Bivariate Kernel Deconvolution with Panel Data

Guillermo Basulto-Elias, Alicia L. Carriquiry, Kris De Brabanter and
Daniel J. Nordman
Iowa State University, Ames, USA

Abstract

We consider estimation of the density of a multivariate response, that is not observed directly but only through measurements contaminated by additive error. Our focus is on the realistic sampling case of bivariate panel data (repeated contaminated bivariate measurements on each sample unit) with an unknown error distribution. Several factors can affect the performance of kernel deconvolution density estimators, including the choice of the kernel and the estimation approach of the unknown error distribution. As the choice of the kernel function is critically important, the class of flat-top kernels can have advantages over more commonly implemented alternatives. We describe different approaches for density estimation with multivariate panel responses, and investigate their performance through simulation. We examine competing kernel functions and describe a flat-top kernel that has not been used in deconvolution problems. Moreover, we study several nonparametric options for estimating the unknown error distribution. Finally, we also provide guidelines to the numerical implementation of kernel deconvolution in higher sampling dimensions.

AMS (2000) subject classification. Primary 62G07; Secondary 65T50.

Keywords and phrases. Deconvolution density estimation, empirical characteristic function, flat-top kernel, measurement error.

1 Introduction

A challenge in many areas of application (e.g., medical sciences) is that responses of interest are not observable; instead, responses are measured with error contamination. Such cases are frequently referred to as “measurement error problems.” An example from nutrition epidemiology is the

Basulto-Elias’ and Carriquiry’s work was partially funded by the Center for Statistics and Applications in Forensic Evidence (CSAFE) through Cooperative Agreement #70NANB15H176 between NIST and Iowa State University, which includes activities carried out at Carnegie Mellon University, University of California Irvine, and University of Virginia. Daniel J. Nordman’s research has been partially funded by NSF DMS-1406747.

following. Suppose that we wish to estimate the joint density of the “usual” or long-run average serum iPTH (intact parathyroid hormone) and 25(OH)D (25-hydroxy vitamin D), associated with bone health, or of the intakes of sodium and potassium, which affect cardiovascular health. While daily values for these substances can be measured reliably, they are noisy estimates of their long-run averages, which are the quantities of interest. Ignoring such measurement error can lead to biased inference. Kernel deconvolution density estimation provides an alternative means to remove measurement error in estimating a target density, without the need for stringent parametric assumptions. Deconvolution typically involves approximating the characteristic function of a target variable through the estimated characteristic functions of noisy measurements and errors. Although kernel deconvolution has received much attention for univariate observations, the case of multivariate observations has received less consideration, especially for the more realistic case where little is known or can be immediately assumed about the error distribution and where the observations consist of one or more (correlated) measurements on subjects (i.e., panel data structure). In particular, there has been limited theoretical development for the case of bivariate panel data case and even less for the multivariate case.

In this paper, we explore the performance of deconvolution methods, in particular when applied to bivariate panel data. The performance of kernel deconvolution estimation in the multivariate setting depends critically on a combination of factors that have not been well studied. These include the choice of kernel, the way in which panel data are pre-processed, the number of measurements obtained for each sample unit, and the approach to estimating the unknown error distribution. We use simulation to study the impact of these potentially important factors in order to better understand the practical issues associated with implementation of kernel deconvolution and to inform possible future theoretical developments in this area.

In addressing bivariate kernel deconvolution, we consider five candidate kernel functions for which the Fourier transform has bounded support, including the sinc kernel (a popular choice in the univariate case) and a trapezoidal flat-top kernel (novel in deconvolution problems). Additionally, we compare different approximations to the (unknown) characteristic function of the errors with panel data. Such approximations involve choices on how to incorporate replicate measurements and combine these with either a direct empirical estimator or an indirect kernel density-based estimator of the characteristic function of the error. Perhaps surprisingly, our findings suggest that in contrast to some claims in the univariate setting, the way in

which panel data are processed (i.e., differenced to estimate the error distribution) has little impact, and that kernels commonly used with univariate measurements do not perform as well as other kernels (e.g., “flat top”) studied here for the multivariate case. Further, kernel deconvolution estimators for multivariate responses pose more computational challenges, in particular for numerical integration. To overcome the computational challenges, we use an algorithm based on a Fast Fourier Transform (FFT) that improves efficiency without compromising accuracy.

We end this section with a brief review of the literature. An extensive discussion of the univariate deconvolution problem can be found in Meister (2009). For the multivariate setting, asymptotic convergence properties when the error distribution is assumed to be known can be found in Masry (1991, 1993a, b). Selection of a single bandwidth parameter (diagonal bandwidth matrix with equal diagonal entries) using cross-validation was proposed by Youndjé and Wells (2008); Comte et al. (2013) also focused on diagonal bandwidth matrices and proposed estimating their diagonal elements as by minimizing an empirical approximation to the asymptotic integrated mean squared error. Johannes (2009) and Comte and Lacour (2011) considered the situation where the error distribution is unknown, but a sample of pure errors is available. None of these papers considered multivariate panel data (i.e., repeated multivariate measurements on each subject, contaminated with error). In the univariate panel data setting, with unknown error distribution, several approaches have been proposed for estimating the characteristic function of the error. Delaigle and Meister (2008) suggested using all possible measurement differences available for each individual with deconvolution applied to the measurement averages per subject. Stirne-mann et al. (2012) and Comte et al. (2013) proposed taking simple pairwise differences of the replicate measurements to approximate the characteristic function of the error; only the first noisy measurement on each subject is then used in the deconvolution steps that follow. Ridge terms or truncation, as in Neumann and Hössjer (1997), are typically applied to control the estimated characteristic function of errors for small values. In much of the literature discussing deconvolution with univariate panel data, the most frequently chosen kernel is the sinc kernel because it has been shown to have good theoretical properties (e.g., its Fourier transform is flat around zero as suggested by Devroye, 1989) and because its form permits applying the FFT for numerical integration.

This paper is organized as follows. Section 2 describes the general density estimation problem and how it can be approached with kernel methods, with special attention to the panel data case introduced in Section 2.2. In

Section 3, we outline several potentially important attributes of the deconvolution method that may impact performance. In particular, Section 3.1 discusses how to approximate the characteristic function of errors from a panel data structure, and Section 3.2 describes candidate kernel functions. Section 3.3 provides numerical integration tools required for bivariate kernel density deconvolution. In Section 4, we present the design of a large simulation study and discuss the numerical results. Section 5 provides a data illustration of the multivariate deconvolution methods to a problem in public health. Finally, conclusions are presented in Section 6.

