

Nonparametric curve estimation in measurement error problems with conditionally heteroscedastic variances

AUORE DELAIGLE^{1,a}, ALEXANDER MEISTER^{2,b} and JIYANG ZHANG^{3,c}

¹*School of Mathematics and Statistics, University of Melbourne, Australia, aurored@unimelb.edu.au*

²*Institut für Mathematik, Universität Rostock, Germany, alexander.meister@uni-rostock.de*

³*School of Mathematics and Statistics, University of Melbourne, Australia, jiyangzhang@student.unimelb.edu.au*

We consider the problem of estimating the density f_X of a latent variable X using replicated observations contaminated by additive errors. Unlike the classical setting, these errors depend on X , which makes standard Fourier deconvolution techniques inappropriate, and we suggest instead a basis expansion approach. Using a basis of Legendre polynomials, we show that it is possible to express the unknown coefficients of the expansion in terms of invertible equations of moments of the observed data. We deduce a nonparametric estimator of f_X which attains optimal minimax convergence rates in the case where the errors are conditionally Gaussian. To implement our density estimator in practice, we introduce data-driven approaches for choosing a truncating parameter and computing an upper bound to the support of f_X . We illustrate the numerical performance of our estimator on simulated examples and on replicated data from the National Health and Nutrition Examination Survey. We show how to use the same ideas in related errors-in-variables regression problems and to estimate the variance function.

MSC2020 subject classifications: Primary 62G07; secondary 62G08

Keywords: Contaminated data; deconvolution; inverse problems; minimax convergence rates; non-standard errors-in-variables

1. Introduction

The problem of estimating a curve nonparametrically from data observed with measurement errors has received a lot of attention in the literature. In the classical setting, for the n individuals in the sample, we observe J noisy replicates of a variable X in the form

$$W_{i,j} = X_i + \varepsilon_{i,j}, \quad i = 1, \dots, n, \quad j = 1, \dots, J, \quad (1.1)$$

where the unobserved X_i 's are i.i.d. with unknown density f_X , and independent of the unobserved i.i.d. $\varepsilon_{i,j}$'s with even density f_ε . The $\varepsilon_{i,j}$'s are often referred to as measurement errors, but model (1.1) is also used when X_i is a latent variable measured indirectly through the $W_{i,j}$'s. A typical example is when X_i is the long-term trend of a quantity that changes every day (e.g. pollutant concentration, activity level, food intake) and the $W_{i,j}$'s are measurements of that quantity taken on J days.

In model (1.1), the density f_W of the $W_{i,j}$'s satisfies $f_W(w) = f_X * f_\varepsilon(w) \equiv \int f_X(x) f_\varepsilon(w - x) dx$. Letting ϕ_X , ϕ_W and ϕ_ε denote the characteristic functions of X_i , $W_{i,j}$ and $\varepsilon_{i,j}$, we deduce that

$$\text{if } \phi_\varepsilon(t) \neq 0 \text{ for all } t \in \mathbb{R}, \text{ then } \phi_X = \phi_W / \phi_\varepsilon \text{ and } f_X(x) = (2\pi)^{-1} \int e^{-itx} \phi_W(t) / \phi_\varepsilon(t) dt. \quad (1.2)$$

Motivated by this result, nonparametric estimation procedures of X -related functions are often based on Fourier inversion methods known as deconvolution techniques. See e.g. [7] and [36], who constructed a

deconvolution kernel estimator of f_X by deconvolving the kernel density estimator (KDE) of f_W , [18] and [10] for errors-in-variables local polynomial regression estimators, [14] who estimated f_ε from the $W_{i,j}$'s, or [4] for estimating the error variance under constraints on f_X . See [29], [9], [39] and the references therein for more comprehensive overviews.

While those techniques can be extended to the case where the $\varepsilon_{i,j}$'s are independent of the X_i 's but not identically distributed [see 15,16,26], it has been argued that independence between X_i and $\varepsilon_{i,j}$ is too strong and could be relaxed by assuming that $\varepsilon_{i,j} = \sigma(X_i) U_{i,j}$, with σ an unknown smooth positive function and $U_{i,j} \sim f_U$, so that $f_W(w) = \int f_X(x) f_U\{(w-x)/\sigma(x)\} / \sigma(x) dx$. [35], and their improved version in [33], suggested Bayesian semiparametric estimators of f_X designed for this relaxed model, using splines or normal mixtures with priors on the coefficients. Their methods work well in the normal mixture examples they considered but are generally not consistent.

Constructing a consistent nonparametric estimator of f_X is challenging in this case: since $\varepsilon_{i,j} = \sigma(X_i) U_{i,j}$ is not independent of X_i , (1.2) does not hold, and ϕ_X cannot be disentangled from ϕ_ε following the standard deconvolution approach for model (1.1). Another common way to obtain a nonparametric estimator of a density f_X that belongs to a space $\mathcal{F}$ is to use a series expansion $f_X = \sum_{k=0}^{\infty} f_{X,k} \varphi_k$, where the φ_k 's are known basis functions of $\mathcal{F}$ and the $f_{X,k}$'s are unknown coefficients. Then if we can construct consistent estimators of the $f_{X,k}$'s, we can deduce a consistent estimator of f_X . However, when $\varepsilon_{i,j}$ depends on X_i , the $f_{X,k}$'s often depend in a complex way on f_X , $\sigma(X)$ and f_U , and it is not clear how to estimate them from the $W_{i,j}$'s. We show that if we take a basis of scaled Legendre polynomials, then the $f_{X,k}$'s can be obtained by inverting equations whose only unknowns are moments estimable from the $W_{i,j}$'s, and use this result to develop a consistent nonparametric estimator of f_X . While our main focus is on f_X , we show that the same ideas can be used to construct a nonparametric estimator of the function σ and that they can be applied to the errors-in-variables regression problem.

Related models and their extensions to the regression and multivariate settings have also received increasing attention in the literature; see e.g. [34] for a Bayesian procedure in the multivariate case, and [32] and [25] for Bayesian and semiparametric procedures in the regression case. Model (2.1) also falls into the more general class of so-called nonclassical measurement errors which have been analyzed in the regression context (i.e. when an additional variable related to X is observed) in, for example, [22].

Other different but related problems include the one from [2], who consider a scaled version of the model in the case where $X \geq 0$ and $\sigma(x) = \sqrt{x}$ is known, and the ones from [1] and [3], who consider a multiplicative error model. There also exist general approaches for measurement errors when the data come from certain algebraic groups, see e.g. [24] or [23].

This paper is organised as follows. We introduce our model, assumptions and basis expansions in Section 2. In Section 3 we describe our strategy for constructing a nonparametric estimator of the density f_X , and establish the asymptotic theory for its statistical risk in Section 4. In Section 5 we suggest a data-driven procedure for implementing our estimator, and in Section 6, we suggest an approach for computing it when the support of f_X is unknown. Numerical simulations and application to data from the National Health and Nutrition Examination Survey (NHANES) are provided in Section 7. In Section 8, our density estimator is extended to the case where we have more than two replicates per individual. We also show that a similar procedure can be used to estimate the function σ and in related errors-in-variables regression models. Proofs and technical details are deferred to appendices.

2. Model and challenges

We consider nonparametric estimation of the unknown density f_X of a continuous variable X , but we only observe repeated contaminated data $W_{i,j}$ coming from the model

$$W_{i,j} = X_i + \varepsilon_{i,j}, \quad \varepsilon_{i,j} = \sigma(X_i) U_{i,j}, \quad U_{i,j} \sim f_U, \quad \text{for } i = 1, \dots, n, \quad j = 1, 2, \quad (2.1)$$

with σ an unknown smooth positive function, the unobserved X_i 's and $U_{i,j}$'s totally independent, $X_i \sim f_X$, f_X and σ^2 square integrable and f_U known and symmetric around 0. We let W , X and U denote generic versions of $W_{i,j}$, X_i and $U_{i,j}$.

The dependence between $\varepsilon_{i,j} = \sigma(X_i) U_{i,j}$ and X_i increases the difficulty of constructing a non-parametric estimator of f_X from the $W_{i,j}$'s. In particular, we cannot apply standard deconvolution techniques since they require independence of X_i and $\varepsilon_{i,j}$. For example, under (1.1), the standard deconvolution KDE of [7] and [36] discussed in the introduction relies on (1.2), which does not hold under (2.1), and it is not clear how to/if one could find an inversion formula useful for a kernel estimator.

An alternative is to use an orthogonal series expansion, which we describe next for square integrable densities. For any continuous random variable T , let $\mathcal{S}_T$ denote the support of its density f_T and assume that $f_X \subseteq L_2(\mathcal{I})$, the set of square integrable functions on an interval $\mathcal{I} \supseteq \mathcal{S}_X$. Letting $\varphi_0, \varphi_1, \dots$ denote an orthonormal basis of $L_2(\mathcal{I})$, write $f_X = \sum_{k=0}^{\infty} f_{X,k} \varphi_k$, where the $f_{X,k}$'s are unknown coefficients. A series estimator of f_X is defined by $\hat{f}_X = \sum_{k=0}^K \hat{f}_{X,k} \varphi_k$, where $K = K(n) \rightarrow \infty$ as $n \rightarrow \infty$ is a truncating parameter and the $\hat{f}_{X,k}$'s are consistent estimators of the $f_{X,k}$'s.

Constructing the $\hat{f}_{X,k}$'s is easy in the error-free case: since $f_{X,k} = \int \varphi_k(x) f_X(x) dx = E\{\varphi_k(X)\}$, we can simply take $\hat{f}_{X,k} = n^{-1} \sum_{i=1}^n \varphi_k(X_i)$. However, they cannot be computed in the error case since the X_i 's are not available. In the classical error model at (1.1), [21] showed that, if f_X is compactly supported and has been rescaled so that $\mathcal{S}_X \subseteq \mathcal{I} = [0, \pi]$, taking the φ_k 's to be sequences of $\sin(r \cdot)$ and $\cos(r \cdot)$ for $r = 1, 2, \dots$, one can estimate the $f_{X,k}$'s consistently from the $W_{i,j}$'s. For that particular basis, the $f_{X,k}$'s are sequences of $\theta_{1r} = \pi^{-1} E\{\cos(rX)\}$ and $\theta_{2r} = \pi^{-1} E\{\sin(rX)\}$, for $r = 1, 2, \dots$ and since in their case $\varepsilon_{i,j}$ is independent of X_i , then $E\{\sin(rW)\} = \pi \theta_{2r} E\{\cos(r\varepsilon)\}$, $E\{\cos(rW)\} = \pi \theta_{1r} E\{\cos(r\varepsilon)\}$ and they deduce estimators of θ_{1r} and θ_{2r} from the empirical versions of the moments $E\{\sin(rW)\}$ and $E\{\cos(rW)\}$. However, in our model at (2.1) and with that basis, we have $E\{\sin(rW)\} = E[\sin(rX) \cos\{rU\sigma(X)\}] + E[\cos(rX) \sin\{rU\sigma(X)\}]$ and $E\{\cos(rW)\} = E[\cos(rX) \cos\{rU\sigma(X)\}] - E[\sin(rX) \sin\{rU\sigma(X)\}]$, from which we can extract neither θ_{1r} nor θ_{2r} .

This example illustrates that a challenge in our case is to find a basis for which we can find an inversion formula expressing the $f_{X,k}$'s in terms of the distribution of W . [35] have considered a basis expansion of this type (albeit non orthogonal); they used a spline basis with a fixed K , a method which they refer to as flexible parametric. However, it is not clear how to/if one could obtain an inversion formula from which to construct consistent estimates in that basis, and instead, [35] proposed Bayesian estimators. While their method can work well in some cases, it produces estimators that are generally not consistent. In the next section we construct an orthogonal series estimator of a compactly supported density f_X and suggest a basis from which we can derive consistent estimators of the $f_{X,k}$'s.

3. Methodology for estimating f_X

In this section, assuming that f_X is supported on $\mathcal{S}_X \subseteq [A_X, B_X]$ with A_X and B_X finite, we introduce an orthogonal series estimator of f_X that can be computed from the $W_{i,j}$'s generated by model (2.1). Note that while it is common in statistics to assume that densities are supported on $\mathbb{R}$, many variables representing real phenomena take their values in a compact set. Following [21], we assume first that $[A_X, B_X]$ is known. Without loss of generality, $[A_X, B_X] = [-1, 1]$, so that $f_X \in L_2([-1, 1])$; see Section 6 for more general cases.

We propose to take the φ_k 's to be orthonormal polynomials. With that choice, the $f_{X,k}$'s are linear combinations of moments of X . Since the moments of W can be expressed as sums of products of moments X , $\sigma(X)$ and U , this suggests that it may be possible to find an inversion formula expressing the $f_{X,k}$'s in terms of quantities estimable from the $W_{i,j}$'s. For $f, g \in L_2([-1, 1])$, let $\langle f, g \rangle = \int_{-1}^1 f g$

and $\|f\| = (\int_{-1}^1 f^2)^{1/2}$. We use the basis of Legendre polynomials scaled by the factor $\sqrt{k+1/2}$, that is

$$\varphi_k(x) = 2^k \sqrt{k+1/2} \sum_{\ell=0}^k \binom{k}{\ell} \binom{(k+\ell-1)/2}{k} \cdot x^\ell, \quad k = 0, 1, 2, \dots,$$

with $\binom{(k+\ell-1)/2}{k}$ the generalised form of the binomial coefficient. For $k = 0, 1, 2, \dots$, the φ_k 's form an orthonormal basis of $L_2([-1, 1])$, so that we can write

$$f_X = \frac{1}{2} + \sum_{k=1}^{\infty} f_{X,k} \cdot \varphi_k, \quad (3.1)$$

where we used the fact that $\varphi_0 = \sqrt{1/2}$ and $\langle f_X, \varphi_0 \rangle = \sqrt{1/2}$, and where, for $k \geq 1$,

$$f_{X,k} = \langle f_X, \varphi_k \rangle = 2^k \sqrt{k+1/2} \sum_{\ell=0}^k \binom{k}{\ell} \binom{(k+\ell-1)/2}{k} \cdot \mathbb{E}(X^\ell). \quad (3.2)$$

Therefore, and since $\mathbb{E}(X^0) = 1$, to estimate $f_{X,k}$ it suffices to estimate $\mathbb{E}(X^\ell)$ for $\ell = 1, \dots, k$.

This problem might seem trivial at first sight: since $W_{i,j} = X_i + \varepsilon_{i,j}$, it may seem that we can deduce the $\mathbb{E}(X^\ell)$'s from the $\mathbb{E}(W^\ell)$'s, which can be estimated from the $W_{i,j}$'s. However, while this is true in the classical model at (1.1), it is not possible in our case because X_i and $\varepsilon_{i,j}$ are not independent. Specifically, using (2.1), for any integer $\ell \geq 1$ such that $\mu_\ell \equiv \mathbb{E}(U^\ell)$ is finite, we have

$$\mathbb{E}(W^\ell) = \sum_{\kappa=0, \dots, \ell} \binom{\ell}{\kappa} \mathbb{E}\{X^{\ell-\kappa} \sigma^\kappa(X)\} \mu_\kappa = \sum_{\kappa=0, \dots, \lfloor \ell/2 \rfloor} \binom{\ell}{2\kappa} \mathbb{E}\{X^{\ell-2\kappa} \sigma^{2\kappa}(X)\} \mu_{2\kappa}, \quad (3.3)$$

since $\mu_\kappa = 0$ for κ odd. In (3.3), for each ℓ , $\mathbb{E}(W^\ell)$ can be estimated from the $W_{i,j}$'s, but the right hand side comprises $\lfloor \ell/2 \rfloor + 1$ unknowns: $\mathbb{E}(X^\ell)$ and $\mathbb{E}\{X^{\ell-2\kappa} \sigma^{2\kappa}(X)\}$'s for $\kappa = 1, \dots, \lfloor \ell/2 \rfloor$. Thus there are too many unknowns in (3.3) and we can't identify the $\mathbb{E}(X^\ell)$'s from the $\mathbb{E}(W^\ell)$'s.

Now computing moments of W alone does not exploit the relationship between $W_{i,1}$ and $W_{i,2}$. Yet, considered together, these could contain useful information since they are two replicates of the same X_i . To consider these jointly, we write (2.1) in the matrix form $C_i = \mathcal{A}_i \mathcal{B}_i$, where $C_i = (W_{i,1}, W_{i,2})$ and $\mathcal{A}_i = (X_i, \sigma(X_i))$ are independent row vectors and $\mathcal{B}_i$ is a 2×2 matrix with j th column equal to $(1, U_{i,j})^T$, for $j = 1, 2$. Then to obtain equations in the $\mathbb{E}(X^\ell)$'s we seek to estimate, we use the mixed-product property of the Kronecker product to get

$$\mathbb{E}C_i^{\otimes, \ell} = \mathbb{E}\mathcal{A}_i^{\otimes, \ell} \mathbb{E}\mathcal{B}_i^{\otimes, \ell}, \quad (3.4)$$

where $\mathcal{A}^{\otimes, \ell}$ denotes the ℓ -fold Kronecker power of a matrix $\mathcal{A}$. There are many redundancies and zeros in (3.4), which induces only $\lfloor \ell/2 \rfloor + 1$ unique equations corresponding to the unique elements $\mathbb{E}(W_{i,1}^{\ell-k} W_{i,2}^k)$ ($k = 0, \dots, \lfloor \ell/2 \rfloor$) of $\mathbb{E}C_i^{\otimes, \ell}$, and it suffices to study those equations.

To do this, let W_1 and W_2 denote generic $W_{i,1}$ and $W_{i,2}$. Since X and U are independent, applying the changes of variables $\iota = \kappa + \kappa'$, $r = \iota/2$ and $r' = \kappa/2$, we deduce from (2.1) that for $0 \leq \nu \leq \lfloor \ell/2 \rfloor$,

$$\mathbb{E}(W_1^{\ell-\nu} W_2^\nu) = \sum_{\kappa=0}^{\nu} \sum_{\kappa'=0}^{\ell-\nu} \binom{\nu}{\kappa} \binom{\ell-\nu}{\kappa'} \mu_\kappa \mu_{\kappa'} \mathbb{E}\{X^{\ell-\kappa-\kappa'} \sigma^{\kappa+\kappa'}(X)\}$$

$$\begin{aligned}
&= \sum_{i=0}^{\ell} \sum_{\kappa=\max(0, i-\ell+\nu)}^{\min(i, \nu)} \binom{\nu}{\kappa} \binom{\ell-\nu}{i-\kappa} \mu_{\kappa} \mu_{i-\kappa} \mathbb{E}\{X^{\ell-i} \sigma^i(X)\} \\
&= \sum_{i=0}^{\ell} \sum_{\kappa=0}^i 1\{\kappa \leq \nu\} 1\{i-\kappa \leq \ell-\nu\} \binom{\nu}{\kappa} \binom{\ell-\nu}{i-\kappa} \mu_{\kappa} \mu_{i-\kappa} \mathbb{E}\{X^{\ell-i} \sigma^i(X)\} \\
&= \sum_{r=0}^{\lfloor \ell/2 \rfloor} \sum_{r'=0}^r 1\{2r' \leq \nu\} 1\{2r-2r' \leq \ell-\nu\} \binom{\nu}{2r'} \binom{\ell-\nu}{2r-2r'} \mu_{2r'} \mu_{2r-2r'} \mathbb{E}\{X^{\ell-2r} \sigma^{2r}(X)\}, \quad (3.5)
\end{aligned}$$

since $\mu_j = 0$ for all j odd. Combining (3.5) for $\nu = 0, \dots, \lfloor \ell/2 \rfloor$, we obtain $\lfloor \ell/2 \rfloor + 1$ equations with $\lfloor \ell/2 \rfloor + 1$ unknowns: $\mathbb{E}(X^{\ell})$ we want to estimate, and $\mathbb{E}\{X^{\ell-2r} \sigma^{2r}(X)\}$ for $r = 1, \dots, \lfloor \ell/2 \rfloor$.

Since we can estimate $\mathbb{E}(W_1^{\ell-\nu} W_2^{\nu})$ from the $W_{i,j}$'s, this suggests that we could estimate $\mathbb{E}(X^{\ell})$ by inverting the $\lfloor \ell/2 \rfloor + 1$ equations induced by (3.5). To check whether this is the case, let

$$\begin{aligned}
\mathcal{M}_W(\ell) &= (\mathbb{E}(W_1^{\ell}), \mathbb{E}(W_1^{\ell-1} W_2), \dots, \mathbb{E}(W_1^{\lfloor \ell/2 \rfloor} W_2^{\lfloor \ell/2 \rfloor})), \\
\mathcal{M}_{X,\sigma}(\ell) &= \left(\mathbb{E}(X^{\ell}), \mathbb{E}\{\sigma^2(X) X^{\ell-2}\}, \dots, \mathbb{E}\{\sigma^{2\lfloor \ell/2 \rfloor}(X) X^{\ell-2\lfloor \ell/2 \rfloor}\} \right)^T,
\end{aligned}$$

and let $\mathcal{M}_U(\ell)$ be the known $(\lfloor \ell/2 \rfloor + 1) \times (\lfloor \ell/2 \rfloor + 1)$ matrix with $(\nu + 1, r + 1)$ th element

$$\mathcal{M}_{U,\nu+1,r+1}(\ell) = \sum_{r'=0}^r 1\{2r' \leq \nu\} 1\{2r-2r' \leq \ell-\nu\} \binom{\nu}{2r'} \binom{\ell-\nu}{2r-2r'} \mu_{2r'} \mu_{2r-2r'}, \quad (3.6)$$

for $r, \nu = 0, \dots, \lfloor \ell/2 \rfloor$. Using (3.5), we get

$$\mathcal{M}_W(\ell) = \mathcal{M}_U(\ell) \mathcal{M}_{X,\sigma}(\ell). \quad (3.7)$$

In Lemma 1 in Section 4, we will show that for any integer $k \geq 1$, if μ_k is finite then, for $\ell = 1, \dots, k$,

$$\mathcal{M}_U(\ell) \text{ is invertible} \quad (3.8)$$

so that for $\ell = 1, \dots, k$,

$$\mathcal{M}_{X,\sigma}(\ell) = \{\mathcal{M}_U(\ell)\}^{-1} \mathcal{M}_W(\ell). \quad (3.9)$$

Defining the $(\lfloor \ell/2 \rfloor + 1)$ -dimensional vector $\mathbf{e}_{0,\ell} = (1, 0, \dots, 0)$, we deduce that for $\ell = 1, \dots, k$,

$$\mu_{X,\ell} \equiv \mathbb{E}(X^{\ell}) = \mathbf{e}_{0,\ell} \{\mathcal{M}_U(\ell)\}^{-1} \mathcal{M}_W(\ell). \quad (3.10)$$

For $\ell = 1, \dots, k$, we can estimate $\mathcal{M}_W(\ell)$ by the empirical moments

$$\hat{\mathcal{M}}_W(\ell) = \frac{1}{n} \sum_{i=1}^n \left(W_{i,1}^{\ell}, W_{i,1}^{\ell-1} W_{i,2}, \dots, W_{i,1}^{\lfloor \ell/2 \rfloor} W_{i,2}^{\lfloor \ell/2 \rfloor} \right)^T, \quad (3.11)$$

and since $\mathcal{M}_U(\ell)$ is known, we propose to estimate $\mu_{X,\ell}$ by

$$\hat{\mu}_{X,\ell} = \mathbf{e}_{0,\ell} \{\mathcal{M}_U(\ell)\}^{-1} \hat{\mathcal{M}}_W(\ell). \quad (3.12)$$

Letting $\hat{\mu}_{X,0} = \mu_{X,0} = 1$, for any integer $k \geq 1$ such that μ_k is finite, we can estimate $f_{X,k}$ by

$$\hat{f}_{X,k} = 2^k \sqrt{k+1/2} \sum_{\ell=0}^k \binom{k}{\ell} \binom{(k+\ell-1)/2}{k} \cdot \hat{\mu}_{X,\ell}. \quad (3.13)$$

Finally, if K is a truncating parameter such that $\mu_{\lceil K \rceil}$ is finite, we can estimate f_X by

$$\hat{f}_X = \frac{1}{2} + \sum_{k=1}^{\lceil K \rceil} w_k \cdot \hat{f}_{X,k} \cdot \varphi_k, \quad (3.14)$$

where $w_k = 1\{k < K\} + (K+1-k) \cdot 1\{K \leq k < K+1\}$. If K is an integer, these are the hard-thresholding weights $w_k = 1\{k \leq K\}$. There can be some advantages in allowing K to be non integer, taking the last weight $w_{\lceil K \rceil}$ to be fractional, which truncates the sum less abruptly; see our data-driven implementation in Section 5. For example, $w_{\lceil K \rceil} = (K+1 - \lceil K \rceil)$ is close to 0 if $\lceil K \rceil \approx K$ and close to 1 if $\lceil K \rceil \approx K+1$. In what follows, when we want to highlight the dependence of $\hat{f}_X$ on K we use the notation $\hat{f}_X(\cdot; K)$.

