CRC.

[46] Yi, G. Y. (2017). *Statistical analysis with measurement error or misclassification: strategy, method and application*. Springer. MR3676914