2 Background

We first outline estimation for multivariate kernel density deconvolution in general. Section 2.1 introduces some initial background in the context of a simplified version of kernel deconvolution density estimation in the multivariate setting. Section 2.2 then discusses the case of multivariate panel data and some related estimation issues that arise in this situation. For illustrative purposes, we focus on the case of bivariate measurements $\mathbf{X} \in \mathbb{R}^d$, $d = 2$, although generalizations to any dimension d of the observation vector are possible.

2.1. A Simplified Estimation Framework Suppose that we are interested in estimating the distribution of a two-dimensional random variable $\mathbf{X}$ that cannot be directly observed. Instead we observe $\mathbf{X}$ convoluted with random noise, say, $\mathbf{Y} = \mathbf{X} + \boldsymbol{\epsilon}$, where $\boldsymbol{\epsilon} \in \mathbb{R}^2$ is the random noise vector. The standard deconvolution estimation problem consists of estimating the density of $\mathbf{X}$ using a random sample of measurements

$$\mathbf{Y}_i = \mathbf{X}_i + \boldsymbol{\epsilon}_i, \quad (2.1)$$

for $i = 1, \dots, n$. In Eq. 2.1, the collection $\{\mathbf{X}_i\}_{i=1}^n$ is independent and identically distributed (iid) and so is the variable set $\{\boldsymbol{\epsilon}_i\}_{i=1}^n$. Furthermore, for each measurement $\mathbf{Y}_i$, the response $\mathbf{X}_i$ of interest is assumed to be independent of the error $\boldsymbol{\epsilon}_i$.

We require some notation before introducing the deconvolution formulas. Let $\phi_{\mathbf{Y}}(\cdot)$ denote the characteristic function (CF) of the random sample $\mathbf{Y}_1, \dots, \mathbf{Y}_n$, and let $\hat{\phi}_{\mathbf{Y},n}(\cdot)$ denote the empirical CF. That is, for any $\mathbf{t} \in \mathbb{R}^2$,

$$\phi_{\mathbf{Y}}(\mathbf{t}) = \mathbb{E}(e^{i\mathbf{t} \cdot \mathbf{Y}_1}) \quad \text{and} \quad \hat{\phi}_{\mathbf{Y},n}(\mathbf{t}) = \frac{1}{n} \sum_{j=1}^n e^{i\mathbf{t} \cdot \mathbf{Y}_j},$$

where the dot product $\cdot$ above represents the usual inner product in $\mathbb{R}^2$ and $\imath = \sqrt{-1}$. For a real-valued function $g : \mathbb{R}^2 \rightarrow \mathbb{R}$, denote its Fourier transform by g^* , where

$$g^*(\mathbf{t}) = \int e^{\imath \mathbf{t} \cdot \mathbf{x}} g(\mathbf{x}) d\mathbf{x}. \quad (2.2)$$

When g is a (multivariate) density function, then g^* represents its CF.

By the independence of $\mathbf{X}_i$ and $\boldsymbol{\epsilon}_i$ in Eq. 2.1, the CF $\phi_{\mathbf{X}}(\cdot)$ of the uncontaminated measurements $\mathbf{X}$ can be obtained from that of $\phi_{\mathbf{Y}}(\cdot)$, the CF of the observed measurements, as

$$\phi_{\mathbf{X}}(\mathbf{t}) = \frac{\phi_{\mathbf{Y}}(\mathbf{t})}{\phi_{\boldsymbol{\epsilon}}(\mathbf{t})}, \quad \mathbf{t} \in \mathbb{R}^2,$$

whenever the CF of the errors $\phi_{\boldsymbol{\epsilon}}(\cdot)$, is non-zero (i.e., $\mathbf{t} \in \mathbb{R}^2$ with $\phi_{\boldsymbol{\epsilon}}(\mathbf{t}) \neq 0$). Consequently, by applying the Fourier inversion theorem, we can compute the density of $\mathbf{X}$ at $\mathbf{x} \in \mathbb{R}^2$ as

$$f_{\mathbf{X}}(\mathbf{x}) = \frac{1}{(2\pi)^2} \int e^{-\imath \mathbf{t} \cdot \mathbf{x}} \frac{\phi_{\mathbf{Y}}(\mathbf{t})}{\phi_{\boldsymbol{\epsilon}}(\mathbf{t})} d\mathbf{t}, \quad (2.3)$$

which is the density of the uncontaminated response $\mathbf{X}$. Because $\phi_{\mathbf{Y}}$ is typically unknown, we may replace it with its empirical version, $\hat{\phi}_{\mathbf{Y},n}$, but the resulting function in Eq. 2.3 might not be integrable. However, by substituting a smoothed estimate instead, say $\hat{\phi}_{\mathbf{Y},n}(\mathbf{t})K^*(B_n\mathbf{t})$, we obtain a valid (kernel) estimator of the density of $\mathbf{X}$ as

$$\hat{f}_{\mathbf{X},n}(\mathbf{x}) = \frac{1}{(2\pi)^2} \int e^{-\imath \mathbf{t} \cdot \mathbf{x}} \hat{\phi}_{\mathbf{Y},n}(\mathbf{t}) \frac{K^*(B_n\mathbf{t})}{\phi_{\boldsymbol{\epsilon}}(\mathbf{t})} d\mathbf{t}, \quad \mathbf{x} \in \mathbb{R}^2. \quad (2.4)$$

Here K^* denotes the Fourier transform of a smoothing kernel K and B_n denotes a 2×2 (positive definite) bandwidth matrix.

Note that Eq. 2.4 assumes that the CF of the underlying error terms, $\phi_{\boldsymbol{\epsilon}}$, is known. Under the general data model (2.1), it is frequently assumed that either the error distribution is known or that there exists a sample of pure errors, obtained independently of $\mathbf{Y}_1, \dots, \mathbf{Y}_n$, that can be used to estimate the error CF (see Johannes, 2009 and Comte and Lacour, 2011). These assumptions are often unrealistic in many applications. Therefore, in this paper we will not make such assumptions when describing the estimation with panel data in Section 2.2.

2.2. *Kernel Deconvolution Estimation with Panel Data* For bivariate panel data, an additive noise model is given by

$$\mathbf{Y}_{ij} = \mathbf{X}_i + \boldsymbol{\epsilon}_{ij}, \quad (2.5)$$

for $i = 1, \dots, n$ and $j = 1, \dots, m$. Here $\mathbf{Y}_{ij}$ represents the $d = 2$ observations for the i -th individual on the j -th measurement occasion. As in model (2.1), the iid uncontaminated response variables $\{\mathbf{X}_i\}_{i=1}^n$ are independent of the iid errors $\{\boldsymbol{\epsilon}_{ij} : i = 1, \dots, n; j = 1, \dots, m\}$.

For simplicity, and without loss of generality, we assume that the number m of observations is the same for every sample unit (individual). Note that while sample units are assumed to be independent, the components among the d -dimensional measurements for each unit are typically not independent, and the replicate measurements obtained on each unit are also correlated.