Remark 1. The general framework with Kronecker moments we used to find the equations for estimating the $\mathbb{E}(X^\ell)$'s opens the perspective to treat broader classes of statistical inverse problems when the product of two independent random matrices is observed and the distribution of one of them is known.

4. Theoretical properties

In this section we study theoretical properties of $\hat{f}_X$ at (3.14). We first prove (3.8), and derive a lower bound on the singular values of $\mathcal{M}_U$ in the case where $U \sim N(0, 1)$, which will be useful to derive convergence rates in that case.

Lemma 1. (a). When f_U is symmetric around 0 and U has ℓ finite moments, $\mathcal{M}_U(\nu)$ is invertible for $\nu = 1, \dots, \ell$. (b). Moreover, if $U \sim N(0, 1)$, for all $\lceil \ell/2 \rceil + 1$ columns vectors x , we have $|\mathcal{M}_U(\ell)x|^2 \geq \ell^{-1} 4^{-\ell} |x|^2 / C^*$, where $C^* = 2 \sum_{r=0}^{\infty} 16^r \cdot r!^2 / (2r)!^2$.

Next, we study the behaviour of the mean integrated squared error

$$\text{MISE}_n(f_X) = \mathbb{E} \int_{-1}^1 |\hat{f}_X(x) - f_X(x)|^2 dx \quad (4.1)$$

in the case where f_X belong to $\mathcal{F}(c, m)$, the class of functions f whose support is included in $[-1, 1]$, which are m times continuously differentiable on $(-1, 1)$ and satisfy the Sobolev condition $\|f^{(m)}\|^2 \leq c$. In the next theorem we provide a general upper bound of MISE_n which is valid for any sample size n , and then give a more specific upper bound in the case where $U \sim N(0, 1)$.

Theorem 1. For any given integer $m \geq 1$, let $\hat{f}_X$ be the estimator at (3.14) computed with $\lceil K \rceil \geq m$. (a). If f_U is even and $\mu_{\lceil K \rceil} < \infty$, then letting $\mathcal{W}_{1,\ell} = (W_{1,1}^\ell, W_{1,1}^{\ell-1} W_{1,2}, \dots, W_{1,1}^{\lceil \ell/2 \rceil} W_{1,2}^{\lfloor \ell/2 \rfloor})^T$ and $\|\sigma\|_\infty = \sup_{x \in [-1, 1]} \sigma(x)$, for any finite positive constants c and d we have

$$\sup_{f_X \in \mathcal{F}(c, m), \|\sigma\|_\infty \leq d} \text{MISE}_n(f_X) \leq c \cdot \{1 + m/(\lceil K \rceil + 3/2)\} \cdot \prod_{i=1}^m (\lceil K \rceil + 3/2 - i)^{-2}$$

$$+ \sup_{\|\sigma\|_\infty \leq d} \frac{1}{n} \sum_{k=1}^{\lceil K \rceil} w_k^2 4^k (k+1/2) \left[\sum_{\ell=1}^k \binom{k}{\ell} \cdot \{\mathbb{E}|\mathbf{e}_{0,\ell} \{\mathcal{M}_U(\ell)\}^{-1} \mathcal{W}_{1,\ell}|^2\}^{1/2} \right]^2.$$

(b). If $U \sim N(0, 1)$, for any finite positive constants c and d we have, for C^* is as in Lemma 1,

$$\begin{aligned} \sup_{f_X \in \mathcal{F}(c, m), \|\sigma\|_\infty \leq d} \text{MISE}_n(f_X) &\leq c \cdot \{1 + m/(\lfloor K \rfloor + 3/2)\} \cdot \prod_{i=1}^m (\lfloor K \rfloor + 3/2 - i)^{-2} \\ &\quad + \frac{1}{n} \cdot C^* \sum_{k=1}^{\lceil K \rceil} w_k^2 \cdot k^2 (k+1/2) \cdot (6+4d)^{2k} \cdot (2k-1)!! . \end{aligned}$$

In Theorem 1 (a), the first and second terms on the right hand side of the inequality are upper bounds to, respectively, the squared bias and the variance of $\hat{f}_X$. The upper bound to the variance depends on the matrices $\mathcal{M}_U(1), \dots, \mathcal{M}_U(\lceil K \rceil)$, which change with f_U and so are difficult to study. Using Lemma 1 (b), in Theorem 1 (b) we provide a more explicit upper bound in the particular case where $U \sim N(0, 1)$, which is often considered one of the most important ones from a practical point a view.

It is well known in the deconvolution literature that, in the Gaussian error case, minimax convergence rates of nonparametric curve estimators are logarithmic. For example, in the classical model at (1.1), that rate is $(\log n)^{-m}$, see [17]. In the next theorem, we show that the convergence rate of $\hat{f}_X$ is also logarithmic, but it is faster than $(\log n)^{-m}$ because in our case, f_X is compactly supported; see [20], [28] and [19] for similar phenomena.

Theorem 2. For any given integer $m \geq 1$, if $U \sim N(0, 1)$ and $\hat{f}_X$ at (3.14) is computed with $K = K_n = \lfloor c_0(\log n)/(\log \log n) \rfloor$ for some constant $c_0 \in (0, 1)$, then for any finite positive constants c and d , as $n \rightarrow \infty$ we have

$$\sup_{f_X \in \mathcal{F}(c, m), \|\sigma\|_\infty \leq d} \text{MISE}_n(f_X) = O\{(\log n / \log \log n)^{-2m}\}.$$

In Appendix A.3, using results from [28] and [16], we prove that the convergence rate in Theorem 2 is minimax optimal. As in the homoscedastic case (e.g. [31]), our estimator can converge at polynomial rates if f_X is supersmooth. Specifically, consider the case where f_X belong to the class of infinite-fold differentiable functions $\mathcal{G}(c, \xi) = \bigcap_{m \geq 1} \mathcal{F}(cm!^{2\xi}, m)$, where $c > 0$ and $\xi \in (0, 1)$. Then taking $K = K_n$ such that $K_n! \asymp n^{1/(3-2\xi)}$ and noting that $\mathcal{G}(c, \xi) \subseteq \mathcal{F}(cm!^{2\xi}, m)$ for $m = c_0 \cdot K_n$ with $c_0 > 0$ sufficiently small, we deduce from Theorem 1(b) that the convergence rate of the supremum of MISE_n over $\mathcal{G}(c, \xi)$ has the upper bound $n^{-\chi}$ for all $\chi < 2(1 - \xi)/\{2(1 - \xi) + 1\}$.

Remark 2. Consistency of the estimator $\hat{f}_X$ requires that $K \rightarrow \infty$ as $n \rightarrow \infty$, which requires that all moments of U are finite. However, in practice, n and K are always finite, and to compute $\hat{f}_X$ we only need U to have $\lceil K \rceil$ finite moments. Theorem 1, which is valid for all n , provides an upper bound of the MISE of $\hat{f}_X$ in that finite sample setting.

Remark 3. In applications, most variables naturally take their values on a non-compact set, but it can be of theoretical interest to discuss the case where the support of f_X is unbounded. We start by proving that f_X is identifiable in this case as long as it fulfils some conditions which are satisfied by, e.g., normal densities. Specifically, assume the conditions of Lemma 1(a) so that all moments of X in model

(2.1) are identified. If, in addition, the moments of X do not grow too fast, in the sense that

$$\limsup_{k \rightarrow \infty} c^k \cdot \mathbb{E}|X|^k / k! = 0 \text{ for all } c > 0,$$

then the characteristic function ϕ_X of X can be expanded into a power series that converges point-wise on the entire real line, with coefficients uniquely determined by the moments of X . Since ϕ_X uniquely determines f_X , we deduce that under those conditions, f_X is identifiable even if its support is unbounded. While our estimator is designed for compactly supported f_X , it is technically possible to extend it to this unbounded scenario. As discussed by [6] in a homoscedastic measurement error problem, even if f_X has unbounded support, in finite samples, $X_1, \dots, X_n$ cover only a finite range that increases with n . Similar to their case, it can be shown that our estimator is consistent if, for each n , we estimate f_X by the estimator of its restriction over a finite interval that grows sufficiently slowly with n . See Section 4.1 in [6] for more details, and [28] for similar results. As in [6] and [28], in this setting our estimator will converge more slowly than the rates from Theorem 2.

5. Data-driven choice of K

The choice of K is crucial for good performance of $\hat{f}_X$ in (3.14). This estimator achieves optimal L_2 rates if computed with $K = K_0(f_X) \equiv \operatorname{argmin}_K \operatorname{MISE}_n(f_X)$, with MISE_n at (4.1) applied to f_X , but $K_0(f_X)$ is not computable in practice and we need a data-driven rule for choosing K . In this section, assuming that the data have been rescaled so that $S_X = [-1, 1]$, we suggest an approach inspired by [15]’s bandwidth selector. In Section 6 and Appendix B.3, we extend it to the case where S_X is unknown. Similar to [13]’s SIMEX procedure, our approach involves simulating noisy versions of the $W_{i,j}$ ’s, tracking the impact of the noise on K , and extrapolating back to find a suitable K for $\hat{f}_X$. As usual, we perform the search for K on a finite set of integer values, $\mathcal{I}_K$ say.

Our method requires estimating the density of the averages $\bar{W}_i = (W_{i,1} + W_{i,2})/2$, which is equal to

$$f_{\bar{W}}(w) = \int_{-1}^1 f_X(x) \frac{1}{\sigma(x)} f_U\left\{\frac{w-x}{\sigma(x)}\right\} dx, \quad (5.1)$$

where $f_{\bar{U}} = f_U/2 * f_U/2$, since $\bar{W}_i = X_i + \sigma(X_i)\bar{U}_i$, with $\bar{U}_i = (U_{i,1} + U_{i,2})/2$.

Since we cannot compute $K_0(f_X)$, we start by choosing a K , $\hat{K}(f_{\bar{W}})$ say, suitable for estimating $f_{\bar{W}}$ by the estimator obtained by plugging $\max\{\hat{f}_X(\cdot; K), 0\}$ into (5.1), that is

$$\hat{f}_{\bar{W}, \text{Leg}}(w; K) = \int_{-1}^1 \max\{\hat{f}_X(x; K), 0\} \frac{1}{\hat{\sigma}_{\text{pilot}}(x)} f_{\bar{U}}\left\{\frac{w-x}{\hat{\sigma}_{\text{pilot}}(x)}\right\} dx, \quad (5.2)$$

where we replaced the unknown σ by a pilot parametric estimator $\hat{\sigma}_{\text{pilot}}$ described in Section C.1.

Unlike for f_X , we have access to error-free data from $f_{\bar{W}}$ (the $\bar{W}_i$ ’s). Therefore, we can also compute a standard KDE $\hat{f}_{\bar{W}, \text{KDE}}$ of $f_{\bar{W}}$ from the $\bar{W}_i$ ’s [37]. Since $\hat{f}_{\bar{W}, \text{KDE}}$ is constructed from error-free data, it is usually closer than $\hat{f}_{\bar{W}, \text{Leg}}$ to $f_{\bar{W}}$, which motivates us to choose

$$\hat{K}(f_{\bar{W}}) = \operatorname{argmin}_{K \in \mathcal{I}_K} \int_{L_W}^{U_W} \{\hat{f}_{\bar{W}, \text{Leg}}(w; K) - \hat{f}_{\bar{W}, \text{KDE}}(w)\}^2 dw, \quad (5.3)$$

where $[L_W, U_W]$ is a large subset of the support of $f_{\bar{W}}$ that excludes the tails where both estimators are too unreliable. Namely,

$$L_W = \min\{-1.2, \min \bar{W}_i + 0.15(\min \bar{W}_i - \max \bar{W}_i)\}, \quad (5.4)$$

$$U_W = \max\{1.2, \max \bar{W}_i + 0.15(\max \bar{W}_i - \min \bar{W}_i)\}, \quad (5.5)$$

where the values ± 1.2 ensure that $[L_W, U_W]$ is significantly larger than $\mathcal{S}_X \subseteq [-1, 1]$.

Now $f_{\bar{W}}$ is a smoothed version of f_X , and so $\hat{K}(f_{\bar{W}})$ is generally too large for $\hat{f}_X$. Inspired by [15], we construct an approximation $\hat{C}_K$ of the factor $C_K = K_0(f_X)/\hat{K}(f_{\bar{W}})$, and approximate $K_0(f_X)$ by

$$\hat{K} = \hat{C}_K \cdot \hat{K}(f_{\bar{W}}). \quad (5.6)$$

Note that $\hat{K}$ is not necessarily an integer, and instead of truncating it to its integer part, we found we got better results by using the fractional $w_{\lceil \hat{K} \rceil}$ defined below (3.14).

To obtain $\hat{C}_K$, similar to [15] we mimic $K_0(f_X)/\hat{K}(f_{\bar{W}})$ by $K_0(f_{X^*})/\hat{K}(f_{\bar{W}^*})$, where f_{X^*} (resp., $f_{\bar{W}^*}$) is the density of a contaminated version X^* of X (resp., $\bar{W}^*$ of $\bar{W}$). Paraphrasing [13], X^* and $\bar{W}^*$ are generated from the data such that $\bar{W}^*$ measures $\bar{W}$ in the same way as X^* measures X . Therefore we can expect that $K_0(f_X)/\hat{K}(f_{\bar{W}}) \approx K_0(f_{X^*})/\hat{K}(f_{\bar{W}^*})$. Unlike f_X , we can choose f_{X^*} so as to have access to both error-free data $X_i^* \sim f_{X^*}$ and contaminated $W_{i,j}^*$'s. We can compute two estimators of f_{X^*} : the standard KDE from the X_i^* 's and our Legendre series estimator from the $W_{i,j}^*$'s, and we can estimate $K_0(f_{X^*})$ by minimising a distance between those two estimators, which is not possible for $K_0(f_X)$.

It is easy to choose contaminated version X_i^* 's of the X_i 's to which we have access from the data and which mimic the X_i 's: we can take $X_i^* = \bar{W}_i$. However, since we will also need to compute the Legendre estimator of f_{X^*} from contaminated versions of the X_i^* 's, we must choose f_{X^*} so that $\mathcal{S}_{X^*} \subseteq [-1, 1]$, whereas $\mathcal{S}_{\bar{W}} \supsetneq [-1, 1]$, and can even be unbounded if f_U is supported on $\mathbb{R}$. To ensure that our choice satisfies $\mathcal{S}_{X^*} \subseteq [-1, 1]$, we rescale the $\bar{W}_i$'s, and then keep only the rescaled values that are in $[-1, 1]$, creating in that way a sample $X_1^*, \dots, X_{n_1}^*$ of size $n_1 \leq n$ of X^* values that satisfy our requirements; see Appendix C.2. Then to create contaminated versions $W_{i,j}^*$ of the X_i^* 's by mimicking the contamination process of the X_i 's, for $i = 1, \dots, n_1$ and $j = 1, 2$, we generate i.i.d. $U_{i,j}^* \sim f_U$ and take $W_{i,j}^* = X_i^* + \sigma^*(X_i^*)U_{i,j}^*$, where $\sigma^* = \hat{\sigma}_{\text{pilot}}$.

Constructed like this, the $(X_i^*, W_{i,1}^*, W_{i,2}^*)$'s mimic the $(X_i, W_{i,1}, W_{i,2})$'s but for the $(X_i^*, W_{i,1}^*, W_{i,2}^*)$'s, both the error-free X_i^* and their contaminated versions $W_{i,j}^*$ are observed, so that we can compute both the standard KDE $\hat{f}_{X^*, \text{KDE}}$ using the X_i^* 's, and the version $\hat{f}_{X^*}$ of $\hat{f}_X$ at (3.14) obtained by replacing the $W_{i,j}$'s by the $W_{i,j}^*$'s. Since $\hat{f}_{X^*, \text{KDE}}$ is computed from error-free data, it converges to f_{X^*} faster than $\hat{f}_{X^*}$. Similar to [15] and [13], we generate B such samples by taking, for $b = 1, \dots, B$, $W_{i,j,b}^* = X_i^* + \sigma^*(X_i^*)U_{i,j,b}^*$, where the $U_{i,j,b}^*$'s are i.i.d. $\sim f_U$. For each b , we compute a version $\hat{f}_{X^*,b}$ of $\hat{f}_{X^*}$ from the $W_{i,j,b}^*$'s, and compute a data-driven version $\hat{K}_0(f_{X^*})$ of $K_0(f_{X^*})$ by taking

$$\hat{K}_0(f_{X^*}) = \operatorname{argmin}_{K \in \mathcal{I}_K} \frac{1}{B} \sum_{b=1}^B \int_{-1}^1 [\max\{\hat{f}_{X^*,b}(x; K), 0\} - \hat{f}_{X^*, \text{KDE}}(x)]^2 dx. \quad (5.7)$$

Next, letting $f_{\overline{W}^*}$ denote the density of $\overline{W}_{i,b}^* = \sum_{j=1}^2 W_{i,j,b}^*/2$, we mimic (5.3) by taking

$$\hat{K}(f_{\overline{W}^*}) = \operatorname{argmin}_{K \in \mathcal{I}_K} \frac{1}{B} \sum_{b=1}^B \int_{L_{W,b}^*}^{U_{W,b}^*} \{\hat{f}_{\overline{W}^*, \text{Leg}, b}(x; K) - \hat{f}_{\overline{W}^*, \text{KDE}, b}(x)\}^2 dx, \quad (5.8)$$

with $L_{W,b}^*$ and $U_{W,b}^*$ like L_W and U_W in (5.4) and (5.5) but with the $\overline{W}_i$'s replaced by the $\overline{W}_{i,b}^*$'s, and $\hat{f}_{\overline{W}^*, \text{Leg}, b}$ and $\hat{f}_{\overline{W}^*, \text{KDE}, b}$ defined as $\hat{f}_{\overline{W}, \text{Leg}}$ and $\hat{f}_{\overline{W}, \text{KDE}}$, but with the $W_{i,j}$'s and $\overline{W}_i$'s replaced by the $W_{i,j,b}^*$'s and $\overline{W}_{i,b}^*$'s. We take $\hat{C}_K = \hat{K}_0(f_{X^*})/\hat{K}(f_{\overline{W}^*})$ in (5.6).

6. Case where $\mathcal{S}_X \not\subseteq [-1, 1]$

6.1. Rescaled data and estimator

Using (3.2), our procedure estimates the coefficients $f_{X,k}$ by estimating the $\mathbb{E}(X^\ell)$'s. However, if $\mathcal{S}_X \not\subseteq [-1, 1]$, then (3.2) does not hold because, for $\ell \geq 0$, $\int_{-1}^1 x^\ell f_X(x) dx \neq \mathbb{E}(X^\ell)$. To apply our estimator in the case where $\mathcal{S}_X \subseteq [A_X, B_X] \not\subseteq [-1, 1]$, where $[A_X, B_X]$ is either $\mathcal{S}_X$ if we know it, or else a bound we have on $\mathcal{S}_X$, we need to rescale the $W_{i,j}$'s by computing

$$W'_{i,j} = (W_{i,j} - \mu)/\lambda = X'_i + \sigma'(X'_i)U_{i,j}, i = 1, \dots, n, j = 1, 2, \quad (6.1)$$

$$\text{where } X'_i = (X_i - \mu)/\lambda, \quad \sigma'(X'_i) = \sigma(X'_i\lambda + \mu)/\lambda, \quad \lambda = (B_X - A_X)/2 \quad \text{and} \quad \mu = (B_X + A_X)/2. \quad (6.2)$$

With these transformed data, we have $\mathcal{S}_{X'} \subseteq [-1, 1]$ and the $f_{X',k}$'s in the expansion $f_{X'} = 1/2 + \sum_{k=1}^\infty f_{X',k} \cdot \varphi_k$ satisfy (3.2) with X replaced by X' . Therefore, we can estimate $f_{X'}$ by

$$\hat{f}_{X'}(\cdot, K) = \frac{1}{2} + \sum_{k=1}^{[K]} w_k \cdot \hat{f}_{X',k} \varphi_k, \quad (6.3)$$

with $\hat{f}_{X',k}$ computed as $\hat{f}_{X,k}$ in (3.13), but replacing there X by X' and with the estimators computed from the $W'_{i,j}$'s instead of the $W_{i,j}$'s. Finally, we estimate $f_X(x)$ by $\tilde{f}_X(x; K) = \lambda^{-1} \hat{f}_{X'}\{(x - \mu)/\lambda; K\}$.