If the error distribution is unknown under a panel data structure, the CF of the error is approximated based on differences of errors. We require then additional assumptions about the error distribution, specifically, we assume that the distribution of the panel data errors is symmetric and its CF is never zero. More details are provided in Section 3.

We wish to estimate the joint density $f_{\mathbf{X}}$ of the uncontaminated variables $\{\mathbf{X}_i\}_{i=1}^n$. Under a panel data structure, the standard deconvolution density estimator (2.4) needs to be modified to account for the dependence among replicate measurements on each sample unit. That is, measurements made on the same subject are no longer independent, as was the case in Section 2.1 (i.e., instead two measurements $\mathbf{Y}_{ij}$ and $\mathbf{Y}_{i'j'}$ are independent only if $i \neq i'$).

As in the univariate panel data setting, there are several ways to account for the panel data structure when using the estimator (2.4) of the density $f_{\mathbf{X}}$. Comte et al. (2013) use only the first measurement for each subject $\{\mathbf{Y}_{i1}\}_{i=1}^n$ to estimate the common CF of the measurements, $\phi_{\mathbf{Y}}$. However, this approach ignores the other $m - 1$ measurements available for each subject and therefore may entail some information loss. Instead, we propose making use of all observations obtained on each subject to compute an average measurement for each individual; we then re-formulate the deconvolution estimator to be based on the CF of the averages, say $\phi_{\bar{\mathbf{Y}}}$, rather than on the CF of individual responses $\phi_{\mathbf{Y}}$ in Eq. 2.4. That is, we first average observations in Eq. 2.5 per individual,

$$\bar{\mathbf{Y}}_i = \mathbf{X}_i + \bar{\boldsymbol{\epsilon}}_i, \quad i = 1, \dots, n. \quad (2.6)$$

When the response variable is the average of m replicate measurements, the error term $\bar{\boldsymbol{\epsilon}}_i$ in Eq. 2.6 represents an average of m iid error variables with

CF given by $\phi_{\bar{\epsilon}}(\mathbf{t}) = [\phi_{\epsilon}(\mathbf{t}/m)]^m$ in terms of the CF ϕ_{ϵ} of an original (i.e., single) error term in the panel data model (2.5). Therefore, instead of using the kernel density deconvolution estimator (2.4), we use a modified estimator given by

$$\begin{aligned}\hat{f}_{\mathbf{X},n}(\mathbf{x}) &= \frac{1}{(2\pi)^2} \int e^{-i\mathbf{t}\cdot\mathbf{x}} \hat{\phi}_{\mathbf{Y},n}(\mathbf{t}) \frac{K^*(B_n\mathbf{t})}{\phi_{\bar{\epsilon}}(\mathbf{t})} d\mathbf{t} \\ &= \frac{1}{(2\pi)^2} \int e^{-i\mathbf{t}\cdot\mathbf{x}} \hat{\phi}_{\mathbf{Y},n}(\mathbf{t}) \frac{K^*(B_n\mathbf{t})}{[\phi_{\epsilon}(\mathbf{t}/m)]^m} d\mathbf{t},\end{aligned}\quad (2.7)$$

for $\mathbf{x} \in \mathbb{R}^2$ based on the CF of error averages, $[\phi_{\epsilon}(\mathbf{t}/m)]^m$ and the empirical CF of subject averages is $\{\hat{\mathbf{Y}}_i\}_{i=1}^n$, $\hat{\phi}_{\mathbf{Y},n}(\mathbf{t}) = n^{-1} \sum_{j=1}^n e^{i\mathbf{t}\cdot\hat{\mathbf{Y}}_j}$, $\mathbf{t} \in \mathbb{R}^2$.

There are, however, several issues that require examination when implementing an estimator of the form (2.7) with panel data. These can impact the performance of the kernel deconvolution estimator $\hat{f}_{\mathbf{X},n}$, which we numerically investigate in Section 4. Firstly, the error distribution is typically unknown so that the associated CF ϕ_{ϵ} in Eq. 2.7 needs to be estimated. By exploiting the repeated measurements in the panel data structure, we can construct at least two different estimators, both of which are based on differences between measurements within subjects. We elaborate on this issue in Section 3. Potential formulations for constructing differences between measurements in panel data are also outlined in that section. A second issue that arises is that the deconvolution formula (2.7) requires specification of a multivariate kernel function and the bandwidth matrix for that kernel. Section 3.2 discusses several types of kernels that may be considered, including “flat-top” kernels that have been successfully used for spectral density estimation but that have not been applied in deconvolution problems (cf. Politis and Romano, 1995, 1999; Politis, 2003). Finally, the deconvolution estimator includes integrals that are often numerically approximated; in contrast to the univariate data case, the multivariate setting introduces further challenges for the evaluation of a density estimator $\hat{f}_{\mathbf{X},n}(\mathbf{x})$ at vector points $\mathbf{x} \in \mathbb{R}^2$. In Section 3.3, we describe some numerical recipes that greatly improve computational efficiency in the multivariate deconvolution setting.

3 Challenges in Deconvolution Estimation with Panel Data

3.1. Estimation of the Error Distribution Assume for now that the number m of observations for each individual is even for the panel data (2.5). In the univariate setting, Comte et al. (2013) proposed constructing pairwise differences between repeated measurements (within individuals) as

$$\mathbf{Y}_{i,2l-1} - \mathbf{Y}_{i,2l} = \epsilon_{i,2l-1} - \epsilon_{i,2l}, \quad (3.1)$$

for $i = 1, \dots, n$ and $l = 1, \dots, m/2$. The reason for formulating differences within individuals is that the measurement differences correspond to differences in error terms (i.e., common unobserved variables $\mathbf{X}_i$ drop out). Each measurement is used only once to create a difference in Eq. 3.1, and therefore the $nm/2$ differences across all individuals are iid. Another possibility is to consider all possible within-individual differences, as in Delaigle and Meister (2008); that is

$$\mathbf{Y}_{i,j} - \mathbf{Y}_{i,j'} = \boldsymbol{\epsilon}_{i,j} - \boldsymbol{\epsilon}_{i,j'}, \quad (3.2)$$

for $i = 1, \dots, n$ and $j < j'$. This approach results in a higher number of measurement differences (i.e., $nm(m-1)/2$ differences) which, though identically distributed, are not independent. Finally, we might also consider an intermediate method, which leads to an intermediate sized collection of measurement differences, with a weaker dependence structure relative to Eq. 3.2. In this approach, the last $m-1$ measurements for each sample unit are deviated from the first one as follows:

$$\mathbf{Y}_{i,j} - \mathbf{Y}_{i,1} = \boldsymbol{\epsilon}_{i,j} - \boldsymbol{\epsilon}_{i,1}, \quad (3.3)$$

for $i = 1, \dots, n$ and $j = 2, 3, \dots, m$. In this case, there are $n(m-1)$ measurement differences which are not independent. In contrast to Eq. 3.1, the differencing strategies (3.2) and (3.3) do not require an even number of observations per individual. The simulation experiments in Section 4 compare the performance of deconvolution estimators based on these three types of measurement differences.