6.2. Importance of the bound on $\mathcal{S}_X$

In the orthogonal series literature with compactly supported functions, it is standard to assume that we have an upper bound $[A_X, B_X]$ to the support of the curve to estimate, even when the data are observed with errors [see e.g. 21]. Then the data are rescaled for the support of the rescaled curve to be a subset of the interval on which the series is constructed, like we did in Section 6.1. In real applications, using physical limitations of the process under study, we can often find such an upper bound.

In theory, any crude bound $[A_X, B_X]$ usually suffices to obtain a consistent estimator; for example, in our case, all we need of $\mathcal{S}_{X'}$ for $\hat{f}_{X'}$ at (6.3) to be consistent is that $\mathcal{S}_{X'} \subseteq [-1, 1]$. However, a crude bound often entails poorer practical performance of estimators. For example, in our case, if $[A_X, B_X]$ is crude, $\mathcal{S}_{X'} \setminus [-1, 1]$ is large and $\hat{f}_{X'}$ spends significant effort estimating $f_{X'}(x) = 0$ on that region. As a result, to capture the fine details of $f_{X'}(x) > 0$ on $\mathcal{S}_{X'}$ itself, the expansion at (6.3) typically needs to be computed with K larger than for a tighter bound. Now, in finite samples, the number $[K]$ of

coefficients $f_{X,k}$ we can estimate well is restricted by the sample size n . Since, for a same K , $\hat{f}_X$ can capture more details of f_X if we have a tighter bound $[A_X, B_X]$ on S_X , then the tighter the bound, the better the performance of our estimator. To get a good bound to S_X , we could estimate the left and right endpoints of S_X , say by $\hat{L}_X$ and $\hat{R}_X$, and take $[A_X, B_X]$ to be a slightly enlarged version of $[\hat{L}_X, \hat{R}_X]$, to ensure that $S_X \subseteq [A_X, B_X]$, which is needed for the rescaled estimator at (6.3) to be consistent.

Developing a formal method for estimating the endpoints of S_X and proving its consistency would require more work than is reasonable within this paper. Even in simpler homoscedastic error settings, entire papers have been dedicated to this problem [e.g. 11, 12, 27], and their method cannot be directly applied to our heteroscedastic context. Therefore, in Section 6.3 we propose an informal approach (without theoretical justification) for getting an upper bound to S_X , with which (6.3) already worked well in our numerical work. We leave the development of a more sophisticated procedure for future work.

6.3. Upper bound to S_X and final estimator of f_X

To construct an upper bound $[A_X, B_X]$ to S_X , we proceed in two steps:

- (1) In Appendix B.1, we calculate a crude bound $[A_{X,1}, B_{X,1}]$ easily computable from the data.
- (2) In Appendix B.2, using the $W'_{i,j}$'s in (6.1), temporarily taking there $A_X = A_{X,1}$ and $B_X = B_{X,1}$, we obtain a pilot *undersmoothed* estimator $\hat{f}_{X',p}$ of $f_{X'}$ by computing (6.3) with a too large K . If the bound $[A_{X,1}, B_{X,1}]$ is too crude, $f_{X'}$ vanishes on a large boundary region of $[-1, 1]$, where $\hat{f}_{X',p}$ usually oscillates a lot. We construct a more precise bound $[A_{X,2}, B_{X,2}]$ to S_X by detecting these boundary oscillations, and update our bound $[A_{X,1}, B_{X,1}]$ to $[A_{X,2}, B_{X,2}]$.

Then, we compute the $W'_{i,j}$'s in (6.1) with $[A_X, B_X] = [A_{X,2}, B_{X,2}]$, and the estimator $\hat{f}_{X'}(\cdot, \hat{K})$ at (6.3) of $f_{X'}$ with $K = \hat{K}$ chosen from the data using a generalised version of the method in Section 5; see Appendix B.3. For $x \in [A_X, B_X]$, we estimate $f_X(x)$ by $\tilde{f}_X(x) = \tilde{f}_X(x; \hat{K})$ obtained from $\hat{f}_{X'}(\cdot, \hat{K})$ as explained below (6.3). We make a final refinement by checking if f_X has large spurious oscillations near A_X and B_X . Specifically, our final estimator of f_X is $\tilde{f}_X$ truncated to the interval $[A_X^\dagger, B_X^\dagger]$ obtained by removing from $[A_X, B_X]$ the boundary regions where $\tilde{f}_X$ oscillates widely; see Section B.4.

7. Numerical properties

7.1. Simulations

We applied our method to data $W_{i,j}$ simulated from model (2.1) with four densities f_X supported on $S_X = [-1, 1]$ and defined by

$$f_1(x) = 15f(15x + 15)/F(30) \cdot 1\{x \in [-1, 1]\}, \text{ with } f \text{ a } \chi^2(30) \text{ density and } F(x) = \int_0^x f;$$

$$f_2(x) = 1.25f(1.25x + 1.25)/F(2.5) \cdot 1\{x \in (-1, 1]\}, \text{ with } F(x) = \int_0^x f \text{ and } f \text{ a density from}$$

an odd Burr BX model (see [5]) with parameters $(\nu, \theta, \gamma, m) = (0.4, 0.95, 2.5, 2)$, defined on $\mathbb{R}^+$

by $f_{\nu, \theta, \gamma, m} = \nu \theta \pi_{\gamma, m} \Pi_{\gamma, m}^{\nu-1} \bar{\Pi}_{\gamma, m}^{\nu \theta - 1} \left(\bar{\Pi}_{\gamma, m}^\nu + \Pi_{\gamma, m}^\nu \right)^{-1-\theta}$ with $\Pi_{\gamma, m} = (1 - g_m)^\gamma$, $\bar{\Pi}_{\gamma, m} = 1 - \Pi_{\gamma, m}$,

$\pi_{\gamma, m}(w) = 2\gamma m^2 w g_m(w) \{1 - g_m(w)\}^{\gamma-1}$ and $g_m(w) = \exp\{-(mw)^2\}$;

$$f_3(x) = 2.5 \cdot (x+1)(x^2+0.1)(x-1)^2 \cdot 1\{x \in [-1, 1]\};$$

$f_4(x) = f(x)/\int_{-1}^1 f(u) du \cdot 1\{x \in [-1, 1]\}$, with $f = 0.6\phi_{-0.3, 0.2} + 0.4\phi_{0.3, 0.2}$ and $\phi_{\mu, \nu}$ the density of a $N(\mu, \nu^2)$.

For each f_k , $k = 1, \dots, 4$, we took $\sigma(x) = c_k \exp(x/2)$, with c_k so that the noise-to-signal ratio $\text{NSR} = \mathbb{V}\text{ar}\{\sigma(X)U\}/\mathbb{V}\text{ar}(X) = \mathbb{E}\{\sigma^2(X)\}/\mathbb{V}\text{ar}(X)$ was equal of 10% or 20%. We generated $W_{i,j}$,

Table 1. Simulation results for $\tilde{f}_X$ with unknown S_X and $\hat{f}_{X,\text{Sarkar}}$: median (interquartile range) of $\text{ISE} \times 10^3$ for densities f_1 to f_4 , computed from 200 samples.

f_X	NSR	$\tilde{f}_X$				$\hat{f}_{X,\text{Sarkar}}$			
		$n = 250$	$n = 500$	$n = 1000$	$n = 5000$	$n = 250$	$n = 500$	$n = 1000$	$n = 5000$
f_1	10%	13.50(15.65)	7.65(9.25)	4.35(6.35)	1.20(1.35)	20.90(19.35)	14.40(11.05)	9.15(7.20)	3.45(2.25)
	20%	19.70(26.17)	10.40(13.62)	5.55(8.23)	1.60(2.22)	24.60(25.55)	17.60(16.98)	11.85(7.02)	5.25(5.22)
f_2	10%	24.55(17.35)	15.20(10.70)	11.55(9.68)	4.85(3.50)	30.20(19.82)	24.75(14.10)	18.30(11.10)	8.80(4.43)
	20%	30.80(19.77)	23.50(11.40)	15.70(11.57)	7.70(5.98)	38.30(29.85)	30.55(22.58)	26.70(14.90)	15.85(8.65)
f_3	10%	14.65(11.43)	9.45(6.65)	5.70(3.73)	3.30(2.60)	35.95(25.10)	26.10(16.73)	17.30(9.32)	9.90(4.80)
	20%	26.45(19.23)	18.90(16.20)	11.65(9.38)	5.60(3.32)	45.60(31.60)	37.95(23.18)	27.55(18.43)	17.50(9.90)
f_4	10%	26.85(21.27)	16.50(13.15)	8.25(6.62)	3.55(2.90)	16.60(16.85)	7.00(7.47)	3.40(3.12)	0.65(0.60)
	20%	38.75(19.83)	24.85(18.82)	16.60(11.35)	6.35(5.50)	22.30(22.05)	11.35(11.05)	4.50(5.35)	1.00(0.80)

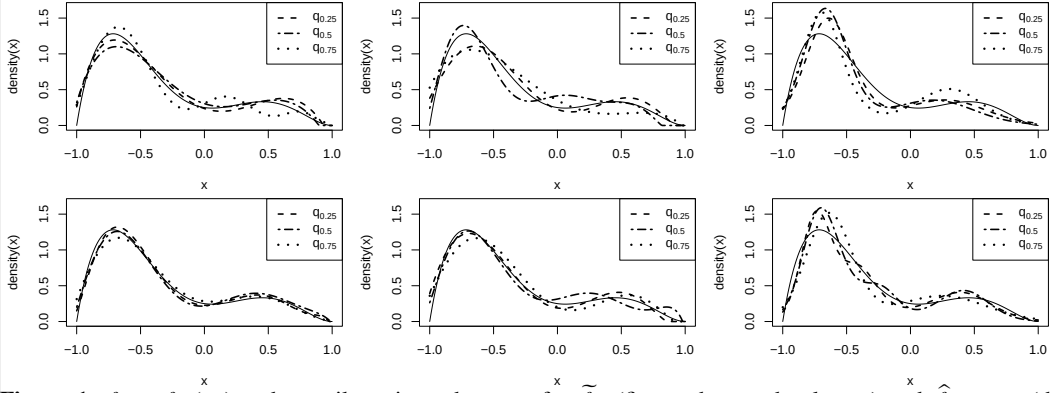


Figure 1. $f_X = f_3$ (—) and quartile estimated curves for $\tilde{f}_X$ (first and second columns) and $\hat{f}_{X,\text{Sarkar}}$ (third column) computed from 200 samples of size $n = 250$ (first row) or 1000 (second row) when NSR = 10% (first column) or 20% (second and third columns).

$i = 1, \dots, n$ and $j = 1, 2$ from model (2.1) for all combinations of the above densities f_X , NSR, and sample sizes $n = 250, 500, 1000$ and 5000 , and with $U_{i,j} \sim N(0, 1)$ (see Appendix D.1 for simulation results where we misspecified f_U by assuming that the $U_{i,j}$'s are averages of two uniforms).

We estimated f_X with our fully data-driven procedure, under two scenarios. In the first, we treated S_X as known, took $[A_X, B_X] = [-1, 1]$ and obtained the estimator $\hat{f}_X(\cdot, \hat{K})$ with $\hat{f}_X$ at (3.14) and $\hat{K}$ at (5.6). In the second, we treated S_X as unknown, computed an upper bound $[A_X, B_X]$ to S_X using the approach from Section 6.3 and used the estimator $\tilde{f}_X = \tilde{f}_X(\cdot; \hat{K})$ defined below (6.3), with $\hat{K}$ at (B.2). In both cases, since a density is always non negative, we truncated negative values to zero. That is, we used $\max(\hat{f}_X, 0)$ and $\max(\tilde{f}_X, 0)$, but to simplify notation, below we denote these truncated estimators by $\hat{f}_X$ and $\tilde{f}_X$. To compute $\hat{K}$, we used the search grid $\mathcal{I}_K = \{3, 4, \dots, 13\}$ and computed all standard KDEs with a plug-in (PI) bandwidth and a standard normal kernel (see [37]). To illustrate the usefulness of our consistent nonparametric estimator over approximation methods, we also computed [33]'s semiparametric Bayesian estimator $\hat{f}_{X,\text{Sarkar}}$. The latter works really well when f_X can be well approximated by a small number of normal mixtures, like density f_4 . On the other hand, we will see that it can also produce undersmoothed estimators in other scenarios.

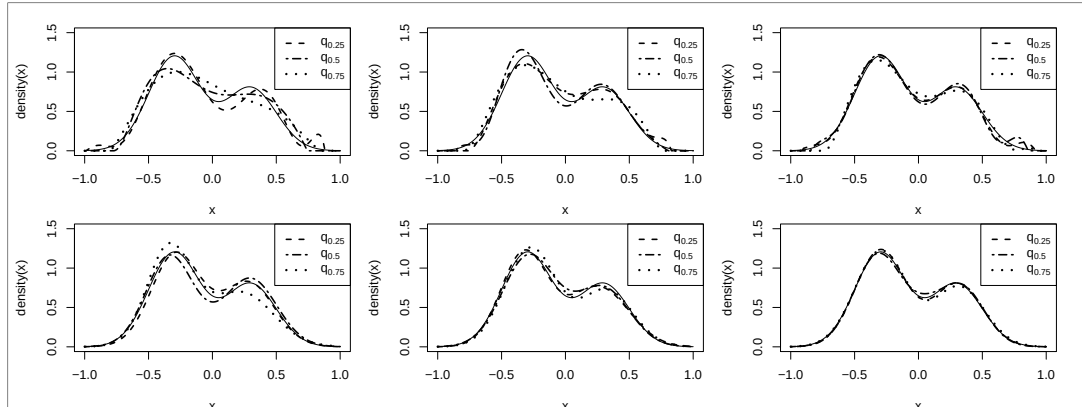


Figure 2. $f_X = f_4$ (—) and quartile estimated curves for $\hat{f}_X$ (first row) or $\hat{f}_{X,\text{Sarkar}}$ (second row) computed from 200 samples of size $n = 500$ (first column), 1000 (second column) or 5000 (third column) when NSR = 10%.

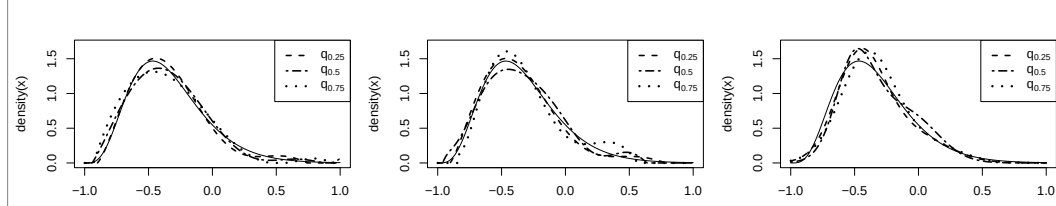


Figure 3. $f_X = f_1$ (—) and quartile estimated curves for $\hat{f}_X$ (first column), $\tilde{f}_X$ (second column) and $\hat{f}_{X,\text{Sarkar}}$ (third column) computed from 200 samples of size $n = 500$ when NSR = 20%.

Let $\check{f}_X$ denote any of those estimators. To assess their performance, for each configuration above we generated 200 samples of $W_{i,j}$'s, applied $\check{f}_X$ to each sample and computed the integrated squared error $\text{ISE} = \int_{-1}^1 (\check{f}_X - f_X)^2$, obtaining 200 ISEs (one per sample) per configuration and estimator.

Table 1 shows the median and interquartile range of the 200 $\text{ISE} \times 10^3$ for our estimator $\tilde{f}_X$ with unknown $\mathcal{S}_X$ and $\hat{f}_{X,\text{Sarkar}}$; see Table 2 in Appendix D.1 for a comparison of $\tilde{f}_X$ with the version $\hat{f}_X$ of our estimator for the case where $\mathcal{S}_X$ is known. As could have been expected, since $\hat{f}_{X,\text{Sarkar}}$ uses a normal mixture to model f_X , it significantly outperformed both versions of our estimator when $f_X = f_4$, where $\hat{f}_X$ and $\tilde{f}_X$ tended to underestimate the height of the peaks. However, our estimators worked better in the other three cases, where $\hat{f}_{X,\text{Sarkar}}$ tended to overestimate the height of peaks of f_X . All three estimators improved as sample size increased and as NSR decreased. In many cases, $\hat{f}_X$ (which requires knowing $\mathcal{S}_X$) gave better results than $\tilde{f}_X$ (which does not know $\mathcal{S}_X$). However, in some cases, the tail correction from Section 6.3 (which we only applied to $\tilde{f}_X$) was so effective that $\tilde{f}_X$ even outperformed $\hat{f}_X$, for which we did not remove spurious boundary oscillations.

To visualize these results graphically, next we show, in some settings, f_X along with three estimated curves corresponding in each setting and for each estimator, to the samples that gave the first, second and third quartiles out of 200 ISEs computed as above. We refer to these curves as quartile estimated curves, and use the notation $q_{0.25}$, (resp., $q_{0.5}$ and $q_{0.75}$) for the curve corresponding to the first (resp., second and third) quartile.

In Figure 1, we compare $\tilde{f}_X$ and $\hat{f}_{X,\text{Sarkar}}$ in the case where $f_X = f_3$, $n = 250$ or 1000, and NSR = 10% or 20%. In this case, $\tilde{f}_X$ outperformed $\hat{f}_{X,\text{Sarkar}}$, which was able to identify the two modes of f_X but overestimated their height, even for large n . This example is particularly well suited to a Legendre poly-

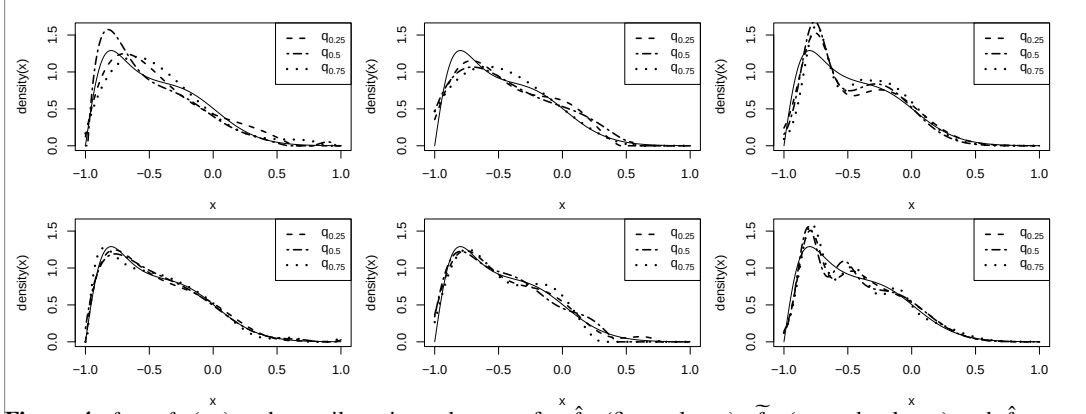


Figure 4. $f_X = f_2$ (—) and quartile estimated curves for $\hat{f}_X$ (first column), $\tilde{f}_X$ (second column) and $\hat{f}_{X,\text{Sarkar}}$ (third column) computed from 200 samples of size $n = 500$ (first row) and 5000 (second row) when NSR = 20%.

nomial approximation since f_3 is a polynomial, and so in Figure 2, we consider an example favourable to $\hat{f}_{X,\text{Sarkar}}$. We compare $\tilde{f}_X$ and $\hat{f}_{X,\text{Sarkar}}$ when $f_X = f_4$ (a truncated normal mixture), with NSR = 20% and $n = 500, 1000$ or 5000. In this case, $\hat{f}_{X,\text{Sarkar}}$ worked the best, although $\tilde{f}_X$, which is consistent, gave reasonable results. As expected, both estimators improved as n increased.

To illustrate the loss incurred by computing an upper bound to S_X from the data, in Figure 3 we compare $\hat{f}_X$ that assumes S_X known and its version $\tilde{f}_X$ with unknown S_X , for density f_1 when NSR = 10% and $n = 500$. While $\hat{f}_X$ worked the best, $\tilde{f}_X$ performed only slightly worse. Here too, $\hat{f}_{X,\text{Sarkar}}$, also shown in Figure 3, overestimated the height of the mode of f_X . Likewise, in Figure 4 where we depict $\hat{f}_X$, $\tilde{f}_X$ and $\hat{f}_{X,\text{Sarkar}}$ for density f_2 when NSR = 20% and $n = 500$ or 5000, we see that while the estimators improved as n increased, $\hat{f}_{X,\text{Sarkar}}$ looks like a sharpened version of f_X , even for $n = 5000$.

Overall, these examples illustrate that, in finite samples, non consistent approximation methods such as $\hat{f}_{X,\text{Sarkar}}$ can give a very good approximation to f_X , especially in smooth multimodal situations, but can also lead to biased estimators in other cases (e.g. with overestimated peaks in the case of $\hat{f}_{X,\text{Sarkar}}$). On the other hand, our consistent estimator $\tilde{f}_X$ can work very well in finite samples, but for smooth multimodal densities such as f_4 , it can be outperformed by an approximation method such as $\hat{f}_{X,\text{Sarkar}}$.

7.2. Application to NHANES data

We analysed nutrition data from the 2017-March 2020 Pre-Pandemic Dietary NHANES dataset (see [30]), which includes detailed daily intake information collected from 2017 to 2020. Our interest was in the ratio of dietary fibre (in grams) over energy intake (in 10^4 kcal) for individuals who reported positive intakes on two 24-hour dietary recall (24-HR) interviews. For the i th individual, we took $W_{i,j} = m(\text{dietary fibre}_{i,j} / \text{energy}_{i,j})$ in (2.1), where $m(x) = 2(\sqrt{x} - 1)$ is the Box-Cox transformation with parameter $\lambda = 1/2$ and dietary fibre $_{i,j}$ and energy $_{i,j}$ are the i th individual's reported dietary fibre and energy intake at the j th 24-HR, for $i = 1, \dots, n$ and $j = 1, 2$. Our goal was to estimate the density f_X of the transformed long-term ratio intake X_i from the sample of size $n = 10,341$ obtained after deleting the outlying individuals for which $W_{i,j} > \hat{q}_{W,0.75} + 2\text{IQR}$, with $\text{IQR} = \hat{q}_{W,0.75} - \hat{q}_{W,0.25}$ and $\hat{q}_{W,\alpha}$ the empirical α quantile of the $W_{i,j}$'s.