Regardless of the chosen approach, the processed data consist of a collection of differenced measurements or error differences, say $\boldsymbol{\delta}_1, \dots, \boldsymbol{\delta}_k$, where the number of differences k depends on the differencing approach (3.1)-(3.3). We propose a methodology to estimate the error CF ϕ_ϵ in Eq. 2.7 with these differences. By the iid property of individual error terms, observe that, for $\mathbf{t} \in \mathbb{R}^2$ and $j \neq j'$, the CF ϕ_δ of a within individual measurement difference is given by

$$\phi_\delta(\mathbf{t}) = \phi_{\boldsymbol{\epsilon}_{i,j} - \boldsymbol{\epsilon}_{i,j'}}(\mathbf{t}) = \phi_{\boldsymbol{\epsilon}_{i,j}}(\mathbf{t})\phi_{\boldsymbol{\epsilon}_{i,j'}}(-\mathbf{t}) = \phi_\epsilon(\mathbf{t})\phi_\epsilon(-\mathbf{t}) = |\phi_\epsilon(\mathbf{t})|^2. \quad (3.4)$$

If we assume that the distribution of the panel data errors is symmetric and its CF is never zero, then

$$\phi_\epsilon(\mathbf{t}) = \sqrt{\phi_\delta(\mathbf{t})} \quad (3.5)$$

holds when ϕ_ϵ is real-valued (by symmetry) and also positive (by $\phi_\epsilon(0) = 1$ and by the assumption that ϕ_ϵ is never zero). Thus, under the multivariate

panel data structure, an estimator for the CF of an error term, ϕ_ϵ , is given by

$$\hat{\phi}_\epsilon(\mathbf{t}) = \sqrt{\left| \Re \left(\hat{\phi}_\delta(\mathbf{t}) \right) \right|}, \quad \mathbf{t} \in \mathbb{R}^2, \quad (3.6)$$

where $\Re(\cdot)$ denotes the real part of a complex number and $\hat{\phi}_\delta(\mathbf{t}) = k^{-1} \sum_{j=1}^k \exp(i\mathbf{t} \cdot \delta_j)$, $\mathbf{t} \in \mathbb{R}^2$, is the empirical CF of the measurement differences $\{\delta_j\}_{j=1}^k$.

Note that the estimator (3.6) of the (single) error term does not require stringent assumptions on the distribution of the errors. To our knowledge, empirical CFs seem to be the main nonparametric approach for estimating such error CFs in the deconvolution literature (albeit often in the non-panel data setting where an independent sample of pure errors is assumed to be available). If one knows the distribution of the error, or that the variables are measured without error, other estimators, like a parametric estimator, could be used instead of the estimator in Eq. 3.6, which is likely to improve the overall behavior and the computational efficiency in practice.

In the simulations of Section 4, we also consider an alternative approach for estimating the error CF, for comparison to the empirical CF (3.6) of differences. This alternative approach uses measurement (or error) differences $\{\delta_j\}_{j=1}^k$ to formulate a kernel density estimator for the distribution of the differences; a Fourier transform of this density then provides an estimator $\hat{\phi}_\delta$ of the CF ϕ_δ . Instead of the empirical CF of differences, this estimator $\hat{\phi}_\delta$ can then be substituted in Eq. 3.6 to approximate the error term CF ϕ_ϵ , which generally produces a smoother version of the empirical CF based on differences. Suppose a kernel density estimator is used to approximate the density of δ , using a (bivariate) kernel L and bandwidth matrix Σ . The corresponding CF of the density estimator is then given by

$$\hat{\phi}_\delta(\mathbf{t}) = \frac{1}{M} \sum_{j=1}^M \exp(i\mathbf{t} \cdot \delta_j) L^* \left(\Sigma^{1/2} \mathbf{t} \right), \quad \mathbf{t} \in \mathbb{R}^2. \quad (3.7)$$

In particular, if L is the product kernel (see Section 3.2) generated by the normal distribution, we have

$$\hat{\phi}_\delta(\mathbf{t}) = \frac{1}{M} \sum_{j=1}^M \exp \left(i\mathbf{t} \cdot \delta_j - \frac{1}{2} \mathbf{t}' \Sigma \mathbf{t} \right), \quad \mathbf{t} \in \mathbb{R}^2. \quad (3.8)$$

We choose a normal kernel above because its Fourier transform (3.8) has an efficient, closed form expression. Another possibility, not considered here,

would be to use the CF of a normal mixture with a data-driven number of components. In Section 4, we specify how we select the bandwidth matrix Σ .

Ultimately, an estimated CF $\hat{\phi}_\epsilon$ of the error (or an estimator $\hat{\phi}_{\bar{\epsilon}}$) appears in the denominator of the deconvolution formula (2.7). This means that numerical approximations or estimates of this CF can potentially lead to instabilities for small values of the CF. Several authors have proposed ways to avoid computational issues when using an estimated error CF in the denominator of the deconvolution estimator. For example, Meister (2009) describes two regularization methods: either use a ridge parameter or truncate the denominator in the deconvolution formula when the value of the estimated CF is very small. In our simulations (see Section 4), we truncate the integrand when the denominator in Eq. 2.7 has small values, as in Neumann and Hössjer (1997). Namely, the term $1/\phi_{\bar{\epsilon}}(\mathbf{t}) = 1/[\phi_\epsilon(\mathbf{t}/m)]^m$ is approximated by

$$\frac{1}{\hat{\phi}_{\bar{\epsilon}}(\mathbf{t})} I_{\{|\hat{\phi}_{\bar{\epsilon}}(\mathbf{t})| > b_{n,m}\}}, \quad (3.9)$$

for $\hat{\phi}_{\bar{\epsilon}}(\mathbf{t}) = [\hat{\phi}_\epsilon(\mathbf{t}/m)]^m$ based on the estimated error CF $\hat{\phi}_\epsilon$ (i.e., using one of the approaches described above) and for $b_{n,m} = 1/\sqrt{k}$ based on the number k of differences used to determine $\hat{\phi}_\epsilon$ (i.e., k is equal to $nm(m-1)/2$, $n(m-1)$ or $nm/2$).