We computed our estimator $\tilde{f}_X$ implemented as in Section 7.1 and $\hat{f}_{X,\text{Sarkar}}$, assuming that $U_{i,j} \sim N(0, 1)$. We also computed the naive standard KDE $\hat{f}_{\bar{W},\text{KDE}}$ of the density of $\bar{W}_i = (W_{i,1} + W_{i,2})/2$

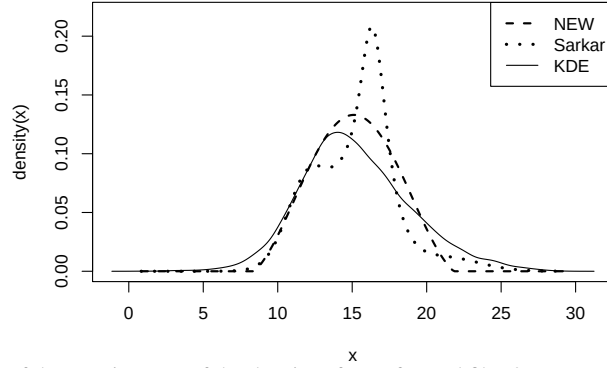


Figure 5. Comparison of three estimators of the density of transformed fibre/energy ratio for NHANES data: naive $\hat{f}_{\bar{W},\text{KDE}}$ (KDE), $\hat{f}_{\bar{X}}$ (NEW) and $\hat{f}_{\bar{X},\text{Sarkar}}$ (Sarkar).

with PI bandwidth, which ignores the measurement errors. The three estimators are shown in Figure 5. Unsurprisingly, we see that $\hat{f}_{\bar{X}}$ and $\hat{f}_{\bar{X},\text{Sarkar}}$ were both sharpened versions of the naive KDE, which is consistent for the smoothed out version $f_{\bar{W}}$ of $f_{\bar{X}}$. While we do not know the true density of the population, in this example, $\hat{f}_{\bar{X},\text{Sarkar}}$ seems to be undersmoothed, similar to the simulated examples shown in Figures 1 and 4, displaying two modes of dietary fibre consumption instead of one for $\hat{f}_{\bar{X}}$.

8. Other extensions

In this section we show how to extend our method to the case where more than two replicates are observed per individual, and we show that the same ideas can be used to estimate the variance function σ^2 and in the related errors-in-variable regression estimation problem.

8.1. Case with more than 2 replicates

To extend our procedure to the case with $J > 2$ replicates $W_{i,j}$, $j = 1, \dots, J$ per individual, we apply it to a sample of averages $(\tilde{W}_{i,1}, \tilde{W}_{i,2})$ defined, for $i = 1, \dots, n$, by

$$\begin{aligned}\tilde{W}_{i,1} &= \frac{1}{\lceil J/2 \rceil} \sum_{j=1}^{\lceil J/2 \rceil} W_{i,j} = X_i + \sigma(X_i) \tilde{U}_{i,1}, \text{ where } \tilde{U}_{i,1} = \frac{1}{\lceil J/2 \rceil} \sum_{j=1}^{\lceil J/2 \rceil} U_{i,1}, \\ \tilde{W}_{i,2} &= \frac{1}{J - \lceil J/2 \rceil} \sum_{j=\lceil J/2 \rceil+1}^J W_{i,j} = X_i + \sigma(X_i) \tilde{U}_{i,2}, \text{ where } \tilde{U}_{i,2} = \frac{1}{J - \lceil J/2 \rceil} \sum_{j=\lceil J/2 \rceil+1}^J U_{i,2}.\end{aligned}$$

To do this, for $j = 1, 2$, let $\tilde{W}_j$ and $\tilde{U}_j$ denote generic $\tilde{W}_{i,j}$'s and $\tilde{U}_{i,j}$'s. Similar to (3.5), we have

$$\begin{aligned}\mathbb{E}(\tilde{W}_1^{\ell-\nu} \tilde{W}_2^\nu) &= \sum_{r=0}^{\lfloor \ell/2 \rfloor} \sum_{r'=0}^r 1\{2r' \leq \nu\} 1\{2r - 2r' \leq \ell - \nu\} \binom{\nu}{2r'} \binom{\ell - \nu}{2r - 2r'} \\ &\quad \times \mathbb{E}(\tilde{U}_2^{2r'}) \mathbb{E}(\tilde{U}_1^{2r-2r'}) \mathbb{E}\{X^{\ell-2r} \sigma^{2r}(X)\},\end{aligned}$$

so that we can write $\mathcal{M}_{\tilde{W}}(\ell) = \mathcal{M}_{\tilde{U}}(\ell) \mathcal{M}_{X,\sigma}(\ell)$ for $\mathcal{M}_{X,\sigma}(\ell)$ as in Section 3, if we define

$$\mathcal{M}_{\tilde{W}}(\ell) = (\mathbb{E}(\tilde{W}_1^\ell), \mathbb{E}(\tilde{W}_1^{\ell-1} \tilde{W}_2), \dots, \mathbb{E}(\tilde{W}_1^{[\ell/2]} \tilde{W}_2^{[\ell/2]}))$$

and let $\mathcal{M}_{\tilde{U}}(\ell)$ be the known $(\lfloor \ell/2 \rfloor + 1) \times (\lfloor \ell/2 \rfloor + 1)$ matrix with $(\nu + 1, r + 1)$ th element

$$\mathcal{M}_{\tilde{U},\nu+1,r+1}(\ell) = \sum_{r'=0}^r 1\{2r' \leq \nu\} 1\{2r - 2r' \leq \ell - \nu\} \binom{\nu}{2r'} \binom{\ell - \nu}{2r - 2r'} \mathbb{E}(\tilde{U}_2^{2r'}) \mathbb{E}(\tilde{U}_1^{2r-2r'}).$$

Letting $\hat{\mathcal{M}}_{\tilde{W}}(\ell) = n^{-1} \sum_{i=1}^n (\tilde{W}_{i,1}^\ell, \tilde{W}_{i,1}^{\ell-1} \tilde{W}_{i,2}, \dots, \tilde{W}_{i,1}^{[\ell/2]} \tilde{W}_{i,2}^{[\ell/2]})^T$, we can estimate f_X by $\hat{f}_X$ at (3.14), where $\hat{f}_{X,k}$ is defined by (3.13), taking there $\hat{\mu}_{X,\ell}$ equal to $\hat{\mu}_{X,\ell} = \mathbf{e}_{0,\ell} \{\mathcal{M}_{\tilde{U}}(\ell)\}^{-1} \hat{\mathcal{M}}_{\tilde{W}}(\ell)$.

For a finite number $J > 2$ of replicates, the rates of convergence for this version of our estimator are identical to those for the case where $J = 2$. It is theoretically interesting to consider the asymptotic behaviour of the estimator when $J \rightarrow \infty$ as $n \rightarrow \infty$. Consider for simplicity the case where J is even, so that for $j = 1, 2$, we can write $\tilde{W}_{i,j} = X_i + \tilde{\sigma}(X_i) \tilde{U}_{i,j}$, where $\tilde{\sigma} = \sqrt{2/J} \sigma$ and $\tilde{U}_{i,j} = \sqrt{J/2} \tilde{U}_{i,j}$. Then the $\tilde{U}_{i,j}$'s are i.i.d. random variables with mean zero and variance 1, and if $J \rightarrow \infty$ as $n \rightarrow \infty$, we have $\|\tilde{\sigma}\|_\infty \rightarrow 0$ as $n \rightarrow \infty$. Therefore, the asymptotic regime where $J \rightarrow \infty$ as $n \rightarrow \infty$ is a particular case of the small error setting where we observe $J = 2$ replicates $W_{i,1}$ and $W_{i,2}$ for $i = 1, \dots, n$ as in model (2.1), and σ is such that $\|\sigma\|_\infty \rightarrow 0$ as $n \rightarrow \infty$. For the sake of completeness, in Appendix A.4 we study the MISE_n of our estimator in the latter setting. However, in that case, our estimator is not needed as many well established estimators can be applied, such as SIMEX methods or standard deconvolution estimators designed for homoscedastic errors. Indeed, while those methods are not consistent when $\|\sigma\|_\infty \not\rightarrow 0$ as $n \rightarrow \infty$, they are consistent when $\|\sigma\|_\infty \rightarrow 0$ as $n \rightarrow \infty$ (see e.g. [8] for results on small error variances).

8.2. Variance estimation

We can extend our procedure for estimating the function σ^2 in (2.1). Assume that $f_X \in L_2[-1, 1]$ and let $h := \sigma^2 \cdot f_X$. If σ is bounded, then $h \in L_2([-1, 1])$ and we can express h as $h = \sum_{k=0}^\infty h_k \cdot \varphi_k$, with

$$h_k = \langle h, \varphi_k \rangle = 2^k \sqrt{k+1/2} \sum_{\ell=0}^k \binom{k}{\ell} \binom{(k+\ell-1)/2}{k} \cdot \mathbb{E}\{X^\ell \sigma^2(X)\}.$$

To estimate h , it suffices to estimate $S_\ell \equiv \mathbb{E}\{X^\ell \sigma^2(X)\}$ for $\ell = 0, 1, \dots$. Similar to (3.10), using (3.9), if $\mu_{\ell+2}$ is finite we have $S_\ell = \mathbf{e}_{1,\ell+2} \{\mathcal{M}_U(\ell+2)\}^{-1} \mathcal{M}_W(\ell+2)$ for $\ell = 0, 1, \dots$, where $\mathbf{e}_{1,\ell+2}$ denote the $(\lfloor \ell/2 + 1 \rfloor + 1)$ -dimensional row vector whose second component equals 1 and all others are 0. Taking $\mathcal{M}_W$ as in (3.11), for $\ell = 0, 1, \dots$ such that $\mu_{\ell+2}$ is finite, we can estimate S_ℓ by

$$\hat{S}_\ell = \mathbf{e}_{1,\ell+2} \{\mathcal{M}_U(\ell+2)\}^{-1} \hat{\mathcal{M}}_W(\ell+2). \quad (8.1)$$

Then we can estimate h_k by

$$\hat{h}_k = 2^k \sqrt{k+1/2} \sum_{\ell=0}^k \binom{k}{\ell} \binom{(k+\ell-1)/2}{k} \cdot \hat{S}_\ell. \quad (8.2)$$

Finally, we can estimate $\sigma^2(x)$ by

$$\hat{\sigma}^2(x) = \hat{h}(x) / \max\{\hat{f}_X(x), \zeta/2\}, \text{ where } \hat{h} = \sum_{k=0}^{\lceil K \rceil} w_{h,k} \cdot \hat{h}_k \cdot \varphi_k, \quad (8.3)$$

with $\hat{f}_X$ as in (3.14), $w_{h,k} = 1\{k < K\} + (K+1-k) \cdot 1\{K \leq k < K+1\}$ and where K is a truncation parameter and $\zeta > 0$ is a ridge parameter that prevents the denominator from getting too close to zero.

We can extend the result of Theorem 2 to the estimator $\hat{\sigma}^2$ of $\sigma^2 = h/f_X$ on intervals $[a, b] \subseteq \mathcal{S}_X$ such that $\inf_{x \in [a, b]} f_X(x) \geq \zeta$. That is, similar to nonparametric curve estimators that take the form of a ratio of two curve estimators (e.g., kernel regression estimators), we can derive properties of $\hat{\sigma}^2$ in regions where the denominator f_X of h/f_X is sufficiently far from zero. Specifically, we prove the following theorem in Appendix A.2.

Theorem 3. *For any given integer $m \geq 1$, if $U \sim N(0, 1)$ and $\hat{\sigma}^2$ at (8.3) is computed with $K = K_n = \lfloor c_0(\log n)/(\log \log n) \rfloor$ for some constant $c_0 \in (0, 1)$, and $[a, b] \subseteq \mathcal{S}_X$ is an interval such that $\zeta \leq \inf_{x \in [a, b]} f_X(x)$, then for any finite positive constants c and d , as $n \rightarrow \infty$ we have*

$$\sup_{f_X, f_X \sigma^2 \in \mathcal{F}(c, m), \|\sigma\|_\infty \leq d} \mathbb{E} \int_a^b |\hat{\sigma}^2(x) - \sigma^2(x)|^2 dx = O\{(\log n / \log \log n)^{-2m}\}. \quad (8.4)$$

Computing $\hat{\sigma}^2$ in practice requires choosing K and ζ . Proposing a sophisticated data-driven selection rule of these parameters, appropriate for the estimation of σ^2 , is beyond the scope of this paper, but in Appendix D.2, we report a small simulation study where we used simple rules, described there, to select ζ and K . There, instead of assuming that the support $\mathcal{S}_X$ is known, we estimated an upper bound $[A_X, B_X]$ to it as in Section 6.3 and applied the estimator of σ^2 to the data rescaled as at (6.1).

8.3. Errors-in-variables regression

Assuming again that that $f_X \in L_2[-1, 1]$, the same ideas can be used to estimate some bounded regression curve $g(x) = \mathbb{E}(Y|X=x)$ from i.i.d. observations $(W_{i,1}, W_{i,2}, Y_i)$, $i = 1, \dots, n$, where the $W_{i,j}$'s are as in (2.1) and the Y_i 's are independent of the $U_{i,j}$'s and have two finite moments. In the same spirit as above, we write $g = g f_X / f_X$ and express $g f_X = \sum_{k=0}^{\infty} (g f_X)_k \cdot \varphi_k$, where

$$(g f_X)_k = \langle g f_X, \varphi_k \rangle = 2^k \sqrt{k+1/2} \sum_{\ell=0}^k \binom{k}{\ell} \binom{(k+\ell-1)/2}{k} \cdot \mathbb{E}\{X^\ell g(X)\}. \quad (8.5)$$

We estimate $\mathbb{E}\{X^0 g(X)\}$ by $\hat{\mathbb{E}}\{X^0 g(X)\} = \sum_{i=1}^n Y_i / n$. To estimate $\mathbb{E}\{X^\ell g(X)\}$ for $\ell \geq 1$, similar to (3.5) we have $\mathcal{M}_{YW}(\ell) = \mathcal{M}_U(\ell) \mathcal{M}_{YX, \sigma}(\ell)$ for $\ell = 1, 2, \dots$ such that μ_ℓ is finite, where

$$\begin{aligned} \mathcal{M}_{YW}(\ell) &= (\mathbb{E}(Y W_1^\ell), \mathbb{E}(Y W_1^{\ell-1} W_2), \dots, \mathbb{E}(Y W_1^{\lceil \ell/2 \rceil} W_2^{\lfloor \ell/2 \rfloor}))^T \\ \mathcal{M}_{YX, \sigma}(\ell) &= (\mathbb{E}(Y X^\ell), \mathbb{E}(Y \sigma^2(X) X^{\ell-2}), \dots, \mathbb{E}(Y \sigma^{2\lfloor \ell/2 \rfloor}(X) X^{\ell-2\lfloor \ell/2 \rfloor}))^T. \end{aligned}$$

We can estimate $\mathcal{M}_{YW}(\ell)$ by

$$\hat{\mathcal{M}}_{YW}(\ell) = \frac{1}{n} \sum_{i=1}^n Y_i \left(W_{i,1}^\ell, W_{i,1}^{\ell-1} W_{i,2}, \dots, W_{i,1}^{\lceil \ell/2 \rceil} W_{i,2}^{\lfloor \ell/2 \rfloor} \right)^T,$$

and under (3.8), we deduce an estimator of $\mathbb{E}\{X^\ell g(X)\}$ for $\ell \geq 1$ by taking

$$\hat{\mathbb{E}}\{X^\ell g(X)\} = \mathbf{e}_{0,j} \{\mathcal{M}_U(\ell)\}^{-1} \hat{\mathcal{M}}_{YW}(\ell).$$

Finally we can estimate g by $\hat{g} = g\hat{f}_X / \hat{f}_X$, with $\hat{f}_X$ in (3.14) and $g\hat{f}_X = \sum_{k=0}^{[K]} w_{g,k} \cdot (g\hat{f}_X)_k \cdot \varphi_k$, where $(g\hat{f}_X)_k$ is obtained by replacing $\mathbb{E}\{X^\ell g(X)\}$ in (8.5) by $\hat{\mathbb{E}}\{X^\ell g(X)\}$ and where $w_{g,k} = 1\{k < K\} + (K+1-k) \cdot 1\{K \leq k < K+1\}$ with K a truncation parameter.

Acknowledgments

The authors would like to thank the editor, the associate editor and the referee for their review and their insightful and constructive comments.

Funding

Delaigle is supported by a discovery project (DP230100147) from the Australian Research Council. Meister is supported by the Research Unit 5381, granted by the German Research Foundation (DFG).

References

- [1] BELOMESTNY, D. and GOLDENSHLUGER, A. (2020). Nonparametric density estimation from observations with multiplicative measurement errors. *Ann. Inst. Henri Poincaré Probab. Stat.* **56** 36–67.
- [2] BELOMESTNY, D. and PANOV, V. (2018). Semiparametric estimation in the normal variance-mean mixture model. *Statistics* **52** 571–589.
- [3] BRENNER MIGUEL, S., COMTE, F. and JOHANNES, J. (2021). Spectral cut-off regularisation for density estimation under multiplicative measurement errors. *Electron. J. Stat.* **15** 3551–3573.
- [4] BUTUCEA, C. and MATIAS, C. (2005). Minimax estimation of the noise level and of the signal density in a semiparametric convolution model. *Bernoulli* **11** 309–340.
- [5] BUTT, N.S. and KHALIL, M.G. (2020). A new bimodal distribution for modeling asymmetric bimodal heavy-tail real lifetime data. *Symmetry* **12** 2058.
- [6] CARROLL, R.J., DELAIGLE, A. and HALL, P. (2007). Nonparametric regression estimation from data contaminated by a mixture of Berkson and classical errors. *J. R. Stat. Soc. Series B* **69** 859–878.
- [7] CARROLL, R.J. and HALL, P. (1988). Optimal rates of convergence for deconvolving a density. *J. Amer. Statist. Assoc.* **83** 1184–1186.
- [8] DELAIGLE, A. (2008). An alternative view of the deconvolution problem. *Statist. Sinica* **18** 1025–1045.
- [9] DELAIGLE, A. (2014). Nonparametric kernel methods with errors-in-variables: constructing estimators, computing them, and avoiding common mistakes. *Aust. N. Z. J. Stat.* **56** 105–124.
- [10] DELAIGLE, A., FAN, J. and CARROLL, R.J. (2009). A design-adaptive local polynomial estimator for the errors-in-variables problem. *J. Amer. Statist. Assoc.* **104** 348–359.
- [11] DELAIGLE, A. and GIJBELS, I. (2006). Data-driven boundary estimation in deconvolution problems. *Comput. Statist. Data Anal.* **50** 1965–1994.
- [12] DELAIGLE, A. and GIJBELS, I. (2006). Estimation of boundary and discontinuity points in deconvolution problems. *Statist. Sinica* **16** 773–788.
- [13] DELAIGLE, A. and HALL, P. (2008). Using SIMEX for smoothing-parameter choice in errors-in-variables problems. *J. Amer. Statist. Assoc.* **103** 280–287.
- [14] DELAIGLE, A., HALL, P. and MEISTER, A. (2008). On deconvolution with repeated measurements. *Ann. Statist.* **36** 665–685.

- [15] DELAIGLE, A. and MEISTER, A. (2007). Nonparametric regression estimation in the heteroscedastic errors-in-variables problem. *J. Amer. Statist. Assoc.* **102** 1416–1426.
- [16] DELAIGLE, A. and MEISTER, A. (2008). Density estimation with heteroscedastic error. *Bernoulli* **14** 562–579.
- [17] FAN, J. (1991). On the optimal rates of convergence for nonparametric deconvolution problems. *Ann. Statist.* **19** 1257–1272.
- [18] FAN, J. and TRUONG, Y.K. (1993). Nonparametric regression with errors in variables. *Ann. Statist.* **21** 1900–1925.
- [19] GASSIAT, E., LE CORFF, S. and LEHÉRICY, L. (2022). Deconvolution with unknown noise distribution is possible for multivariate signals. *Ann. Statist.* **50** 303–323.
- [20] GOLDENSHLUGER, A. (2002). Density deconvolution in the circular structural model. *J. Multivariate Anal.* **84** 350–375.
- [21] HALL, P. and QIU, P. (2005). Discrete-transform approach to deconvolution problems. *Biometrika* **92** 135–148.
- [22] HU, Y. and SCHENNACH, S. (2008). Instrumental variable treatment of nonclassical measurement error models. *Econometrica* **76** 195–216.
- [23] JEON, J.M., PARK, B.U. and VAN KEILEGOM, I. (2022). Nonparametric regression on Lie groups with measurement errors. *Ann. Statist.* **50** 2973–3008.
- [24] KIM, P.T. and RICHARDS, D.ST.P., Deconvolution density estimation on the space of positive definite symmetric matrices, in: *Nonparametric Statistics and Mixture Models – A Festschrift in Honor of Thomas P Hettmansperger*, 2011, 147–168.
- [25] LI, M., MA, Y. and LI, R. (2019). Semiparametric regression for measurement error model with heteroscedastic error. *J. Multivariate Anal.* **171** 320–338.
- [26] MCINTYRE, J. and STEFANSKI, L.A. (2011). Density estimation with replicate heteroscedastic measurements. *Ann. Instit. Statist. Math.* **63** 81–99.
- [27] MEISTER, A. (2006). Support estimation via moment estimation in presence of noise. *Statistics* **40** 259–275.
- [28] MEISTER, A. (2007). Deconvolving compactly supported densities. *Math. Methods Statist.* **16** 63–76.
- [29] MEISTER, A., *Deconvolution Problems in Nonparametric Statistics*, Lecture Notes in Statistics, Vol. **193**, 2009, Springer, Berlin/Heidelberg/New York.
- [30] [dataset] NHANES (2022). Dietary Interview - Total Nutrient Intakes (First Day, Second Day): Energy and Dietary Fiber. Hyattsville, MD: HHS, CDC. <http://wwwn.cdc.gov/nchs/nhanes/search/datapage.aspx?Component=Dietary&Cycle=2017-2020>.
- [31] PENSKY, M. and VIDAKOVIC, B. (1999). Adaptive wavelet estimator for nonparametric density deconvolution. *Ann. Statist.* **27** 2033–2053.
- [32] SARKAR, A., MALLICK, B.K. and CARROLL, R.J. (2014). Bayesian semiparametric regression in the presence of conditionally heteroscedastic measurement and regression errors. *Biometrics* **70** 823–834.
- [33] SARKAR, A., MALLICK, B.K., STAUDENMAYER, J., PATI, D. and CARROLL, R.J. (2014). Bayesian semiparametric density deconvolution in the presence of conditionally heteroscedastic measurement errors. *J. Comput. Graph. Statist.* **25** 1101–1125.
- [34] SARKAR, A., PATI, D., CHAKRABORTY, A., MALLICK, B.K. and CARROLL, R.J. (2018). Bayesian semiparametric multivariate density deconvolution. *J. Amer. Statist. Assoc.* **113** 401–416.
- [35] STAUDENMAYER, J., RUPPERT, D. and BUONACCORSI, J.P. (2008). Density estimation in the presence of heteroscedastic measurement error. *J. Amer. Statist. Assoc.* **103** 726–736.
- [36] STEFANSKI, L.A. and CARROLL, R.J. (1990). Deconvolving kernel density estimators. *Statistics* **21** 169–184.
- [37] WAND, M. P. and JONES, M.C. (1994). *Kernel Smoothing*. CRC press.
- [38] WANG, H. (2018). A new and sharper bound for Legendre expansion of differentiable functions. *Appl. Math. Lett.* **85** 95–102.
- [39] YI, G., DELAIGLE, A. and GUSTAFSON, P. (2021). *Handbook on Measurement Error Models*. CRC press.

Appendix A: Technical arguments

A.1. Proofs of the main results

Proof of Lemma 1. (a). Write x as $x = (x(0), \dots, x(\ell/2))^T$. For any vector $v = (v(0), \dots, v(\ell))^T$ such that $v(2r) = x(r)$ for $r = 0, \dots, \lfloor \ell/2 \rfloor$ and $v(r) = 0$ for odd r , we have

$$\begin{aligned} x^T \mathcal{M}_U(\ell)^T \mathcal{M}_U(\ell) x &= \sum_{v=0}^{\lfloor \ell/2 \rfloor} \left| \sum_{i=0}^{\lfloor \ell/2 \rfloor} x(i) \mathcal{M}_{U, v+1, i+1}(\ell) \right|^2 \\ &= \sum_{v=0}^{\lfloor \ell/2 \rfloor} \left| \sum_{i=0}^{\lfloor \ell/2 \rfloor} x(i) \sum_{j=0}^i 1\{2j \leq v\} 1\{2i - 2j \leq \ell - v\} \binom{v}{2j} \binom{\ell - v}{2i - 2j} \mu_{2j} \mu_{2i - 2j} \right|^2 \\ &= \sum_{v=0}^{\lfloor \ell/2 \rfloor} \left| \sum_{r=0}^{\ell} v(r) \sum_{\kappa=0}^r 1\{\kappa \leq v\} 1\{r - \kappa \leq \ell - v\} \binom{v}{\kappa} \binom{\ell - v}{r - \kappa} \mu_{\kappa} \mu_{r - \kappa} \right|^2. \end{aligned}$$

(Note that this results holds even if $v(r) \neq 0$ for odd r , since $\mu_{\kappa} \mu_{r - \kappa} = 0$ when r is odd, because $\mu_{r - \kappa} = 0$ if $r - \kappa$ is odd and $\mu_{\kappa} = 0$ if κ is odd. However putting $v(r) = 0$ for odd r will be convenient for some calculations below.)

Moreover,

$$\begin{aligned} &\sum_{v=\lfloor \ell/2 \rfloor + 1}^{\ell} \left| \sum_{r=0}^{\ell} v(r) \sum_{\kappa=0}^r 1\{\kappa \leq v\} 1\{r - \kappa \leq \ell - v\} \binom{v}{\kappa} \binom{\ell - v}{r - \kappa} \mu_{\kappa} \mu_{r - \kappa} \right|^2 \\ &= \sum_{v'=\lfloor \ell/2 \rfloor - 1}^{\ell - \lfloor \ell/2 \rfloor - 1} \left| \sum_{r=0}^{\ell} v(r) \sum_{\kappa'=0}^r 1\{r - \kappa' \leq \ell - v'\} 1\{\kappa' \leq v'\} \binom{\ell - v'}{r - \kappa'} \binom{v'}{\kappa'} \mu_{r - \kappa'} \mu_{\kappa'} \right|^2 \\ &\leq x^T \mathcal{M}_U(\ell)^T \mathcal{M}_U(\ell) x, \end{aligned}$$

where we used the changes of variable $v' = \ell - v$ and $\kappa' = r - \kappa$. Therefore,

$$\begin{aligned} x^T \mathcal{M}_U(\ell)^T \mathcal{M}_U(\ell) x &\geq \frac{1}{2} \sum_{v=0}^{\ell} \left| \sum_{r=0}^{\ell} v(r) \sum_{\kappa=0}^r 1\{\kappa \leq v \leq \kappa + \ell - r\} \binom{v}{\kappa} \binom{\ell - v}{r - \kappa} \mu_{\kappa} \mu_{r - \kappa} \right|^2 \\ &= \frac{1}{2} \sum_{v=0}^{\ell} \binom{\ell}{v}^{-2} \left| \sum_{r=0}^{\ell} v(r) \binom{\ell}{r} \sum_{\kappa=0}^r 1\{\kappa \leq v \leq \kappa + \ell - r\} \binom{r}{\kappa} \binom{\ell - r}{v - \kappa} \mu_{\kappa} \mu_{r - \kappa} \right|^2. \end{aligned}$$

Now for all coefficients α_v , $v = 0, \dots, \ell$, and all $\theta \in \mathbb{R}$, using the Cauchy-Schwarz inequality we have

$$\left| \sum_{v=0}^{\ell} \alpha_v \cdot \theta^v \right|^2 = \left| \sum_{v=0}^{\ell} \alpha_v \binom{\ell}{v}^{-1} \binom{\ell}{v} \cdot \theta^v \right|^2 \leq \sum_{v=0}^{\ell} \alpha_v^2 \binom{\ell}{v}^{-2} \cdot \sum_{v=0}^{\ell} \binom{\ell}{v}^2 \cdot \theta^{2v}.$$

Moreover,

$$\sum_{v=0}^{\ell} \binom{\ell}{v}^2 \cdot \theta^{2v} \leq 2^{\ell} \sum_{v=0}^{\ell} \binom{\ell}{v} \cdot \theta^{2v} = (2 + 2\theta^2)^{\ell},$$

so that

$$\sum_{v=0}^{\ell} \alpha_v^2 \binom{\ell}{v}^{-2} \geq \left| \sum_{v=0}^{\ell} \alpha_v \cdot \theta^v \right|^2 \cdot (2 + 2\theta^2)^{-\ell}.$$

Taking

$$\alpha_v = \sum_{r=0}^{\ell} v(r) \binom{\ell}{r} \sum_{\kappa=0}^r 1_{\{\kappa \leq v \leq \kappa + \ell - r\}} \binom{r}{\kappa} \binom{\ell - r}{v - \kappa} \mu_{\kappa} \mu_{r - \kappa},$$

we deduce that,

$$\begin{aligned} 2x^T \mathcal{M}_U(\ell)^T \mathcal{M}_U(\ell) x &= \sum_{v=0}^{\ell} \binom{\ell}{v}^{-2} \alpha_v^2 \\ &\geq \sup_{\theta \in \mathbb{R}} \left| \sum_{v=0}^{\ell} \theta^v \sum_{r=0}^{\ell} v(r) \binom{\ell}{r} \sum_{\kappa=0}^r 1_{\{\kappa \leq v \leq \kappa + \ell - r\}} \binom{r}{\kappa} \binom{\ell - r}{v - \kappa} \mu_{\kappa} \mu_{r - \kappa} \right|^2 \cdot (2 + 2\theta^2)^{-\ell} \\ &= \sup_{\theta \in \mathbb{R}} \left| \sum_{r=0}^{\ell} v(r) \binom{\ell}{r} \sum_{\kappa=0}^r \binom{r}{\kappa} \mu_{\kappa} \mu_{r - \kappa} \sum_{v=\kappa}^{\kappa + \ell - r} \binom{\ell - r}{v - \kappa} \theta^v \right|^2 \cdot (2 + 2\theta^2)^{-\ell} \\ &= \sup_{\theta \in \mathbb{R}} \left| \sum_{r=0}^{\ell} v(r) \binom{\ell}{r} \sum_{\kappa=0}^r \binom{r}{\kappa} \theta^{\kappa} \mu_{\kappa} \mu_{r - \kappa} \sum_{v=0}^{\ell - r} \binom{\ell - r}{v} \theta^v \right|^2 \cdot (2 + 2\theta^2)^{-\ell} \\ &\geq 2^{-\ell} \sup_{\theta > -1} \left(\frac{(1 + \theta)^2}{1 + \theta^2} \right)^{\ell} \left| \sum_{r=0}^{\ell} v(r) \binom{\ell}{r} \mathbb{E} \left(\frac{\theta}{1 + \theta} U_1 + \frac{1}{1 + \theta} U_2 \right)^r \right|^2, \end{aligned} \quad (\text{A.1})$$

where we used the fact that $\sum_{v=0}^{\ell - r} \binom{\ell - r}{v} \theta^v = (1 + \theta)^{\ell - r}$.

Therefore, and since as long as we take $\theta > -1$ away from -1 , the contribution of the term $(1 + \theta)^2 / (1 + \theta^2)$ is strictly positive, then letting $t = \theta / (1 + \theta)$, $x^T \mathcal{M}_U(\ell)^T \mathcal{M}_U(\ell) x = 0$ implies that

$$\sum_{r=0}^{\ell} v(r) \binom{\ell}{r} \mathbb{E} \{ t U_1 + (1 - t) U_2 \}^r = 0,$$

for all $t > t_0$, for some finite $t_0 > 0$. The left side of this equation is a polynomial in t of degree $\leq \ell$. The leading monomial t^{ℓ} has the coefficient $v(\ell) \cdot \mathbb{E}(U_1 - U_2)^{\ell}$, which must be equal to zero. For odd ℓ , this is satisfied by definition of v since then we have $v(\ell) = 0$. For ℓ even, since $\mathbb{E}(U_1 - U_2)^{\ell} > 0$, we must also have $v(\ell) = 0$. Thus $v(\ell) = 0$ for both ℓ even and odd, which implies that the leading monomial is $t^{\ell - 1}$, which has the coefficient $v(\ell - 1) \binom{\ell}{\ell - 1} \mathbb{E}(U_1 - U_2)^{\ell - 1}$. Treating again the odd and even case separately, we deduce that $v(\ell - 1) = 0$. The same argument can be iterated to show that $v(r) \binom{\ell}{r} \mathbb{E}(U_1 - U_2)^r = 0$ for $r = 0, \dots, \ell$, from which we deduce, treating r odd and r even separately as above, that $v(r) = 0$ for $r = 0, \dots, \ell$ and, hence that $v = 0$. Thus, $\mathcal{M}_U$ is invertible.

(b). In the particular case where $U_1, U_2 \sim \mathbf{N}(0, 1)$, using the fact that

$$\frac{\theta}{1 + \theta} U_1 + \frac{1}{1 + \theta} U_2 \sim \mathbf{N} \left(0, \frac{1 + \theta^2}{(1 + \theta)^2} \right),$$

the r th moment of this term vanishes for odd r and equals

$$\mathbb{E}\left(\frac{\theta}{1+\theta}U_1 + \frac{1}{1+\theta}U_2\right)^r = \left(\frac{1+\theta^2}{(1+\theta)^2}\right)^{r/2} \cdot (r-1)!! ,$$

for even r . Combining this with the result $(r-1)!! = r!/(2^{r/2}(r/2)!)$, and letting

$$w(r) = \begin{cases} v(r) \binom{\ell}{r} \binom{r}{r/2} (r/2)! , & \text{for even } r , \\ 0 , & \text{for odd } r . \end{cases}$$

$\mu = \sqrt{(1+\theta^2)/(1+\theta)^2} \geq \sqrt{1/2}$ for $\theta > -1$, $\mu^* = \sqrt{2}\mu \geq 1$ and $\kappa = 1/\mu^*$, we find that (A.1) is equal to

$$\begin{aligned} 2^{-\ell} \sup_{\mu \geq 1/\sqrt{2}} \mu^{-2\ell} \left| \sum_{r=0}^{\ell} w(r) 2^{-r/2} \mu^r \right|^2 &= \sup_{\mu^* \geq 1} (\mu^*)^{-2\ell} \left| \sum_{r=0}^{\ell} w(r) 2^{-r} (\mu^*)^r \right|^2 \\ &= \sup_{\kappa \in [0,1]} \left| \sum_{r=0}^{\ell} w(r) 2^{-r} \kappa^{\ell-r} \right|^2 \\ &= \sup_{\kappa \in [0,1]} \left| \sum_{j=0}^{\ell} w(\ell-j) 2^{j-\ell} \kappa^j \right|^2 \\ &= \sup_{\kappa \in [-1,1]} \left| \sum_{j=0}^{\ell} w(\ell-j) 2^{j-\ell} \kappa^j \right|^2 . \end{aligned} \quad (\text{A.2})$$

In the last step we have used the fact that either all terms with even exponent or all terms with odd exponent in the above polynomial in κ vanish because $v(\ell-j) = w(\ell-j) = 0$, for $\ell-j$ odd, i.e. for j even (if ℓ is odd) or for j odd (if ℓ is even). Working with a symmetric interval in κ will be useful later when we apply Markov brothers' inequality.

Now taking $\kappa = 0$, a first lower bound for (A.2) is $4^{-\ell} \cdot w^2(\ell)$. To find a second lower bound we use Markov brothers' inequality and Taylor's theorem for polynomials. For this, let $P(\kappa) = \sum_{j=0}^{\ell} w(\ell-j) 2^{j-\ell} \cdot \kappa^j$ be the polynomial of order ℓ in κ and note that for all $i = 1, \dots, \ell$, the derivative of order i with respect to κ is equal to $P^{(i)}(\kappa) = i! \sum_{j=i}^{\ell} w(\ell-j) 2^{j-\ell} \cdot \kappa^{j-i} / (j-i)!$, which is such that $\sup_{\kappa \in [-1,1]} |P^{(i)}(\kappa)| \geq |P^{(i)}(0)| = i! |w(\ell-i)| 2^{i-\ell}$. By Markov brothers' inequality we have

$$(2i-1)!! \sup_{\kappa \in [-1,1]} |P^{(i)}(\kappa)| \leq \ell^2(\ell^2-1)(\ell^2-2^2) \cdot (\ell^2-(i-1)^2) \sup_{\kappa \in [-1,1]} |P(\kappa)|$$

which implies that

$$i!(2i-1)!! |w(\ell-i)| 2^{i-\ell} \leq \ell^2(\ell^2-1)(\ell^2-2^2) \cdot (\ell^2-(i-1)^2) \sup_{\kappa \in [-1,1]} |P(\kappa)| ,$$

that is

$$\sup_{\kappa \in [-1,1]} |P(\kappa)| \geq \frac{i! 2^{i-\ell} (2i-1)!! |w(\ell-i)|}{\ell^2(\ell^2-1)(\ell^2-2^2) \cdot (\ell^2-(i-1)^2)}$$

$$\begin{aligned}
&= \frac{i!2^{i-\ell}(2i-1)!!|w(\ell-i)|}{\ell(\ell-1)(\ell-2) \cdot (\ell-(i-1)) \times \ell(\ell+1)(\ell+2) \cdot (\ell+i-1)} \\
&= \frac{i!2^{i-\ell}(2i-1)!!(\ell-i)!(\ell-1)!|w(\ell-i)|}{\ell!(\ell+i-1)!} \\
&= |w(\ell-i)|2^{1-\ell} \frac{i(2i-1)!(\ell-i)!}{\ell(\ell+i-1)!},
\end{aligned}$$

where we have used $(2i-1)!! = (2i-1)!/\{2^{i-1}(i-1)!\}$. Thus

$$\begin{aligned}
\sup_{\kappa \in [-1,1]} |P(\kappa)|^2 &\geq 4^{1-\ell} w^2(\ell-i) \frac{i^2(2i-1)!^2(\ell-i)!^2}{\ell^2(\ell+i-1)!^2} \\
&= 4^{1-\ell} v^2(\ell-i) \binom{\ell}{\ell-i}^2 \binom{\ell-i}{(\ell-i)/2}^2 (\ell/2-i/2)!^2 \frac{i^2(2i-1)!^2(\ell-i)!^2}{\ell^2(\ell+i-1)!^2} \\
&= 4^{1-\ell} v^2(\ell-i) \frac{\ell!^2}{(\ell-i)!^2 i!^2} \frac{(\ell-i)!^2}{(\ell/2-i/2)!^4} (\ell/2-i/2)!^2 \frac{i^2(2i-1)!^2(\ell-i)!^2}{\ell^2(\ell+i-1)!^2} \\
&= 4^{1-\ell} v^2(\ell-i) \frac{(\ell-1)!^2}{(\ell/2-i/2)!^2} \frac{(2i-1)!^2(\ell-i)!^2}{(i-1)!^2(\ell+i-1)!^2} \\
&= 4^{1-\ell} v^2(\ell-i) \left\{ \frac{(\ell-i)!}{(\ell/2-i/2)!} \binom{2i-1}{i} / \binom{\ell+i-1}{i} \right\}^2
\end{aligned}$$

for all $i = 1, \dots, \ell$ for which $\ell-i$ is even. We use the inequalities

$$\binom{\ell+i-1}{i} \leq 2^{\ell+i-1}, \quad \binom{2i-1}{i} \geq 4^{i-1}/\sqrt{i},$$

and conclude that for all $i = 1, \dots, \ell$ for which $\ell-i$ is even,

$$\sup_{\kappa \in [-1,1]} |P(\kappa)|^2 \geq 4^{1-\ell} v^2(\ell-i) \left\{ \frac{(\ell-i)!}{(\ell/2-i/2)!} \frac{4^{i-1}/\sqrt{i}}{2^{\ell+i-1}} \right\}^2 = i^{-1} 4^{i-2\ell} v^2(\ell-i) \left\{ \frac{(\ell-i)!}{(\ell/2-i/2)!} \right\}^2.$$

We use this term as well as $4^{-\ell} \cdot w^2(\ell)$ to get lower bounds on (A.2); and these bounds can be endowed with weights $\beta_0, \dots, \beta_\ell$ so that (A.2) is larger or equal to

$$\ell^{-1} 16^{-\ell} \sum_{i=0}^{\ell} v^2(\ell-i) \left\{ \frac{(\ell-i)!}{(\ell/2-i/2)!} \right\}^2 4^i \cdot \beta_i / \sum_{i=0}^{\ell} \beta_i$$

for all non-negative coefficients $\beta_0, \dots, \beta_\ell$. We choose

$$\beta_i = (\ell/2-i/2)!^2 \cdot 4^{-i} / \{(\ell-i)!^2\},$$

for all i such that $\ell-i$ is even; and $\beta_i = 0$ otherwise. As

$$\sum_{i=0}^{\ell} \beta_i \leq 4^{-\ell} \sum_{r=0}^{\infty} 16^r \cdot r!^2 / (2r)!^2,$$

the lower bound is equal to

$$\ell^{-1} 4^{-\ell} \sum_{i=0}^{\ell} v^2(\ell-i) / \sum_{r=0}^{\infty} 16^r \cdot r!^2 / (2r)!^2 = \ell^{-1} 4^{-\ell} \sum_{i=0}^{\lfloor \ell/2 \rfloor} x^2(i) / \sum_{r=0}^{\infty} 16^r \cdot r!^2 / (2r)!^2$$

as claimed in the lemma. $\square$

Proof of Theorem 1. (a). For $k > \lceil K \rceil$, define $w_k = 0$ and $\hat{f}_{X,k} = 1$. By Parseval's identity with respect to the orthonormal basis $\{\varphi_k\}_k$, we have

$$\text{MISE}_n(f_X) = \sum_{k=1}^{\infty} \mathbb{E} |w_k \cdot \hat{f}_{X,k} - f_{X,k}|^2 = \sum_{k=1}^{\infty} |w_k \mathbb{E}(\hat{f}_{X,k}) - f_{X,k}|^2 + \sum_{k=1}^{\lceil K \rceil} w_k^2 \text{Var}(\hat{f}_{X,k}). \quad (\text{A.3})$$

We study the squared bias (first sum in (A.3)) and the variance (second sum in (A.3)) terms separately.