3.2. Choice of Kernel Function In standard kernel density estimation, it is common to construct multivariate kernel functions starting with univariate kernels. The simplest approach is to define a multivariate kernel as a product of univariate kernels

$$K(\mathbf{x}) = K_1(x_1)K_1(x_2), \quad \mathbf{x} = (x_1, x_2) \in \mathbb{R}^2, \quad (3.10)$$

where K_1 is a univariate kernel function. Such multivariate kernels (3.10) are called *product kernels* and their corresponding Fourier transform is given by

$$K^*(\mathbf{t}) = K_1^*(t_1)K_1^*(t_2), \quad \mathbf{t} = (t_1, t_2) \in \mathbb{R}^2. \quad (3.11)$$

Another alternative, not pursued here, is the family of radial-symmetric kernels (cf. Wand and Jones, 1995).

Computation of the deconvolution kernel density estimator (2.4) or (2.7) involves integration over the domain $\mathbb{R}^2$ (i.e., for bivariate data). If a bivariate kernel K is chosen in such way that its Fourier transform K^* has bounded support, then the integral in Eq. 2.7 is restricted to a finite region. In what follows, we choose univariate kernels $K_1 = K_2$ that have

Fourier transforms with support on an finite interval, say $[-1, 1]$. Then, without loss of generality, the resulting bivariate product kernel has Fourier transform K^* supported on $[-1, 1] \times [-1, 1]$. Because this bounded-support property simplifies integration and is also associated with kernels that have good properties such as ensuring integrability conditions required for Fourier inversion, this type of kernel function is commonly used in deconvolution applications (see Meister, 2009 for a deeper discussion of kernels with Fourier transform with bounded support). Table 1 displays various univariate kernel functions K_1 and their respective Fourier transforms K_1^* for application in Eq. 3.11 and then Eq. 2.7.

Among univariate kernels K_1 with a Fourier transform with bounded support, the most common is the sinc kernel, whose Fourier transform K_1^* corresponds to a simple indicator function. The sinc kernel has well-known theoretical advantages (see below), but is not non-negative. Consequently, density estimates based on the sinc kernel tend to exhibit wiggly behavior at the tails, with an unattractive appearance; this behavior is illustrated with numerical examples in Section 4.

As pointed out by Devroye (1989), kernels K_1 which perform well in estimation can often be associated with corresponding Fourier transforms that are flat around zero. The sinc kernel has this feature, but other kernels are often adopted because they produce smoother estimates of the target density. An alternative to the sinc kernel is the de la Vallée Poussin kernel (VP kernel), defined in Table 1. This kernel is actually a probability density (i.e., is non-negative), but its Fourier transform is sharp around zero and

Table 1: Kernel functions with their Fourier transforms, where $x, t \in \mathbb{R}$ and K_1 is continuously extended at zero in every case

Name	Kernel (K_1)	Fourier Transform (K_1^*)
Sinc	$\frac{\sin(x)}{\pi x}$	$I_{[-1,1]}(t)$
de la Vallée-Poussin	$\frac{1 - \cos(x)}{\pi x^2}$	$(1 - t) I_{[-1,1]}(t)$
Triweight Ft	$\frac{48(1 - 15x^{-2}) \cos x - \frac{\pi x^4}{144(2 - 5x^{-2})} \sin x}{\pi x^5}$	$(1 - t^2)^3 I_{[-1,1]}(t)$
Tricube Ft	$\frac{648(-3x^4 + 90x^2 - 560) \sin(x) + \frac{x^9 \pi}{162(x^6 - 80x^4 + 1120x^2 - 2240) \cos(x) + \frac{18(20160 - x^6)}{x^{10} \pi}}}{\pi x^5}$	$(1 - t ^3)^3 I_{[-1,1]}(t)$
Flat-top Ft	$\frac{2}{\pi x^2} \left[\cos\left(\frac{x}{2}\right) - \cos(x) \right]$	$I_{\{ t \leq 1/2\}} + 2(1 - t) I_{\{1/2 \leq t \leq 1\}}$

the numerical findings in Section 4 suggest that this kernel exhibits poor performance for bivariate deconvolution density estimation.

Between the sinc and VP kernels, there are other kernel functions whose Fourier transforms are flat around zero to varying degrees. We consider three of them (triweight, tricube, trapezoidal flat-top), described in Table 1. Figure 1 displays all the kernels we have mentioned and their Fourier transforms. The kernel with Fourier transform proportional to the triweight kernel was used by Delaigle and Gijbels (2004a, b), and resembles the kernel whose Fourier transform is proportional to the tricube kernel in Table 1. The flat-top kernel has been seemingly unknown in the deconvolution literature but has established properties in density estimation for time series (cf. Politis, 2003). Section 4 demonstrates that the flat-top kernel produces density estimates with attractive shape properties while maintaining good (integrated squared error) performance relative to the other commonly used kernels for deconvolution estimation.

3.3. Numerical Implementation The deconvolution kernel density estimator $\hat{f}_{\mathbf{X},n}$ in Eq. 2.7 requires the computation of an integral that does not have a closed expression, and the oscillatory nature of the integrand introduces challenges for numerical approximations. We provide numerical integration recipes to address this practical issue. General-purpose software and algorithms for evaluating integrals in multiple dimensions do exist

Figure 1: Plots of kernel functions and their Fourier transforms: Sinc, de la Vallée-Poussin (VP), triweight Fourier transform (triw), tricube Fourier transform (tric) and Flat-top Fourier transform kernels. The kernel function that oscillates the most is the sinc kernel

(mostly Monte Carlo methods); an example is the package called **CUBA**, described in Hahn (2005). However, when applying such methods to compute the estimator $\hat{f}_{\mathbf{X},n}(\mathbf{x})$ at different values $\mathbf{x} \in \mathbb{R}^2$, the exponential term in the integrand in Eq. 2.7 is evaluated repeatedly for the same $\mathbf{x}$ values, which is time-consuming and computationally wasteful. Therefore, we propose approximating the integrals in Eq. 2.4 using faster and computationally more efficient approaches.

One simple approximation to the integral (2.7) is based on Riemann sums as

$$\begin{aligned} \hat{f}_{\mathbf{X},n}(\mathbf{x}) &= \frac{1}{(2\pi)^2} \int_{[-1,1]^2} e^{-i\mathbf{t} \cdot \mathbf{x}} \hat{\phi}_{\mathbf{Y},n}(\mathbf{t}) \frac{K^*(\mathbf{B}_n \mathbf{t})}{\phi_{\epsilon}(\mathbf{t})} d\mathbf{t} \\ &\approx \frac{\Delta}{(2\pi)^2} \sum_{\mathbf{t} \in G} e^{-i\mathbf{t} \cdot \mathbf{x}} \hat{\phi}_{\mathbf{Y},n}(\mathbf{t}) \frac{K^*(\mathbf{B}_n \mathbf{t})}{\phi_{\epsilon}(\mathbf{t})}, \end{aligned} \quad (3.12)$$

where G denotes a regular grid on $[-1,1]^2$ of $L_1 \cdot L_2$ (L_i points in the i -th coordinate) and Δ denotes the area between grid points. Note that the integration limits are $[-1,1]^2$ rather than $\mathbb{R}^2$ when we focus on kernel functions with Fourier transforms supported on $[-1,1]^2$. (Furthermore, in practice, one would estimate $1/\phi_{\epsilon}(\cdot)$ as described in Section 3.1.)

In application, we also need to evaluate the density estimator $\hat{f}_{\mathbf{X},n}(\cdot)$ over a grid of points of $\mathbf{X}$ so that probability estimates can be obtained by numerically integrating $\hat{f}_{\mathbf{X},n}(\cdot)$. This is a different matter than the problem of computing $\hat{f}_{\mathbf{X},n}(\cdot)$ at a single point $\mathbf{x}$, see Comte and Lacour (2010). For that purpose, we carry out density estimation $\hat{f}_{\mathbf{X},n}(\mathbf{x})$ on a regular grid of points $\mathbf{x}$, say on $[c_1, d_1] \times [c_2, d_2]$ with $L_3 \cdot L_4$ points; here c_1, d_1, c_2, d_2 denote the bounds of a rectangular region for computing density estimators, selected by the user.

We considered two approaches for computing the Riemann sums in Eq. 3.12. The first approach computes all quantities in Eq. 3.12 which do not depend on $\mathbf{x}$ (requiring evaluation only on the subgrid $G \subset [-1,1]^2$) and then evaluates exponential terms at $\mathbf{x}$ -points on the $[c_1, d_1] \times [c_2, d_2]$ to determine the Riemann sums. Another alternative is to use the Fast Fourier Transform (FFT) to compute the approximating sums in Eq. 3.12. We used the FFT method developed in Inverarity (2002), which is the generalization to the multivariate case of Bailey and Swarztrauber (1994).

The FFT imposes the restrictions $L_1 = L_3$ and $L_2 = L_4$, which are not required by the plain Riemann sum method. However, the FFT approach is much faster, as can be seen in Fig. 2. Figure 2 compares the time required to obtain density estimates on a grid (using simulated panel data described in Section 4) based on both the FFT and plain Riemann sum approaches. (For

Figure 2: Speed comparison of integrating procedures. The Fast Fourier Transform (FFT) is substantially faster than plain Riemann sums for increasingly larger approximation grids

the sake of simplicity, to produce the Fig. 2, we considered the same resolution for evaluating both methods, that is, $L_1 = L_2 = L_3 = L_4$). Note that both approaches have the same numerical accuracy as each computes the same Riemann approximation (3.12) to the integral. However, more accurate approximations to this integral require higher resolution grids $G \subset [-1, 1]^2$, creating larger computational demands where the FFT has speed advantages. Hence, we recommend the use of the FFT-based approach in practice to approximate integrals for bivariate deconvolution estimators. In what follows, we implemented the FFT approximation to carry out all calculations in the simulation studies to follow.

We have also developed a software package called `fourierin`, made available in R, for computing multivariate deconvolution estimators on a grid through the FFT-based numerical approximations of the required integrations; see Basulto-Elias et al. (2017).

4 Simulation Study

4.1. Design of the Simulation Study In this section, we numerically examine several factors that have an impact on the performance of bivariate

kernel deconvolution. We begin by describing distributions for data generation and for the underlying target density. Fan (1991) characterizes distributions according to the smoothness of their CF. In our context, the smoother the error distribution, the harder it becomes to estimate the target density of the underlying signal variable. Formally, a (possibly multivariate) random variable $\mathbf{X}$ is said to be *ordinary smooth* (OS) if there exist positive constants β , d_0 and d_1 such that

$$d_0|\mathbf{t}|^{-\beta} \leq |\phi_{\mathbf{X}}(\mathbf{t})| \leq d_1|\mathbf{t}|^{-\beta}, \quad \text{as } |\mathbf{t}| \rightarrow \infty,$$

and it is said to be *supersmooth* (SS) if there exist positive constants β , δ , γ , d_0 and d_1 such that

$$d_0|\mathbf{t}|^{-\beta} \exp\left(-\gamma|\mathbf{t}|^\delta\right) \leq |\phi_{\mathbf{X}}(\mathbf{t})| \leq d_1|\mathbf{t}|^{-\beta} \exp\left(-\gamma|\mathbf{t}|^\delta\right), \quad \text{as } |\mathbf{t}| \rightarrow \infty,$$

where $|\cdot|$ is the Euclidean norm. Unlike OS variables, SS random variables have densities which are infinitely differentiable. In the univariate case, examples of OS random variables include those with the gamma distribution and the Laplace distribution while examples of SS distributions include Gaussian and Cauchy densities. To understand the impact of distribution smoothness on deconvolution estimators here, note that the empirical characteristic function (CF) of the contaminated samples appears in the numerator of Eq. 2.4, while the CF of the error appears in the denominator. This same ratio of CFs also arises in asymptotic expansions for mean squared error of deconvolution density estimators. As a consequence, supersmooth error terms then lead to slower convergence rates in density estimation, because they introduce an additional exponential term into such expansions.

In the simulation study, we explored the impact of several factors on the performance of the bivariate deconvolution. We included four combinations of ordinary smooth (OS) and supersmooth (SS) distributions across the signal and the noise, where random errors are generated either from a bivariate normal distribution (SS error) or a bivariate Laplace distribution with the same covariance matrix (OS error), and where signal random variable is generated from either a bivariate normal distribution (SS signal) or from a bivariate gamma distribution (OS signal). The choice of the Laplace and Gamma distributions was pragmatic; the Laplace distribution is often used to model error distributions and happens to be Laplace, and the bivariate Gamma distribution has a flexible shape thanks to its parameters. Table 2 shows two signal distributions (OS/SS) that have the same mean and variance properties, while Fig. 3 shows the contour plot of the distribution of

Table 2: Distributions used in the study for signal and noise

Type	Distribution
OS signal	Bivariate gamma with mean $(10, 2)$ and cov. $\begin{bmatrix} 5 & -1.75 \\ -1.75 & 2 \end{bmatrix}$, whose marginals are gamma with shape parameters 20 and 2 and rate parameters 2 and 1, respectively.
SS signal	Bivariate normal with mean $(10, 2)$ and cov. $\begin{bmatrix} 5 & -1.75 \\ -1.75 & 2 \end{bmatrix}$.
OS noise	Laplace with mean $(0, 0)$ and correl. matrix $\begin{bmatrix} 1 & .5534 \\ .5534 & 1 \end{bmatrix}$.
SS noise	Normal with mean $(0, 0)$ and correl. matrix $\begin{bmatrix} 1 & .5534 \\ .5534 & 1 \end{bmatrix}$.