For the squared bias term, it follows from (3.10) and (3.12) that $\mathbb{E}(\hat{\mu}_{X,\ell}) = \mathbb{E}(X^\ell)$ for $\ell = 1, \dots, \lceil K \rceil$. Therefore, $E(\hat{f}_{X,k}) = f_{X,k}$ for $k = 1, \dots, \lceil K \rceil$ and we have

$$|w_k \mathbb{E}(\hat{f}_{X,k}) - f_{X,k}|^2 = \sum_{k=1}^{\infty} (w_k - 1)^2 f_{X,k}^2 = \sum_{k=\lceil K \rceil+1}^{\infty} (w_k - 1)^2 f_{X,k}^2. \quad (\text{A.4})$$

Next we derive an upper bound for (A.4) in terms of $\|f_X^{(m)}\|^2$. To do this, for $k = 1, 2, \dots$ and $j = 0, \dots, m$, let $s_k = k + 1/2$ and $\rho_k^{[j]} = s_k^{1/2} \langle \varphi_k, f_X^{(j)} \rangle$. Below (A.8), we will prove that for $k \geq m$, we have

$$(\rho_k^{[0]})^2 \leq 2^{-m} \sum_{j=0}^m \binom{m}{j} (\rho_{k+2j-m}^{[m]})^2 \prod_{i=1}^m s_{k-i}^{-2}. \quad (\text{A.5})$$

Therefore, for $k \geq m$,

$$\begin{aligned} f_{X,k}^2 &= s_k^{-1} (\rho_k^{[0]})^2 \leq 2^{-m} s_k^{-1} \sum_{j=0}^m \binom{m}{j} (\rho_{k+2j-m}^{[m]})^2 \prod_{i=1}^m s_{k-i}^{-2} \\ &= 2^{-m} \sum_{j=0}^m (s_{k+2j-m}/s_k) \cdot \binom{m}{j} \langle \varphi_{k+2j-m}, f_X^{(m)} \rangle^2 \prod_{i=1}^m s_{k-i}^{-2} \\ &\leq 2^{-m} \sum_{j=0}^m (s_{k+m}/s_k) \cdot \binom{m}{j} \langle \varphi_{k+2j-m}, f_X^{(m)} \rangle^2 \prod_{i=1}^m s_{k-i}^{-2}. \end{aligned}$$

Using Parseval's identity and the binomial theorem, we deduce that

$$\sum_{k=\lceil K \rceil+1}^{\infty} f_{X,k}^2 \cdot (s_k/s_{k+m}) \cdot \prod_{i=1}^m s_{k-i}^{-2} \leq 2^{-m} \sum_{j=0}^m \binom{m}{j} \sum_{k=\lceil K \rceil+1}^{\infty} \langle \varphi_{k+2j-m}, f_X^{(m)} \rangle^2 \leq \|f_X^{(m)}\|^2,$$

so that

$$\sum_{k=\lceil K \rceil+1}^{\infty} f_{X,k}^2 \leq \|f_X^{(m)}\|^2 \sup_{k \geq \lceil K \rceil+1} (s_{k+m}/s_k) \cdot \prod_{i=1}^m s_{k-i}^{-2}.$$

Therefore, and since $f_X \in \mathcal{F}(c, m)$, the right hand side of (A.4) satisfies

$$\begin{aligned} \sum_{k=\lfloor K \rfloor+1}^{\infty} (w_k - 1)^2 f_{X,k}^2 &\leq \|f_X^{(m)}\|^2 \cdot \sup_{k \geq \lfloor K \rfloor+1} (w_k - 1)^2 \cdot \{1 + m/(k + 1/2)\} \cdot \prod_{i=1}^m (k + 1/2 - i)^{-2} \\ &\leq c \cdot \{1 + m/(\lfloor K \rfloor + 3/2)\} \cdot \prod_{i=1}^m (\lfloor K \rfloor + 3/2 - i)^{-2}. \end{aligned} \quad (\text{A.6})$$

For the variance term in (A.3), we need to bound $\sum_{k=1}^{\lfloor K \rfloor} w_k^2 \mathbb{V}\text{ar}(\hat{f}_{X,k})$. Now, letting

$$\mathcal{W}_{1,\ell} = (W_{1,1}^\ell, W_{1,1}^{\ell-1} W_{1,2}, \dots, W_{1,1}^{\lfloor \ell/2 \rfloor} W_{1,2}^{\lfloor \ell/2 \rfloor})^T$$

and using the Cauchy-Schwarz inequality and the bound

$$\left| \binom{(k + \ell - 1)/2}{k} \right| \leq 1, \quad \forall \ell = 1, \dots, k, \quad (\text{A.7})$$

we have

$$\begin{aligned} \mathbb{V}\text{ar}(\hat{f}_{X,k}) &= \frac{1}{n} 4^k (k + 1/2) \sum_{\ell, \ell'=1}^k \binom{k}{\ell} \binom{(k + \ell - 1)/2}{k} \binom{k}{\ell'} \binom{(k + \ell' - 1)/2}{k} \\ &\quad \cdot \text{cov}\{\mathbf{e}_{0,\ell} \{\mathcal{M}_U(\ell)\}^{-1} \mathcal{W}_{1,\ell}, \mathbf{e}_{0,\ell'} \{\mathcal{M}_U(\ell')\}^{-1} \mathcal{W}_{1,\ell'}\} \\ &\leq \frac{1}{n} 4^k (k + 1/2) \left[\sum_{\ell=1}^k \binom{k}{\ell} \cdot \{\mathbb{E}|\mathbf{e}_{0,\ell} \{\mathcal{M}_U(\ell)\}^{-1} \mathcal{W}_{1,\ell}|^2\}^{1/2} \right]^2, \end{aligned} \quad (\text{A.8})$$

where the sums start from $\ell, \ell' = 1$ since the terms for $\ell = 0$ and $\ell' = 0$ are deterministic. The proof of part (a) follows by combining (A.3), (A.6) and (A.8).

It remains to prove (A.5), note first that since all non scaled Legendre polynomials $\varphi_k(x) s_k^{-1/2}$ are equal to 1 at $x = \pm 1$, then by integration by parts similar to [38], we find

$$\begin{aligned} \rho_{k-1}^{[j+1]}/(2s_{k-1}) - \rho_{k+1}^{[j+1]}/(2s_{k+1}) &= \frac{1}{2} \int_{-1}^1 \{s_{k-1}^{-1/2} \varphi_{k-1}(x) - s_{k+1}^{-1/2} \varphi_{k+1}(x)\} f_X^{(j+1)}(x) dx \\ &= \frac{1}{2} \int_{-1}^1 \{\varphi'_{k+1}(x) s_{k+1}^{-1/2} - \varphi'_{k-1}(x) s_{k-1}^{-1/2}\} f_X^{(j)}(x) dx = \rho_k^{[j]}, \end{aligned}$$

where in the last step we used the fact that non scaled Legendre polynomials satisfy $2s_k^{1/2} \varphi_k(x) = s_{k+1}^{-1/2} \varphi'_{k+1}(x) - s_{k-1}^{-1/2} \varphi'_{k-1}(x)$, for all $x \in (-1, 1)$. Using the result $(a - b)^2 \leq 2a^2 - 2b^2$, it follows that

$$(\rho_k^{[j]})^2 \leq (\rho_{k-1}^{[j+1]})^2/(2s_{k-1}^2) + (\rho_{k+1}^{[j+1]})^2/(2s_{k+1}^2).$$

We complete the proof of (A.5) by induction. Specifically, for $\ell = 0$, for all $k \geq 0$ we have

$$2^{-\ell} \sum_{j=0}^{\ell} \binom{\ell}{j} (\rho_{k+2j-\ell}^{[\ell]})^2 \prod_{i=1}^{\ell} s_{k-i}^{-2} = (\rho_k^{[0]})^2$$

so that (A.5) holds for all $k \geq \ell = 0$. Now assume (A.5) holds for ℓ ; then, for $k \geq \ell + 1$, we have

$$\begin{aligned}
& 2^{-\ell} \sum_{j=0}^{\ell} \binom{\ell}{j} (\rho_{k+2j-\ell}^{[\ell]})^2 \prod_{i=1}^{\ell} s_{k-i}^{-2} \\
& \leq 2^{-\ell-1} \sum_{j=0}^{\ell} \binom{\ell}{j} (\rho_{k+2j-\ell-1}^{[\ell+1]})^2 s_{k+2j-\ell-1}^{-2} \prod_{i=1}^{\ell} s_{k-i}^{-2} + 2^{-\ell-1} \sum_{j=0}^{\ell} \binom{\ell}{j} (\rho_{k+2j-\ell+1}^{[\ell+1]})^2 s_{k+2j-\ell+1}^{-2} \prod_{i=1}^{\ell} s_{k-i}^{-2} \\
& = 2^{-\ell-1} \left\{ (\rho_{k-\ell-1}^{[\ell+1]})^2 s_{k-\ell-1}^{-2} + \sum_{j=1}^{\ell} \binom{\ell}{j} (\rho_{k+2j-\ell-1}^{[\ell+1]})^2 s_{k+2j-\ell-1}^{-2} \right\} \prod_{i=1}^{\ell} s_{k-i}^{-2} \\
& \quad + 2^{-\ell-1} \sum_{v=1}^{\ell+1} \binom{\ell}{v-1} (\rho_{k+2v-\ell-1}^{[\ell+1]})^2 s_{k+2v-\ell-1}^{-2} \prod_{i=1}^{\ell} s_{k-i}^{-2} \\
& = 2^{-\ell-1} \left\{ (\rho_{k-\ell-1}^{[\ell+1]})^2 s_{k-\ell-1}^{-2} + \sum_{j=1}^{\ell} \binom{\ell}{j} (\rho_{k+2j-\ell-1}^{[\ell+1]})^2 s_{k+2j-\ell-1}^{-2} \right\} \prod_{i=1}^{\ell} s_{k-i}^{-2} \\
& \quad + 2^{-\ell-1} \left\{ (\rho_{k+\ell+1}^{[\ell+1]})^2 s_{k+\ell+1}^{-2} + \sum_{v=1}^{\ell} \binom{\ell}{v-1} (\rho_{k+2v-\ell-1}^{[\ell+1]})^2 s_{k+2v-\ell-1}^{-2} \right\} \prod_{i=1}^{\ell} s_{k-i}^{-2} \\
& \leq 2^{-\ell-1} \left(\prod_{i=1}^{\ell+1} s_{k-i}^{-2} \right) \cdot \left\{ (\rho_{k-(\ell+1)}^{[\ell+1]})^2 + \sum_{j=1}^{\ell} \left(\binom{\ell}{j} + \binom{\ell}{j-1} \right) (\rho_{k+2j-(\ell+1)}^{[\ell+1]})^2 + (\rho_{k+\ell+1}^{[\ell+1]})^2 \right\} \\
& = 2^{-\ell-1} \left(\prod_{i=1}^{\ell+1} s_{k-i}^{-2} \right) \cdot \left\{ (\rho_{k-(\ell+1)}^{[\ell+1]})^2 + \sum_{j=1}^{\ell} \binom{\ell+1}{j} (\rho_{k+2j-(\ell+1)}^{[\ell+1]})^2 + (\rho_{k+\ell+1}^{[\ell+1]})^2 \right\} \\
& = 2^{-(\ell+1)} \sum_{j=0}^{\ell+1} \binom{\ell+1}{j} (\rho_{k+2j-(\ell+1)}^{[\ell+1]})^2 \prod_{i=1}^{\ell+1} s_{k-i}^{-2},
\end{aligned}$$

which proves that (A.5) holds for all $k \geq m$.

(b). In the particular case where $U \sim N(0, 1)$, applying Lemma 1 with $x = \mathcal{M}_U(\ell)^{-1} \mathcal{W}_{1,\ell}$, we have

$$|\mathbf{e}_{0,\ell} \{ \mathcal{M}_U(\ell) \}^{-1} \mathcal{W}_{1,\ell}|^2 \leq |\{ \mathcal{M}_U(\ell) \}^{-1} \mathcal{W}_{1,\ell}|^2 \leq C^* \cdot \ell \cdot 4^\ell \cdot |\mathcal{W}_{1,\ell}|^2.$$

Moreover,

$$\begin{aligned}
\mathbb{E} |\mathcal{W}_{1,\ell}|^2 &= \sum_{v=0}^{\lfloor \ell/2 \rfloor} \mathbb{E}(W_{1,1}^{2\ell-2v}) \cdot \mathbb{E}(W_{1,2}^{2v}) \leq (\lfloor \ell/2 \rfloor + 1) \cdot \mathbb{E}(W_{1,1}^{2\ell}) \leq (\lfloor \ell/2 \rfloor + 1) \cdot \mathbb{E}(1 + d|U|)^{2\ell} \\
&\leq (\lfloor \ell/2 \rfloor + 1) \cdot (1 + d)^{2\ell} \cdot (2\ell - 1)!!,
\end{aligned}$$

since $|X| \leq 1$ a.s., $\|\sigma\|_\infty \leq d$ and $\mu_{2\ell} = (2\ell - 1)!!$. We deduce (A.8) and the binomial theorem that

$$\mathbb{V}\text{ar}(\hat{f}_{X,k}) \leq \frac{1}{n} \cdot C^* \cdot 4^k \cdot k^2 (k + 1/2) \cdot \left\{ \sum_{\ell=0}^k \binom{k}{\ell} (2 + 2d)^\ell \right\}^2 \cdot (2k - 1)!!$$

$$\leq \frac{1}{n} \cdot C^* \cdot (6+4d)^{2k} \cdot k^2(k+1/2) \cdot (2k-1)!! . \quad (\text{A.9})$$

The proof of part (b) follows by combining (A.3), (A.6), (A.9) and (A.8). $\square$

Proof of Theorem 2. Note that

$$\{1 + m/(\lfloor K \rfloor + 3/2)\} \cdot \prod_{i=1}^m (\lfloor K \rfloor + 3/2 - i)^{-2} = O(K^{-2m}),$$

as $K \rightarrow \infty$ for any fixed $m \geq 1$. The variance term in Theorem 1 (b) has the following upper bound:

$$\begin{aligned} \frac{1}{n} \cdot C^* \sum_{k=1}^{\lfloor K \rfloor} k^2(k+1/2) \cdot (6+4d)^{2k} \cdot (2k-1)!! \\ \leq \frac{1}{n} \cdot D^K \cdot K! \asymp \frac{1}{n} \cdot \sqrt{2\pi K} \exp\{K(\log D + \log K - 1)\} = O(n^{c_0^*-1}), \end{aligned}$$

for some finite constant $D > 0$ sufficiently large, and all $c_0^* > c_0$, and where we used Stirling's approximation. This provides the desired convergence rate. $\square$

A.2. Proof of Theorem 3

Recalling that $\hat{\sigma}^2(x) = \hat{h}(x)/\max\{\hat{f}_X(x), \zeta/2\}$ and $\sigma^2 = h/f_X$, we write $\mathbb{E} \int_a^b |\hat{\sigma}^2(x) - \sigma^2(x)|^2 dx = I + II$, where

$$\begin{aligned} I &= \mathbb{E} \int_a^b |\hat{h}(x)f_X(x) - h(x)\hat{f}_X(x)|^2 \hat{f}_X^{-2}(x)f_X^{-2}(x) \cdot 1_{\{\hat{f}_X(x) \geq \zeta/2\}} dx, \\ II &= \mathbb{E} \int_a^b |2\hat{h}(x)/\zeta - \sigma^2(x)|^2 \cdot 1_{\{\hat{f}_X(x) < \zeta/2\}} dx. \end{aligned}$$

Recalling the definition of $\text{MISE}_n(f_X)$ and letting $\text{MISE}_n(h) = \mathbb{E} \int_{-1}^1 |\hat{h}(x) - h(x)|^2 dx$, if $\|\sigma\|_\infty \leq d$ and $f_X \in \mathcal{F}(c, m)$, we can bound I by

$$\begin{aligned} I &\leq 2 \mathbb{E} \int_a^b |\hat{h}(x)f_X(x) - h(x)f_X(x)|^2 \hat{f}_X^{-2}(x)f_X^{-2}(x) \cdot 1_{\{\hat{f}_X(x) \geq \zeta/2\}} dx \\ &\quad + 2 \mathbb{E} \int_a^b |h(x)f_X(x) - h(x)\hat{f}_X(x)|^2 \hat{f}_X^{-2}(x)f_X^{-2}(x) \cdot 1_{\{\hat{f}_X(x) \geq \zeta/2\}} dx \\ &\leq \frac{8}{\zeta^2} \mathbb{E} \int_a^b |\hat{h}(x) - h(x)|^2 dx + \frac{8}{\zeta^2} \sup_{x \in [a, b]} \sigma^4(x) \cdot \mathbb{E} \int_a^b |f_X(x) - \hat{f}_X(x)|^2 dx \\ &\leq \text{const.} \{\text{MISE}_n(h) + \text{MISE}_n(f_X)\}, \end{aligned}$$

where here and below, const. denotes a generic constant.

To bound II , recall that $f_X(x) > \zeta$ on $[a, b]$, so that for all $x \in [a, b]$, $\hat{f}_X(x) < \zeta/2$ implies that $|\hat{f}_X(x) - f_X(x)| > \zeta/2$. Therefore, we can write

$$II \leq \frac{8}{\zeta^2} \mathbb{E} \int_a^b |\hat{h}(x) - h(x)|^2 dx + 2 \int_a^b |2h(x)/\zeta - \sigma^2(x)|^2 \cdot \mathbb{P}\{|\hat{f}_X(x) - f_X(x)| > \zeta/2\} dx, \quad (\text{A.10})$$

where by Markov's inequality, the second term in (A.10) is bounded by

$$\frac{8}{\zeta^2} \left(\sup_{x \in [a, b]} \sigma^4(x) \cdot |2f_X(x)/\zeta - 1|^2 \right) \cdot \mathbb{E} \int_a^b |f_X(x) - \hat{f}_X(x)|^2 dx.$$

We deduce that if $\|\sigma\|_\infty \leq d$ and $f_X \in \mathcal{F}(c, m)$, then

$$II \leq \text{const.} \{ \text{MISE}_n(h) + \text{MISE}_n(f_X) \}.$$

Now under the conditions of Theorem 2, we have

$$\sup_{f_X \in \mathcal{F}(c, m), \|\sigma\|_\infty \leq d} \text{MISE}_n(f_X) = O\{(\log n / \log \log n)^{-2m}\},$$

and a similar result can be established for $\text{MISE}_n(h)$ when $U \sim N(0, 1)$. Specifically, following the proof of Theorem 1 but with the following modifications: replace the w_k 's by the $w_{h,k}$'s; start the sums from $k = 0$ and from $\ell, \ell' = 0$ (for the variance term); in the covariance in the line above (A.8), replace ℓ and ℓ' by $\ell + 2$ and $\ell' + 2$, respectively, and $\mathbf{e}_{0,\ell}$ and $\mathbf{e}_{0,\ell'}$ by $\mathbf{e}_{1,\ell+2}$ and $\mathbf{e}_{1,\ell'+2}$ (this does not have any impact on the bound), the following result can be established: for any given integer $m \geq 1$, if $\lceil K \rceil \geq m$ then for any finite positive constants c and d we have, for C^* is as in Lemma 1,

$$\begin{aligned} \sup_{f_X \sigma^2 \in \mathcal{F}(c, m), \|\sigma\|_\infty \leq d} \text{MISE}_n(h) &\leq c \cdot \{1 + m/(\lfloor K \rfloor + 3/2)\} \cdot \prod_{i=1}^m (\lfloor K \rfloor + 3/2 - i)^{-2} \\ &+ \frac{1}{n} \cdot C^* \cdot 16(1+d)^4 \sum_{k=0}^{\lfloor K \rfloor} w_{h,k}^2 \cdot (k+2)^2 (k+1/2) \cdot (6+4d)^{2k} \cdot (2k+3)!! . \end{aligned}$$

Using the proof of Theorem 2, we deduce that if $K = K_n = \lfloor c_0(\log n) / (\log \log n) \rfloor$ for some constant $c_0 \in (0, 1)$, then as $n \rightarrow \infty$,

$$\sup_{f_X \sigma^2 \in \mathcal{F}(c, m), \|\sigma\|_\infty \leq d} \text{MISE}_n(h) = O\{(\log n / \log \log n)^{-2m}\}.$$

Together with the above bounds on I and II , this implies that

$$\sup_{f_X, f_X \sigma^2 \in \mathcal{F}(c, m), \|\sigma\|_\infty \leq d} \mathbb{E} \int_a^b |\hat{\sigma}^2(x) - \sigma^2(x)|^2 dx = O\{(\log n / \log \log n)^{-2m}\}.$$

A.3. Optimality of rate given in Theorem 2

In this section we prove that the rate given in Theorem 2 is optimal. For this, it suffices to prove that, in the model at (2.1) with $U_{i,j} \sim N(0, 1)$, for a subset of functions $f_X \in \mathcal{F}(c, m)$ and for some functions σ such that $\|\sigma\|_\infty \leq d$, where c and d are positive constants, no nonparametric estimator constructed from the $W_{i,j}$'s converges at a rate faster than $O\{(\log n / \log \log n)^{-2m}\}$. Since the lower bound for estimators in the case where σ is known is smaller or equal to that in the case where σ is unknown, it suffices to prove this in the case where σ is known. With this in mind, consider the particular case where σ is equal to a known finite constant σ_ε , so that the $W_{i,j}$'s satisfy the homoscedastic model at

(1.1) with $J = 2$ and $\varepsilon_{i,j} \sim N(0, \sigma_\varepsilon^2)$, which is a particular case of the model considered by [16]. In [16], it was proved that in that setting, the sample of averages

$$(W_{i,1} + W_{i,2})/2 = X_i + (\varepsilon_{i,1} + \varepsilon_{i,2})/2, \quad i = 1, \dots, n$$

is a sufficient statistic for f_X . Therefore, lower bounds of the convergence rate of nonparametric estimators in model (1.1) with $J = 2$ and $\varepsilon_{i,j} \sim N(0, \sigma_\varepsilon^2)$, over a subset of functions $f_X \in \mathcal{F}(c, m)$, are equal to those in the model where we observe $Y_1, \dots, Y_n$ such that

$$Y_i = X_i + V_i, \quad V_i \sim N(0, \sigma_V^2/2). \quad (\text{A.11})$$

Such a lower bound was established in Theorem 2 of [28]. Specifically, Theorem 2 of [28] with $\delta = 0$ and $S_n = 1$ (i.e., taking in that theorem the support of f_X to be included in $[-1, 1]$) establishes the existence of a subset of $\mathcal{F}(c, m)$, with $m \geq 1$, such that no estimator of f_X in that subset can converge at a rate faster than $(\log n / \log \log n)^{-2m}$, which proves that the rate in our Theorem 2 is optimal.

Note that Theorem 2 of [28] was established for densities f_V that do not include normal densities, but an inspection of the proof in that paper shows that the same lower bound holds in the case where V_i is normal. The only change required in that proof is the proof of (28) in [28], which in our case should be established for $\tilde{g} = f_V$, the density of a $N(0, \sigma_V^2)$ for $\sigma_V^2 = \sigma_\varepsilon^2/2$. That is, in our case we should prove that

$$\int |(f_X - f_n) * f_V(x)| dx = O(n^{-1}), \quad (\text{A.12})$$

where $f_n(x) = f_X(x) + \alpha_n(Q_{\mathcal{K}_n} * \varphi_{b_n})(x)$, and $Q_{\mathcal{K}_n}$ and $\varphi_{b_n}(x) = b_n \varphi(b_n x)$ are functions defined in [28] (taking there $S_n = 1$ and $\delta = 0$) which satisfy

$$|(Q_{\mathcal{K}_n}^{f^t})^{(j)}(t)| \leq O(1) |te/(\mathcal{K}_n - j)|^{\mathcal{K}_n - j} \quad \text{for } j = 0, 1, \quad (\text{A.13})$$

$|\varphi^{f^t}(t/b_n)| \leq 1$ and $|(\varphi^{f^t})'(t)| \leq 1$ for all t , for sequences $\mathcal{K}_n \rightarrow \infty$, $\alpha_n \rightarrow 0$ and $b_n \rightarrow \infty$ as $n \rightarrow \infty$.

To prove this, following (29) of [28], but taking there $\tilde{g} = f_V$, we have

$$\begin{aligned} & \int |(f_X - f_n) * f_V(x)| dx \\ &= O(1) \times \left\{ \sum_{(j,k)=(0,0),(1,0),(0,1)} \int t^{2k} |\{(f_X - f_n)^{f^t}(t)\}^{(j)}|^2 \exp(-\sigma_V^2 t^2) dt \right\}^{1/2} \\ &= O(\alpha_n) \left\{ \sum_{(j,k)=(0,0),(1,0),(0,1)} \int |(Q_{\mathcal{K}_n}^{f^t})^{(j)}(t)|^2 t^{2k} \exp(-\sigma_V^2 t^2) dt \right\}^{1/2}. \end{aligned}$$

Now for $(j, k) = (0, 0), (1, 0)$ and $(0, 1)$, using (A.13), we have

$$\begin{aligned} & \int |(Q_{\mathcal{K}_n}^{f^t})^{(j)}(t)|^2 t^{2k} \exp(-\sigma_V^2 t^2) dt = O \left\{ \left(\frac{e}{\mathcal{K}_n - j} \right)^{2(\mathcal{K}_n - j)} \int t^{2(\mathcal{K}_n - j + k)} \exp(-\sigma_V^2 t^2) dt \right\} \\ &= O \left[\{2(\mathcal{K}_n - j + k) - 1\}!! \left\{ \frac{e}{\sqrt{2}\sigma_V(\mathcal{K}_n - j)} \right\}^{2(\mathcal{K}_n - j)} \right] \\ &= O \left(\left\{ \frac{2(\mathcal{K}_n - j + k) - 1}{e} \right\}^{\mathcal{K}_n - j + k} \left\{ \frac{e}{\sqrt{2}\sigma_V(\mathcal{K}_n - j)} \right\}^{2(\mathcal{K}_n - j)} \right), \quad (\text{A.14}) \end{aligned}$$

where we used the fact that for n large enough, we have

$$\left(2(\mathcal{K}_n - j + k) - 1\right)!! \leq O\left(\left\{\frac{2(\mathcal{K}_n - j + k) - 1}{e}\right\}^{\mathcal{K}_n - j + k}\right).$$

Since, for n large enough, (A.14) takes its maximum value when $(j, k) = (1, 0)$, we deduce that for $(j, k) = (0, 0), (1, 0)$ and $(0, 1)$, as $n \rightarrow \infty$,

$$(A.14) = O\left[\{d_1(\mathcal{K}_n - 1)\}^{1 - \mathcal{K}_n}\right],$$

where $d_1 = \sigma_V^2 / e$. Recalling that $\alpha_n = o(1)$, we deduce that

$$\int |(f_X - f_n) * f_V(x)| dx = O\left[\alpha_n \{d_1(\mathcal{K}_n - 1)\}^{(1 - \mathcal{K}_n)/2}\right] = o\left[\{d_1(\mathcal{K}_n - 1)\}^{(1 - \mathcal{K}_n)/2}\right].$$

Taking $\mathcal{K}_n = 1 + c \log n / \log \log n$ for some constant $c > 0$ as in equation (32) from [28], for n large enough we have

$$\begin{aligned} \{d_1(\mathcal{K}_n - 1)\}^{(1 - \mathcal{K}_n)/2} &= \exp\left[\frac{(1 - \mathcal{K}_n)}{2} \log\{d_1(\mathcal{K}_n - 1)\}\right] \\ &= \exp\left\{-\frac{c \log n}{2 \log \log n} \log\left(\frac{d_1 c \log n}{\log \log n}\right)\right\} \\ &\leq \exp\left(-\frac{c \log n}{2 \log \log n} \times \frac{\log \log n}{2}\right) = \exp(\log n^{-c/4}) = n^{-c/4} = O(n^{-1}) \end{aligned}$$

if $c \geq 4$. This proves (A.12) and the rest of the proof of Theorem 2 in [28] can be continued as in [28].

A.4. Small error setting

In this section we discuss the behaviour of the $\text{MISE}_n(f_X)$ of our estimator studied in Theorem 1 in the case where $\|\sigma^2\|_\infty \rightarrow 0$ as $n \rightarrow \infty$, and compare it with the $\text{MISE}_n(f_X)$ in the error-free case where $\sigma \equiv 0$. We assume throughout this section that the conditions of Theorem 1 hold.

Since the bias of the estimator is not affected by the error, we can bound the bias term of $\text{MISE}_n(f_X)$ by (A.6). However, in the case where $\|\sigma^2\|_\infty \rightarrow 0$ as $n \rightarrow \infty$, we can provide a refined version of the upper bound to the variance term at (A.8). Specifically, letting

$$\begin{aligned} C(\ell, \ell') &= \sum_{r=1}^{\lfloor (\ell + \ell')/2 \rfloor} \mathbb{E}\{X^{\ell + \ell' - 2r} \sigma^{2r}(X)\} \\ &\quad \times \sum_{q=0}^{\lfloor \ell/2 \rfloor} \sum_{q'=0}^{\lfloor \ell'/2 \rfloor} \{\mathcal{M}_U(\ell)\}_{1, q+1}^{-1} \{\mathcal{M}_U(\ell')\}_{1, q'+1}^{-1} \cdot \{\mathcal{M}_U(\ell + \ell')\}_{q+q'+1, r+1}, \end{aligned} \quad (A.15)$$

below we show that we have

$$\mathbb{V}\text{ar}(\hat{f}_{X, k}) = \frac{1}{n} \cdot \mathbb{V}\text{ar}\{\varphi_k(X)\} + \frac{1}{n} 4^k \left(k + \frac{1}{2}\right) \sum_{\ell, \ell'=1}^k \binom{k}{\ell} \binom{(k + \ell - 1)/2}{k} \binom{k}{\ell'} \binom{(k + \ell' - 1)/2}{k} C(\ell, \ell'). \quad (A.16)$$

While the first term of the above equation is of order n^{-1} , the second term is very difficult to analyse. In the error-free case where $\sigma \equiv 0$, $C(\ell, \ell') = 0$, so that $\text{Var}(\hat{f}_{X,k}) = O(n^{-1})$. Combining this with (A.3), (A.4) and (A.6), and recalling that $\|\varphi_k(X)\| = 1$, we deduce that in the error-free case, we have

$$\sup_{f_X \in \mathcal{F}(c,m)} \text{MISE}_n(f_X) \leq c \cdot \{1 + m/(\lfloor K \rfloor + 3/2)\} \cdot \prod_{i=1}^m (\lfloor K \rfloor + 3/2 - i)^{-2} + \sum_{k=1}^{\lfloor K \rfloor} w_k^2 \frac{1}{n} \cdot \|f_X\|_\infty.$$

Taking $K \asymp n^{1/(2m+1)}$, we deduce that when $\sigma \equiv 0$, $\sup_{f_X \in \mathcal{F}(c,m)} \text{MISE}_n(f_X) = O(n^{-2m/(2m+1)})$, which is the fastest possible convergence rate of the $\text{MISE}_n(f_X)$ towards zero in the error-free case.

When $\|\sigma^2\|_\infty > 0$, $C(\ell, \ell')$ does not vanish but if $\|\sigma^2\|_\infty \rightarrow 0$ sufficiently fast as $n \rightarrow \infty$, then $C(\ell, \ell')$ is so small that the second term of (A.16) is negligible compared to the first. In that case, if $K \asymp n^{1/(2m+1)}$, we have $\sup_{f_X \in \mathcal{F}(c,m)} \text{MISE}_n(f_X) = O(n^{-2m/(2m+1)})$ as in the error-free case. However, it would be technically very challenging to establish the rate at which $\|\sigma^2\|_\infty$ should tend to zero for this error-free rate to be achieved. This is because the second term of (A.16) is too complex to be analysed precisely, given the quadruple sum (in ℓ, ℓ', q and q') and the inverse of the matrices $\mathcal{M}_U(\ell)$ and $\mathcal{M}_U(\ell')$ within those sums. For the same reason, when $\|\sigma^2\|_\infty \rightarrow 0$ but the second term of (A.16) is not negligible compared to the first, determining the rate at which $\text{MISE}_n(f_X) \rightarrow 0$ is a technically hard problem which we leave open for future research. However, we conjecture that polynomial rates of convergence can only be achieved if $\|\sigma^2\|_\infty \rightarrow 0$ exponentially fast as $n \rightarrow \infty$.

Proof of (A.16). Recall from the equation above (A.8) that

$$\begin{aligned} \text{Var}(\hat{f}_{X,k}) &= \frac{1}{n} 4^k (k+1/2) \sum_{\ell, \ell'=1}^k \binom{k}{\ell} \binom{(k+\ell-1)/2}{k} \binom{k}{\ell'} \binom{(k+\ell'-1)/2}{k} \\ &\quad \cdot \text{cov}\{\mathbf{e}_{0,\ell} \{\mathcal{M}_U(\ell)\}^{-1} \mathcal{W}_{1,\ell}, \mathbf{e}_{0,\ell'} \{\mathcal{M}_U(\ell')\}^{-1} \mathcal{W}_{1,\ell'}\}. \end{aligned}$$

Now we have

$$\begin{aligned} &\text{cov}\{\mathbf{e}_{0,\ell} \{\mathcal{M}_U(\ell)\}^{-1} \mathcal{W}_{1,\ell}, \mathbf{e}_{0,\ell'} \{\mathcal{M}_U(\ell')\}^{-1} \mathcal{W}_{1,\ell'}\} \\ &= \sum_{q=0}^{\lfloor \ell/2 \rfloor} \sum_{q'=0}^{\lfloor \ell'/2 \rfloor} \{\mathcal{M}_U(\ell)\}_{1,q+1}^{-1} \{\mathcal{M}_U(\ell')\}_{1,q'+1}^{-1} \cdot \text{cov}(W_{1,1}^{\ell-q} W_{1,2}^q, W_{1,1}^{\ell'-q'} W_{1,2}^{q'}) \\ &= \sum_{q=0}^{\lfloor \ell/2 \rfloor} \sum_{q'=0}^{\lfloor \ell'/2 \rfloor} \{\mathcal{M}_U(\ell)\}_{1,q+1}^{-1} \{\mathcal{M}_U(\ell')\}_{1,q'+1}^{-1} \cdot \{\mathbb{E}(W_{1,1}^{\ell+\ell'-(q+q')} W_{1,2}^{q+q'}) \\ &\quad - (\mathbb{E} W_{1,1}^{\ell-q} W_{1,2}^q) \cdot (\mathbb{E} W_{1,1}^{\ell'-q'} W_{1,2}^{q'})\}. \end{aligned} \tag{A.17}$$

Since

$$\mathbb{E}(W_{1,1}^{\ell-q} W_{1,2}^q) = \{\mathcal{M}_W(\ell)\}_{q+1} = \{\mathcal{M}_U(\ell) \mathcal{M}_{X,\sigma}(\ell)\}_{q+1},$$

then (A.17) is equal to

$$\sum_{q=0}^{\lfloor \ell/2 \rfloor} \sum_{q'=0}^{\lfloor \ell'/2 \rfloor} \{\mathcal{M}_U(\ell)\}_{1,q+1}^{-1} \{\mathcal{M}_U(\ell')\}_{1,q'+1}^{-1} \cdot \{\mathcal{M}_U(\ell+\ell') \mathcal{M}_{X,\sigma}(\ell+\ell')\}_{q+q'+1}$$

$$\begin{aligned}
& - \sum_{q=0}^{\lfloor \ell/2 \rfloor} \sum_{q'=0}^{\lfloor \ell'/2 \rfloor} \{\mathcal{M}_U(\ell)\}_{1,q+1}^{-1} \{\mathcal{M}_U(\ell')\}_{1,q'+1}^{-1} \cdot \{\mathcal{M}_U(\ell)\mathcal{M}_{X,\sigma}(\ell)\}_{q+1} \{\mathcal{M}_U(\ell')\mathcal{M}_{X,\sigma}(\ell')\}_{q'+1} \\
& = \sum_{r=0}^{\lfloor (\ell+\ell')/2 \rfloor} \mathbb{E}\{X^{\ell+\ell'-2r}\sigma^{2r}(X)\} \cdot \sum_{q=0}^{\lfloor \ell/2 \rfloor} \sum_{q'=0}^{\lfloor \ell'/2 \rfloor} \left[\{\mathcal{M}_U(\ell)\}_{1,q+1}^{-1} \{\mathcal{M}_U(\ell')\}_{1,q'+1}^{-1} \right. \\
& \quad \left. \times \{\mathcal{M}_U(\ell+\ell')\}_{q+q'+1,r+1} \right] - (\mathbb{E}X^\ell)(\mathbb{E}X^{\ell'}). \tag{A.18}
\end{aligned}$$

Now according to (3.6) we have $\{\mathcal{M}_U(\ell+\ell')\}_{q+q'+1,1} = 1$ for all q, q' . This also implies (when $\ell+\ell'$ is replaced by ℓ) that

$$\mathcal{M}_U(\ell)(1, 0, \dots, 0)^T = (1, 1, \dots, 1)^T.$$

Combing this with the invertibility of $\mathcal{M}_U(\ell)$, we deduce that $\sum_{q=0}^{\lfloor \ell/2 \rfloor} \{\mathcal{M}_U(\ell)\}_{1,q+1}^{-1} = 1$, so that

$$\sum_{q=0}^{\lfloor \ell/2 \rfloor} \sum_{q'=0}^{\lfloor \ell'/2 \rfloor} \{\mathcal{M}_U(\ell)\}_{1,q+1}^{-1} \{\mathcal{M}_U(\ell')\}_{1,q'+1}^{-1} \cdot \{\mathcal{M}_U(\ell+\ell')\}_{q+q'+1,1} = 1.$$

Therefore, using the notation at (A.15), (A.18) can be written as $\text{cov}(X^\ell, X^{\ell'}) + C(\ell, \ell')$.

Recalling the definition of φ_k above (3.1), we deduce that

$$\begin{aligned}
\mathbb{V}\text{ar}(\hat{f}_{X,k}) &= \frac{1}{n} 4^k (k+1/2) \sum_{\ell, \ell'=1}^k \binom{k}{\ell} \binom{(k+\ell-1)/2}{k} \binom{k}{\ell'} \binom{(k+\ell'-1)/2}{k} \cdot \{\text{cov}(X^\ell, X^{\ell'}) + C(\ell, \ell')\} \\
&= \frac{1}{n} \cdot \mathbb{V}\text{ar}\{\varphi_k(X)\} + \frac{1}{n} 4^k \left(k + \frac{1}{2}\right) \sum_{\ell, \ell'=1}^k \binom{k}{\ell} \binom{(k+\ell-1)/2}{k} \binom{k}{\ell'} \binom{(k+\ell'-1)/2}{k} C(\ell, \ell').
\end{aligned}$$

□

Appendix B: Case with a data-driven upper bound to $\mathcal{S}_X$

B.1. Crude upper bound

Here we describe how to construct a rough upper bound $[A_{X,1}, B_{X,1}]$ to $\mathcal{S}_X$. Although $\mathcal{S}_X \subseteq \mathcal{S}_W$, $A_{X,1} = \min_{1 \leq i \leq n, j=1,2} W_{i,j}$ and $B_{X,1} = \max_{1 \leq i \leq n, j=1,2} W_{i,j}$ are generally too crude to be useful; for example, if $\mathcal{S}_U = \mathbb{R}$ (e.g., if $U \sim N(0, 1)$), then $|A_{X,1}|$ and $B_{X,1} \rightarrow \infty$ as $n \rightarrow \infty$. On the other hand, for $\alpha > 0$ sufficiently small, we have $\mathcal{S}_X \subseteq [q_{W,\alpha}, q_{W,1-\alpha}]$, where $q_{W,\alpha}$ denotes the α quantile of W . This motivates us to take $A_{X,1} = \hat{q}_{W,0.001} - \Delta_L$ and $B_{X,1} = \hat{q}_{W,0.999} + \Delta_U$, with $\hat{q}_{W,\alpha}$ the empirical α quantile of the $W_{i,j}$'s, $\Delta_L = \hat{\mu}_W - \hat{q}_{W,0.001}$, $\Delta_U = \hat{q}_{W,0.999} - \hat{\mu}_W$, and $\hat{\mu}_W = n^{-1} \sum_{i=1}^n \bar{W}_i$. Here, we use an expanded version of the interval $[\hat{q}_{W,0.001}, \hat{q}_{W,0.999}]$ for cases where f_W is skewed to one side, applying an asymmetric correction Δ_L or Δ_U , which is larger on the skewed side. Except perhaps in cases of marginal practical interest, $[A_{X,1}, B_{X,1}]$ constructed in that way usually contains $\mathcal{S}_X$.

B.2. Refined upper bound

Next we describe a procedure for refining the upper bound $[A_{X,1}, B_{X,1}]$ to $\mathcal{S}_X$ from Appendix B.1. For this, let $W'_{i,j}$, X'_i , μ and λ as in (6.1) and (6.2), taking there $A_X = A_{X,1}$ and $B_X = B_{X,1}$. If $\mathcal{S}_X \subseteq [A_{X,1}, B_{X,1}]$ (see Remark 4 below otherwise), then $\mathcal{S}_{X'} \subseteq [-1, 1]$ and we can compute the Legendre polynomial estimator of $f_{X'}$ at (6.3). Since $f_{X'} = 0$ on $[-1, 1] \setminus \mathcal{S}_{X'}$, then on that range, $\hat{f}_{X'}(\cdot; K)$ estimates the constant value 0, and unless $[A_{X,1}, B_{X,1}]$ is a relatively tight upper bound to $\mathcal{S}_X$, $\hat{f}_{X'}(\cdot; K)$ typically oscillates around zero there. This is particularly true if K is larger than the optimal K for estimating $f_{X'}$ by $\hat{f}_{X'}(\cdot; K)$ because then $\hat{f}_{X'}(\cdot; K)$ is undersmoothed and the oscillations are more pronounced. We exploit these oscillations to refine $[A_{X,1}, B_{X,1}]$. To choose such a large K , we take

$$\hat{K}_1(f_{\overline{W}'}) = \operatorname{argmin}_{K \in I_K} \sum_{\alpha \in \{0.001, 0.002, 0.998, 0.999\}} \{\hat{F}_{\overline{W}', \text{Leg}}^{-1}(\alpha) - \hat{q}_{\overline{W}', \alpha}\}^2, \quad (\text{B.1})$$

where $\hat{q}_{\overline{W}', \alpha}$ is the (interpolated) empirical α quantile of the $\overline{W}'_i$'s and $\hat{F}_{\overline{W}', \text{Leg}}(x) = \int_{-\infty}^x \hat{f}_{\overline{W}', \text{Leg}}(\cdot; K)$, with $\hat{f}_{\overline{W}', \text{Leg}}$ defined like $\hat{f}_{\overline{W}, \text{Leg}}$ in (5.2) but replacing there $\overline{W}$, X and $\hat{\sigma}_{\text{pilot}}$ by $\overline{W}'$, X' and $\hat{\sigma}'_{\text{pilot}}$ from Section C.1.

We define an undersmoothed pilot estimator of $f_{X'}$ by $\hat{f}_{X',p} \equiv \hat{f}_{X'}\{\cdot; \hat{K}_1(f_{\overline{W}'})\}$. Similar to $\hat{K}(f_{\overline{W}'})$ in (5.3), $\hat{K}_1$ is typically too large for estimating $f_{X'}$ because it targets $f_{\overline{W}'}$. Here we use $\hat{K}_1$ instead of $\hat{K}(f_{\overline{W}'})$ because (B.1) focuses on the tails of the estimator of $f_{W'}$, which we found works better when the focus is on the boundary, like here.

As noted above, if $[A_{X,1}, B_{X,1}]$ is a crude upper bound to $\mathcal{S}_X$, we can expect $\hat{f}_{X',p}$ to oscillate around zero on $[-1, 1] \setminus \mathcal{S}_{X'}$. We detect those oscillations near the boundary as follows:

1. Find $x_{\max} = \operatorname{argmax}_{x \in [-1, 1]} \hat{f}_{X',p}(x)$. If $x_{\max}$ is not unique, let $x_{\max_1}$ (resp., $x_{\max_2}$) be the smallest (resp., largest) of them; otherwise, let $x_{\max_1} = x_{\max_2} = x_{\max}$. If it exists, let x_L be the largest $-1 \leq x < x_{\max_1}$ (resp., x_R the smallest $x_{\max_2} < x \leq 1$) such that $\hat{f}_{X',p}(x) = 0$; otherwise, take $x_L = -1$ (resp., $x_R = 1$).

If f_X is unimodal, then typically $\mathcal{S}_{X'} \subseteq [x_L, x_R]$. If f_X is multimodal with well separated modes, $\hat{f}_{X',p}$ may oscillate in an area between the modes, and x_L and/or x_R may correspond to that area (whereas our goal is to detect oscillations near the boundaries of $\mathcal{S}_{X'}$). In that case, often $\int_{x_L}^{x_R} \max\{\hat{f}_{X',p}(x), 0\} dx$ is much smaller than $\int_{-1}^1 \max\{\hat{f}_{X',p}(x), 0\} dx$, which can be exploited to discard the area between two modes, as follows:

2. Compute $r = \int_{x_L}^{x_R} \max\{\hat{f}_{X',p}(x), 0\} dx / \int_{-1}^1 \max\{\hat{f}_{X',p}(x), 0\} dx$.
3. If $r \geq 0.7$, do not change x_L and x_R . If $r < 0.7$, we are likely to be in a multimodal situation with mode(s) to the left of x_L or to the right of x_R . If $f_{X'}$ has a mode to the left of x_L (resp., to the right of x_R), then $[-1, x_L]$ (resp., $[x_R, 1]$) will contain a significant part of $\mathcal{S}_{X'}$ and $I_L^+ = \int_{-1}^{x_L} \max\{\hat{f}_{X',p}(x), 0\} dx$ (resp., $I_R^+ = \int_{x_R}^1 \max\{\hat{f}_{X',p}(x), 0\} dx$) will typically be quite large. We identify a potential mode of $f_{X'}$ to the left of x_L or the right of x_R , depending on which of I_L^+ and I_R^+ is the largest, and recompute x_L or x_R accordingly. Namely:
 - If $I_L^+ > I_R^+$, compute $x_{\max,L} = \operatorname{argmax}_{x \in [-1, x_L]} \hat{f}_{X',p}(x)$ (if $x_{\max,L}$ is not unique, take the largest one). Replace x_L by be the largest $-1 \leq x < x_{\max,L}$ such that $\hat{f}_{X',p}(x) = 0$ if it exists; otherwise take $x_L = -1$.
 - If $I_L^+ \leq I_R^+$, compute $x_{\max,R} = \operatorname{argmax}_{x \in (x_R, 1]} \hat{f}_{X',p}(x)$ (if $x_{\max,R}$ is not unique, take the largest one). Replace x_R by the smallest $x_{\max,R} < x \leq 1$ such that $\hat{f}_{X',p}(x) = 0$ if it exists; otherwise take $x_R = 1$.

- Once x_L or x_R has been updated as above, return to step 2.

Finally, we take our upper bound to $\mathcal{S}_{X'}$ to be $[A_{X'}, B_{X'}] = [x_L, x_R]$. Recalling the notation μ and λ from Section 6.3, our final upper bound to $\mathcal{S}_X$ is $[A_{X,2}, B_{X,2}] = [\mu + \lambda A_{X'}, \mu + \lambda B_{X'}]$.