Both signal distributions have the same mean and covariance matrix and both error distributions have the same covariance matrix. The bivariate gamma is defined here through a Gaussian copula with density is given $f_G(g_1, g_2) = \frac{f_{G_1}(g_1)}{\phi(\Phi^{-1}(F_{G_1}(g_1)))} \cdot \frac{f_{G_2}(g_2)}{\phi(\Phi^{-1}(F_{G_2}(g_2)))} \times f_Z(\Phi^{-1}(F_{G_1}(g_1)), \Phi^{-1}(F_{G_2}(g_2)))$ $g_1, g_2 > 0$ depending on parameters, $\alpha_1, \alpha_2, \beta_1, \beta_2 > 0$ and $-1 < r < 1$. For $i = 1, 2$, F_{G_i} and f_{G_i} denote the distribution and density functions of a gamma variable with shape α_i and rate β_i parameters; Φ^{-1} and ϕ are the quantile function and the density of a standard normal distribution; and f_Z denotes a bivariate normal density with correlation r and standard normal marginal distributions

the signal. With regard to noise to signal ratios, we considered three ratios 0.2, 1, 2 to represent low to high levels of error. These ratios were component-wise fixed for the marginal distributions; that is, for a noise to signal ratio of

Figure 3: Contour plots of the distributions used in the simulation study. Both distributions have the same mean and covariance matrix. Observe that the coordinates are correlated in both distributions. The bivariate gamma was generated according to Table 2

0.2, the first components of the error and signal have a variance ratio of 0.2, and the same holds for the second component. For the kernel density-based approximation to the CF of the errors, we used a normal kernel, as described in Section 3.1, with a diagonal bandwidth matrix obtained with a plug-in method computed using the R package `ks` (see Duong and et al., 2007). In summary, the design factors and the levels of the factors in the simulation study were the following:

1. Four combinations of OS/SS signal/error distributions along with three varying levels of noise to signal variance ratios (0.2, 1, 2), as described further below.
2. Two values for the number of replicates $m = 2$ or $m = 8$ for each individual in a panel data model (2.5) with $n = 150$ individuals.
3. Five kernel functions described in Table 1.
4. Three different approaches (described Section 3.1) for creating “error differences” to approximate the CF of the error.
5. Two different approximations to the characteristic function based on error differences: either the empirical CF from differences or the Fourier transform of a kernel density estimator applied to differences, as described in Section 3.1.

To compute the deconvolution estimates, we used a grid of 64×64 points to approximate the integral in Eq. 2.7 via a fast Fourier transform (as described in Section 3.3) and also for evaluating the density of the signal variable. There was no significant improvement in precision for finer grids. Moreover, the region for the density estimation was fixed at $[3, 18] \times [-2, 9]$, which contains at least 99% of the mass of both OS/SS signal distributions considered (see Fig. 3). For each simulated panel data set and deconvolution estimator, we evaluated the corresponding integrated squared error (ISE) using the approximation

$$\int \left[\hat{f}_{\mathbf{X},n}(\mathbf{x}) - f_{\mathbf{X}}(\mathbf{x}) \right]^2 d\mathbf{x} \approx \Delta \sum_{ij} \left[\hat{f}_{\mathbf{X},n}(\mathbf{x}_{ij}) - f_{\mathbf{X}}(\mathbf{x}_{ij}) \right]^2,$$

where $\Delta = \tau_1 \tau_2 / 64^2$ is the area of every grid rectangle for $\tau_1 = 11$ and $\tau_2 = 15$, and $\mathbf{x}_{ij} = (-2 + i\tau_1, 3 + j\tau_2)$, $i, j \in \{0, \dots, 63\}$, are the grid points. ISEs for all deconvolution estimates were computed over 200 simulation runs per data-generating process, and are summarized in Figs. 4, 5 and 6.

Figure 4: Comparison of estimators for different combinations of smoothness when the noise to signal ratio is equal to 1. Observe that the SS signal and OS noise has the closest performance to the error free case when sinc and flat-top kernels are used. On the other hand, the worst performance overall occurs in the OS signal and SS noise, which is expected. The two background bands represent the three quartiles of the ISE computed for the kernel density estimator of the error free observations (lower band) and the kernel density estimator of the contaminated sample ignoring the noise (upper band). This plot was generated for $m = 2$ replicate observations per individual and noise to signal ratio of one

Finally, we make some comments about bandwidth effects, which we sought to control in order to quantify the impact of the factors mentioned above (e.g., type of kernel) for multivariate panel data with unknown errors. Bandwidth selection is a challenging problem in kernel deconvolution, with a rich literature for the univariate case (see Delaigle and Gijbels, 2004a), but far less is known for the multivariate case. To minimize the confounding

Figure 5: Comparison of estimators for different noise-to-signal ratio and number of observations per individuals when the empirical characteristic function has been used for approximating the error with SS signal and OS noise. The two background bands represent the three quartiles of the ISE computed for the kernel density estimator of the error free observations (lower band) and the kernel density estimator of the contaminated sample ignoring the noise (upper band)

effect of bandwidth choice, we used the diagonal bandwidth matrix for minimizing the integrated squared error (ISE) for deconvolution kernel density estimators based on known (not estimated) measurement and error CFs. That is, we employ a diagonal bandwidth matrix for each kernel by choosing the best approximate bandwidths for these kernels, but we do not assume that the distributions of the errors or the signal are known in the simulation study.

4.2. Simulation Results As a frame of reference, we adopted two “gold standard” or reference estimators against which we compare our results.

Figure 6: Estimation performance based on pre-processing mechanism used to compute the differences among $m = 8$ replicates per individual (for approximating the characteristic function of noise). The left plot considers a noise to signal ratio of 1 and the right plot involves a ratio of 2. The two background bands represent the three quartiles of the ISE computed for the kernel density estimator of the error free observations (lower band) and the kernel density estimator of the contaminated sample ignoring the noise (upper band). This plot was generated using the empirical characteristic function to estimate the error (results were similar the characteristic function of a kernel density estimator, cf. Eq. 3.8, Section 3.1) with OS signal/SS noise data generation

These reference estimators are obtained from kernel density estimators (without deconvolution) applied to samples defined by either Eqs.4.1 or 4.2 as

$$\mathbf{X}_1, \dots, \mathbf{X}_n \quad (\text{Error free}) \quad (4.1)$$

$$\bar{\mathbf{Y}}_1, \dots, \bar{\mathbf{Y}}_n \quad (\text{Contaminated}) \quad (4.2)$$

using the normal kernel and a plug-in bandwidth matrix (from the R package **decon**). The first estimation reference, which serves as an approximate lower bound for the estimation error (*error-free* in plots), is the usual bivariate kernel density estimator applied to the sample $\{\mathbf{X}_i\}_{i=1}^n$ without measurement error, which is unobservable in practice. Ideally, the ISE of a deconvolution estimator should be close to the ISE of this reference estimator, which can be generally expected to exhibit better performance than estimators based on data with true errors in variables. The second reference estimator (*contaminated* in plots) provides an upper bound on the estimation error and corresponds to the usual bivariate kernel density estimator applied to the measurement averages for each individual $\{\bar{\mathbf{Y}}_i\}_{i=1}^n$ from Eq. 2.6. Note that

this reference estimator *ignores* measurement errors inherent to the averages (4.2). Hence, as the purpose of kernel deconvolution is precisely to account for measurement error, it would be ideal for deconvolution density estimators to have smaller ISEs than this reference.