Remark 4. In unusual cases where $\mathcal{S}_{X'} \not\subseteq [-1, 1]$, $A_{X'}$ and $B_{X'}$ are usually close to -1 and 1 . There, while (6.3) is not a consistent estimator of $f_{X'}$, it can usually provide a reasonable approximation.

B.3. Choosing K in practice

Here we show how to extend the method from Section 5 to the case where we compute an upper bound $[A_X, B_X]$ to $\mathcal{S}_X$ from the data, and compute the estimator $\hat{f}_{X'}(\cdot, \hat{K})$ at (6.3) from the rescaled data $W'_{i,j}$ defined in (6.1), with $\lambda = (B_X - A_X)/2$ and $\mu = (B_X + A_X)/2$. The main differences between this case and the one in Section 5 are that (1) we need to rescale the data at each step of the procedure, and (2) to mimic the process of the real world, in the SIMEX world we also need to estimate the support of the SIMEX data. We proceed as follows.

First we compute $\hat{K}(f_{\bar{W}'})$ for estimating the density $f_{\bar{W}'}$ of $\bar{W}'_i = (W'_{i,1} + W'_{i,2})/2 = X'_i + \sigma'(X'_i)\bar{U}_i$, where $\bar{U}_i \sim f_{\bar{U}} = f_U/2 * f_U/2$ and σ' as in (6.1). We take $\hat{K}(f_{\bar{W}'})$ as at (5.3), replacing there and in (5.2), $\bar{W}$, X and $\hat{\sigma}_{\text{pilot}}$ by $\bar{W}'$, X' and $\hat{\sigma}'_{\text{pilot}}$ defined in Section C.1. Next, to construct an approximation $\hat{C}_K$ of $C_K = K_0(f_{X'})/\hat{K}(f_{\bar{W}'})$, we mimic $K_0(f_{X'})/\hat{K}(f_{\bar{W}'})$ using artificial data $(X'^{*}_i, W'^{*}_{i,1}, W'^{*}_{i,2})$ created from the $W'_{i,j}$'s in a way that mimics the $(X'_i, W'_{i,1}, W'_{i,2})$'s as follows:

1. For $i = 1, \dots, n_1$, mimic the X'_i 's by taking X'^{*}_i defined in Appendix C.2, with $n_1 \leq n$ defined there.
2. For $i = 1, \dots, n_1$, and $j = 1, 2$, generate i.i.d. $U^*_{i,j} \sim f_U$ and mimic the $W_{i,j}$'s by taking $W'^{*}_{i,j} = X'^{*}_i + \sigma^*(X'^{*}_i)U^*_{i,j}$, where $\sigma^* = \hat{\sigma}'_{\text{pilot}} \equiv \lambda \hat{\sigma}'_{\text{pilot}} \{(x - \mu)/\lambda\}$.
3. Compute an upper bound $[A_{X^*}, B_{X^*}]$ to $\mathcal{S}_{X^*}$ using the approach from Section 6.3, replacing there the $W_{i,j}$'s by the $W'^{*}_{i,j}$'s. Let $\lambda^* = (B_{X^*} - A_{X^*})/2$ and $\mu^* = (A_{X^*} + B_{X^*})/2$, and for $i = 1, \dots, n_1$, and $j = 1, 2$, mimic the X'_i 's and $W'_{i,j}$'s by taking $X'^{*}_i = (X'^{*}_i - \mu^*)/\lambda^*$ and $W'^{*}_{i,j} = (W'^{*}_{i,j} - \mu^*)/\lambda^*$.

As in Section 5, we generate B such samples $W'^{*}_{i,j,b}$'s, $b = 1, \dots, B$ by repeating steps 2 and 3 B times. In step 3, this requires recomputing an upper bound $[A_{X^*,b}, B_{X^*,b}]$ to $\mathcal{S}_{X^*}$ for each b . Instead, to speed up computations, we use the same bound $[A_{X^*}, B_{X^*}]$ for all b , obtained as follows: we compute $A_{X^*,b}$ and $B_{X^*,b}$ for $b = 1, 2, 3$ and take $A_{X^*} = \bar{A}_{X^*,b_1,b_2} \equiv \sum_{b \in \{b_1,b_2\}} A_{X^*,b}/2$ and $B_{X^*} = \bar{B}_{X^*,b_1,b_2} \equiv \sum_{b \in \{b_1,b_2\}} B_{X^*,b}/2$, where $1 \leq b_1 < b_2 \leq 3$ minimise $\sum_{b \in \{b_1,b_2\}} \{(A_{X^*,b} - \bar{A}_{X^*,b_1,b_2})^2 + (B_{X^*,b} - \bar{B}_{X^*,b_1,b_2})^2\}$. Thus we average the two most similar bounds and discard the most dissimilar one.

Finally, we take $\hat{C}_K = \hat{K}_0(f_{X'^{*}})/\hat{K}(f_{\bar{W}'^*})$, where $\hat{K}_0(f_{X'^{*}})$ and $\hat{K}(f_{\bar{W}'^*})$ are defined as in (5.7) and (5.8) but with X , $\bar{W}$ and $\bar{W}^*$ replaced by X' , $\bar{W}'$ and $\bar{W}'^*$. We approximate $K_0(f_{X'})$ by

$$\hat{K} = \hat{C}_K \cdot \hat{K}(f_{\bar{W}'}). \quad (\text{B.2})$$

B.4. Truncating the estimator

As described in Section 6.3, after computing the estimator $\tilde{f}_X(\cdot, \hat{K})$ on $[A_X, B_X] = [A_{X,2}, B_{X,2}]$, we check if it has large spurious oscillations near A_X and B_X . If it does, we truncate it to an interval $[A_X^\dagger, B_X^\dagger] \subset [A_X, B_X]$, chosen by searching for unusually large values and oscillations near A_X and B_X , as follows. Let $\Delta x = (B_X - A_X)/200$, a small value compared to $B_X - A_X$. Then:

1. Let $M \equiv \max_{x \in [A_X, B_X]} \tilde{f}_X(x)$, and let $m_0 = (A_X + B_X)/2$, the centre of $[A_X, B_X]$.
2. If $\tilde{f}_X(A_X) > \max\{M/20, \tilde{f}_X(A_X + \Delta x)\}$, $\tilde{f}_X$ is unexpectedly large and decreases near A_X , so we look for the oscillation closest to A_X by finding the smallest $m_L \in (A_X, m_0]$ such that $\tilde{f}(m_L) \leq \min\{\tilde{f}_X(m_L - \Delta x), \tilde{f}_X(m_L + \Delta x)\}$. If no such m_L exists or $\tilde{f}_X(A_X) \leq \max\{M/20, \tilde{f}_X(A_X + \Delta x)\}$, we let $m_L = A_X$.
3. Proceed similarly near the right boundary: if $\tilde{f}_X(B_X) > \max\{M/20, \tilde{f}_X(B_X - \Delta x)\}$, search for the largest $m_R \in [m_0, B_X]$ such that $\tilde{f}(m_R) \leq \min\{\tilde{f}_X(m_R - \Delta x), \tilde{f}_X(m_R + \Delta x)\}$. If no such m_R exists or $\tilde{f}_X(B_X) \leq \max\{M/20, \tilde{f}_X(B_X - \Delta x)\}$, let $m_R = B_X$.

This produces a shrunk version $[m_L, m_R]$ of $[A_X, B_X]$ which avoids the most obvious spurious oscillations nearest to A_X and B_X ; we further shrink $[m_L, m_R]$ to $[A_X^\dagger, B_X^\dagger]$ obtained by applying steps 1 to 3 from Appendix B.2, but replacing there -1 by m_L , 1 by m_R , $\hat{f}_{X',p}(x)$ by $\tilde{f}_X(x) \cdot 1_{\{x \in [m_L, m_R]\}}$, X' by X and $[x_L, x_R]$ by $[A_X^\dagger, B_X^\dagger]$.

Appendix C: Further numerical details

C.1. Pilot estimator of σ'

We construct a pilot estimator of σ' defined below (6.2), assuming that the variables were rescaled as at (6.1)–(6.2). If they were not rescaled, we take $W'_{i,j} = W_{i,j}$, $X'_i = X_i$ and $\sigma' = \sigma$ in the description below. Since the pilot estimator is only used to choose K and refine the upper bound to $\mathcal{S}_X$, it only needs to be a rough approximation to σ' . We use a parametric estimator based on the potentially misspecified form $\sigma'_{\text{pilot}}(x) = |\alpha + \beta x|$ for some parameters α and β to estimate. Noting that, for $\ell = 0, 1, \dots$,

$$\mathbb{E}\{\sigma'^2_{\text{pilot}}(X) \cdot X'^\ell\} = \alpha^2 \mathbb{E}(X'^\ell) + 2\alpha\beta \mathbb{E}(X'^{\ell+1}) + \beta^2 \mathbb{E}(X'^{\ell+2}), \quad (\text{C.1})$$

we estimate α and β by solving

$$\begin{cases} \hat{S}'_0 = \alpha^2 + 2\hat{\mu}_{X',1}\alpha\beta + \hat{\mu}_{X',2}\beta^2 \\ \hat{S}'_1 = \alpha^2\hat{\mu}_{X',1} + 2\hat{\mu}_{X',2}\alpha\beta + \hat{\mu}_{X',3}\beta^2, \end{cases} \quad (\text{C.2})$$

where $\hat{S}'_\ell$, $\ell \geq 0$ and $\hat{\mu}_{X',\ell}$, $\ell \geq 1$ are defined like $\hat{S}_\ell$ at (8.1) and $\hat{\mu}_{X,\ell}$ at (3.12), but with $W_{i,j}$ and X_i replaced by $W'_{i,j}$ and X'_i . Exploiting (C.1) with $\ell = 2$, if we find multiple solutions to (C.2), we choose the one that minimises $|\hat{S}'_2 - \hat{\mu}_{X',2}\alpha^2 - 2\hat{\mu}_{X',3}\alpha\beta - \hat{\mu}_{X',4}\beta^2|$. If (C.2) has no solution, we use instead the constant model $\sigma'_{\text{pilot}}(x) = \alpha$; then $\mathbb{E}\{\sigma'^2_{\text{pilot}}(X') \cdot X'^\ell\} = \alpha^2 \mathbb{E}(X'^\ell)$ and we estimate α by $\hat{\alpha} = \hat{S}'_0^{1/2}$.

C.2. Creating the X_i^* 's

We describe how we construct the X_i^* 's used in Section 5 (resp., B.3) from the data.

To construct the X_i^* 's in Section 5, we have two requirements: (1) mimic the X_i 's, which suggests taking $X_i^* = \overline{W}_i$ as they are more concentrated than the $W_{i,j}$'s around the X_i 's, and (2) ensure that $X_i^* \in [-1, 1]$, which is usually not satisfied by the $\overline{W}_i$'s. To satisfy (2) we could think of rescaling the $\overline{W}_i$'s so that they all lie in $[-1, 1]$, but if f_U is not compactly supported, then $f_{\overline{W}}$ and its rescaled versions aren't either. Instead, we rescale the $\overline{W}_i$'s so that most rescaled data are in $[-1, 1]$, and then discard

those that aren't. We proceed as follows. First, we compute the empirical quantiles $a_0 = \hat{q}_{\bar{W}, j_1/(2n)}$ and $b_0 = \hat{q}_{\bar{W}, 1-j_2/(2n)}$ of the $\bar{W}_i$'s, where $j_1 = \sum_{i=1}^n 1\{\bar{W}_i < -1\}$ and $j_2 = \sum_{i=1}^n 1\{\bar{W}_i > 1\}$; then we take

$$\tilde{W}_i = \begin{cases} \bar{W}_i & \text{if } a_0 \geq -1 \text{ and } b_0 \leq 1; \\ \bar{W}_i/|a_0| & \text{if } a_0 < -1 \text{ and } b_0 \leq 1; \\ \bar{W}_i/b_0 & \text{if } a_0 \geq -1 \text{ and } b_0 > 1; \\ \bar{W}_i/\{(b_0 - a_0)/2\} & \text{if } a_0 < -1 \text{ and } b_0 > 1. \end{cases}$$

Computed in this way, most $\tilde{W}_i$'s are in $[-1, 1]$. Let $n_1 = \sum_{i=1}^n I(\tilde{W}_i \in [-1, 1])$; wlog, reorder the $\tilde{W}_i$'s so that the first n_1 are in $[-1, 1]$. We take $X_1^*, \dots, X_{n_1}^*$ to be these first n_1 reordered $\tilde{W}_i$'s.

For Section B.3, we generate the $\tilde{W}_i$'s like the $\bar{W}_i$'s above except that we replace $\bar{W}_i$ by $\bar{W}_i'$. Wlog, reorder the $\bar{W}_i'$'s so that the first n_1 are in $[-1, 1]$. These mimic the X_i' 's at (6.2), that is $X_i' = (X_i - \mu)/\lambda$, where $\lambda = (B_X - A_X)/2$, $\mu = (B_X + A_X)/2$ and $[A_X, B_X]$ is the upper bound to $\mathcal{S}_X$ computed from the data. Therefore, for $i = 1, \dots, n_1$, we mimic the X_i 's by taking $X_i^* = \lambda \bar{W}_i' + \mu$.

Appendix D: Additional simulation results

D.1. Estimation of f_X

Table 2 shows the median and interquartile range (IQR) of the 200 $\text{ISE} \times 10^3$ for the two versions of our estimator: $\hat{f}_X$ (which assumes known $\mathcal{S}_X$) and $\tilde{f}_X$ (which does not assume known $\mathcal{S}_X$) in the case where $U \sim N(0, 1)$. To examine the performance of our estimator with f_U misspecified, we also applied $\hat{f}_X$ and $\tilde{f}_X$ to the samples generated as in Section 7.1 with $U \sim N(0, 1)$, but this time we computed those estimators under the incorrect assumption that $U = \sum_{\ell=1}^2 V_\ell/2$ with the V_ℓ 's i.i.d. $U[-\sqrt{6}, \sqrt{6}]$. The results shown in Table 3 illustrate that while the estimators performed worse in this case with misspecified f_U , the ISEs remained reasonable compared to the correctly specified case from Table 2.

Table 2. Simulation results for $\hat{f}_X$ (known $\mathcal{S}_X$) and $\tilde{f}_X$ (unknown $\mathcal{S}_X$): median (IQR) of $\text{ISE} \times 10^3$ for densities f_1 to f_4 , computed from 200 samples with f_U correctly specified.

f_X	NSR	$\hat{f}_X$				$\tilde{f}_X$			
		$n = 250$	$n = 500$	$n = 1000$	$n = 5000$	$n = 250$	$n = 500$	$n = 1000$	$n = 5000$
f_1	10%	12.65(16.28)	7.20(11.25)	4.00(4.25)	1.10(1.10)	13.50(15.65)	7.65(9.25)	4.35(6.35)	1.20(1.35)
	20%	17.70(21.33)	9.85(15.63)	5.75(9.03)	1.75(1.80)	19.70(26.17)	10.40(13.62)	5.55(8.23)	1.60(2.22)
f_2	10%	24.05(19.50)	14.65(15.08)	6.95(8.12)	1.55(1.52)	24.55(17.35)	15.20(10.70)	11.55(9.68)	4.85(3.50)
	20%	28.25(18.67)	20.40(18.15)	9.20(13.68)	2.80(4.20)	30.80(19.77)	23.50(11.40)	15.70(11.57)	7.70(5.98)
f_3	10%	12.90(17.18)	7.00(11.42)	4.10(5.90)	0.90(1.33)	14.65(11.43)	9.45(6.65)	5.70(3.73)	3.30(2.60)
	20%	19.45(18.30)	10.15(14.48)	5.10(8.85)	1.50(4.62)	26.45(19.23)	18.90(16.20)	11.65(9.38)	5.60(3.32)
f_4	10%	22.65(19.22)	13.65(11.70)	8.45(5.87)	2.65(4.62)	26.85(21.27)	16.50(13.15)	8.25(6.62)	3.55(2.90)
	20%	35.80(24.88)	26.20(19.45)	16.30(14.85)	8.10(5.37)	38.75(19.83)	24.85(18.82)	16.60(11.35)	6.35(5.50)

D.2. Estimation of σ^2

Computing the estimator $\hat{\sigma}^2$ at (8.3) requires choosing the smoothing parameter K , and the parameter ζ which prevents the denominator from being too close to zero. Proposing a sophisticated data-driven

Table 3. Simulation results for $\hat{f}_X$ (known S_X) and $\tilde{f}_X$ (unknown S_X): median (IQR) of $\text{ISE} \times 10^3$ for densities f_1 to f_4 , computed from 200 samples with f_U misspecified.

f_X	NSR	$\hat{f}_X$				$\tilde{f}_X$			
		$n = 250$	$n = 500$	$n = 1000$	$n = 5000$	$n = 250$	$n = 500$	$n = 1000$	$n = 5000$
f_1	10%	13.95(17.90)	7.10(13.48)	4.50(6.72)	1.10(1.10)	15.30(15.93)	7.85(9.23)	4.55(5.92)	1.30(1.40)
	20%	19.45(21.25)	13.20(18.32)	7.45(12.50)	1.90(2.35)	20.25(26.20)	9.85(14.77)	6.95(10.40)	1.85(2.45)
f_2	10%	25.85(17.92)	17.40(17.88)	8.30(11.22)	2.05(2.15)	26.70(18.23)	16.55(13.78)	12.00(9.05)	5.20(5.52)
	20%	32.75(22.40)	25.90(21.22)	17.05(18.65)	5.55(7.82)	30.80(21.00)	22.95(12.05)	15.65(12.58)	8.80(9.18)
f_3	10%	14.45(15.10)	7.40(11.20)	4.65(6.82)	1.20(8.35)	14.60(11.88)	9.45(6.73)	5.85(4.03)	4.00(3.25)
	20%	33.05(55.25)	20.75(28.80)	12.55(17.38)	4.00(10.52)	29.45(29.85)	21.70(20.58)	14.30(17.33)	7.75(10.00)
f_4	10%	25.40(16.25)	15.40(13.8)	8.60(6.85)	2.40(3.12)	26.75(19.15)	16.75(12.82)	8.80(6.82)	3.80(2.90)
	20%	32.95(24.7)	25.05(17.77)	17.60(13.95)	7.55(7.45)	40.95(22.45)	27.05(20.05)	16.05(15)	5.85(5.90)

selection rule of these parameters is beyond the scope of this paper. However, to illustrate the estimator, we ran a simulation study where we computed $\hat{\sigma}^2$ using simple rules for choosing ζ and K .

For ζ we took the maximum value of a normalised version of the density estimator $\hat{f}_X$ in a tail region determined by two empirical quantiles, as follows. Let $\tilde{f}_X^+ = \hat{f}_X^+ / \int_{-1}^1 \hat{f}_X^+$, where $\hat{f}_X^+ = \max(\hat{f}_X, 0)$ and $\hat{f}_X$ at (3.14) is computed with $\hat{K}$ from Section 5. We took $\zeta = \min(\hat{\zeta}_1, \hat{\zeta}_2)$, where $\hat{\zeta}_1 = 2\max_{x \leq q_{1/(2n)}} \hat{f}_X^+(x)$, $\hat{\zeta}_2 = 2\max_{x \geq q_{1-(1/(2n))}} \hat{f}_X^+(x)$, and, for $0 < \alpha < 1$, $q_\alpha = \hat{F}_X^{-1}(\alpha)$ with $\hat{F}_X(x) = \int_{-\infty}^x \hat{f}_X(u) du$.

To compute $\hat{h}$ and $\hat{f}_X$ in (8.3), motivated by the fact that $\sigma^2 = h/f_X$ can only be estimated reasonably well where f_X is not too small, we chose K by minimising an estimator of the integrated squared error $\int_{-1}^1 \{\hat{h}(x) - h(x)\}^2 dx$, which naturally puts a small weight on regions where f_X is small, since $h = \sigma^2 f_X$. Since $\int_{-1}^1 h^2(x) dx$ does not depend on K , minimising $\int_{-1}^1 \{\hat{h}(x) - h(x)\}^2 dx$ with respect to K is equivalent to minimising $\int_{-1}^1 \{\hat{h}(x)\}^2 dx - 2 \int_{-1}^1 \hat{h}(x)h(x) dx$ with respect to K . Noting that

$$\int_{-1}^1 \{\hat{h}(x)\}^2 dx - 2 \int_{-1}^1 \hat{h}(x)h(x) dx = \sum_{k=0}^K w_{h,k}^2 \cdot \hat{h}_k^2 - 2 \sum_{k=0}^K w_{h,k} \cdot \hat{h}_k h_k, \quad (\text{D.1})$$

where $w_{h,k} = 1\{k < K\} + (K+1-k) \cdot 1\{K \leq k < K+1\}$ as in Section 8.2, we estimated the second term of the right hand side of (D.1) by the cross-validation estimator $n^{-1} \sum_{i=1}^n \sum_{k=0}^K w_{h,k} \cdot \hat{h}_{k,-i} \hat{h}_{k,i}$, where $\hat{h}_{k,-i}$ and $\hat{h}_{k,i}$ denote, respectively, the estimator $\hat{h}_k$ at (8.2) computed without the i -th individual and with only the i -th individual. Then we chose K to be the first local minimum of

$$CV(K) = \sum_{k=0}^K w_{h,k}^2 \cdot \hat{h}_k^2 - \frac{2}{n} \sum_{i=1}^n \sum_{k=0}^K w_{h,k} \cdot \hat{h}_{k,-i} \hat{h}_{k,i}.$$

To illustrate the estimator $\hat{\sigma}^2$, we applied it to data generated as in our simulation study in Section 7.1, for $f_X = f_2$ and $f_X = f_4$. Recall that for each f_k , we had $\sigma(x) = c_k \exp(x/2)$, with c_k so that the noise-to-signal ratio $\text{NSR} = \mathbb{E}\{\sigma^2(X)\}/\text{Var}(X)$ was equal of 10% or 20%. Instead of assuming that we knew S_X , we estimated an upper bound $[A_X, B_X]$ to it as in Section 6.3 and applied $\hat{\sigma}^2$ to the data rescaled as at (6.1), where we chose K and ζ as described above.

In each case we applied $\hat{\sigma}^2$ to 200 samples and for each sample we computed the integrated squared error $\text{ISE} = \int_a^b (\hat{\sigma}^2 - \sigma^2)^2$, this time on an interval $[a, b]$ with a and b , respectively the 0.025 and 0.975 theoretical quantiles of f_X (recall that the denominator of $\hat{\sigma}^2$ is $\hat{f}_X$, and so we cannot go too far in the tails of f_X or else the denominator gets too close to zero). In Figure 6 we show, on the interval $[a, b]$