Figure 4 provides a comparison of average ISEs for the different kernel choices in deconvolution estimation with panel data across the four combinations of ordinary smooth/supersmooth (OS/SS) signal and OS/SS noise (with $m = 2$ replicates per individual and noise to signal ratio of one). The figure also considers the type of estimation for the error/noise characteristic function, based on either the empirical characteristic function (ecf) or the characteristic function of a kernel density estimator (kde) (formed by differences between measurement replicates). As expected, SS error distributions tend to induce larger ISEs on average compared to OS errors and the smoothness of the underlying signal also impacts performance of deconvolution estimators. However, the effect of the kernel choice is qualitatively similar in Fig. 4, where the sinc kernel has the best performance closely followed by the flat-top kernel in every scenario. The estimation approach for the error characteristic function (i.e., ecf or kde) typically resulted in only small differences in performance, with a small but consistent edge for the kde-based method. Note also that deconvolution estimators often substantially improve upon the contaminated reference in Fig. 4 (upper horizontal bar) which ignores measurement error in density estimation, but as expected perform worse than the estimation reference when no measurement error is present (lower horizontal bar).

A similar pattern emerges in Fig. 5, which considers the effect of both differing noise to signal ratios and varying amounts of replication ($m = 2, 8$). From the figure, the expected estimation error (ISE) decreases when the number of replicates grows from $m = 2$ to $m = 8$, and errors also increase as the noise to signal ratio increases. However, regardless of these factors, the results are again qualitatively similar across types of kernel choices, and the flat-top and sinc kernels emerge as the best performers. In particular, with higher numbers of replicates and low noise to signal ratios (ideal inference situations with panel data), deconvolution estimators based on flat-top and sinc kernels achieve error rates similar to the oracle type reference kernel estimation based on observations with no measurement error (lower horizontal bar in Fig. 5), which also supports these kernel choices. Again, ignoring measurement error (simply averaging observations over replicates and applying ordinary kernel density estimation techniques) results in larger estimation errors compared to deconvolution approaches (upper horizontal bar in Fig. 5), as is evident in particular for small replication $m = 2$.

We mentioned in Section 3.1 that there are several strategies to obtain differences to approximate the characteristic function of the error. Specifically, taking all the possible difference of observations for the same individual (cf. Eq. 3.2), which leads to many correlated observations of differences of errors; taking differences of paired columns, obtaining fewer but independent error differences (cf. Eq. 3.3), and everything minus the first repetition for individual, which is a compromise between number of observations and independence (cf. Eq. 3.1). Figure 6 compares performances for deconvolution estimators based on this type of pre-processing used for creating error differences from replicate measurements in the panel data structure. This illustration considers the $m = 8$ replicate case (as the type of differencing is not an issue with $m = 2$ replications). Essentially from this figure, the type of pre-processing has little effect, even as the noise ratio increases.

4.3. Kernel Choices Among the previous simulation findings, the choice of kernel emerges as the single most influential factor for the performance of deconvolution estimators to be considered in application. While the sinc kernel (sinc) has been widely used in the literature for univariate deconvolution and also exhibited consistently good performance in our simulations, the flat-top kernel emerged as a competitor, with a performance similar to the sinc kernel in terms of average ISE. In fact, in many simulation cases, the mean ISEs of deconvolution estimators from both sinc and flat-top kernels were close to the reference estimator based on error free samples. However, an additional potential advantage of the flat-top kernel is that this tends to produce density estimators without the “wiggly” (and sometimes heavily negative) tails associated with the sinc kernel. Such artifacts are a well-known and undesirable feature of sinc kernel-based estimates, which becomes a compounding issue in the multivariate setting. See Delaigle and Hall (2006) for a discussion on this issue.

To illustrate, Fig. 7 presents the deconvolution density estimates based on different kernel types, produced from a simulated data set of OS noise and OS signal with a noise to signal ratio of 0.2, two replicates per individual ($m = 2$) and 150 individuals ($n = 150$). From the figure, the sinc kernel yields an estimate with the most pronounced oscillations, leading to a substantial negative part in the density estimate unlike the other kernel approaches shown. In contrast, the flat-top kernel provides a density estimate with more stable tails and fewer negative artifacts. While less prominent, some negative portions in the density estimate do still appear with the flat-top kernel in Fig. 7, but it would be difficult to do much better. For comparison and

Figure 7: Perspective plots of the true density and kernel deconvolution density estimates based on a sample size of $n = 150$ varying the kernel function. The sinc kernel (upper left corner) possess the greatest oscillations

perspective, among the kernels in Fig. 7, only the de la Vallée-Poussin kernel (VP) produces a non-negative density estimate. However, from the previous simulations with deconvolution estimators, the VP kernel also resulted in the worst performances in terms of average ISE, sometimes no better than the contaminated reference estimator that ignored measurement error. In this sense, the flat-top kernel in multivariate deconvolution estimation may lead to comparable performances in terms of ISE relative to the sinc kernel (being the best among all the kernel functions considered here), but with more attractive properties in the shape and appearance of its density estimator.

5 Density of the Log-Ratio of Sodium and Potassium

In this section, we illustrate the methods that we have discussed by applying these to a problem in public health. Here the problem of interest is to quantify the proportion of adults in the United States who are meeting current medical recommendations regarding ratio of sodium to potassium intake. High usual intake of sodium is associated with high blood pressure. Potassium tends to have the opposite effect, so recent research suggests that the usual ratio of sodium to potassium intake may be a better biomarker for high blood pressure than sodium intake itself (NRC, 2019).

We use intake information collected by the National Health and Nutrition Examination Survey (NHANES): 2009-2010, 2011-2012, 2013-2014. NHANES is a representative survey of the non-institutionalized population in the contiguous United States. We restrict the analysis to adults aged 19 to 50 years. Across all three survey waves, there were 8823 subjects 19-50 years of age with at least one measurement of sodium and potassium daily intake. For 7662 of them, there was a second measurement, obtained a few days after the first. In all, 1161 participants have a single observation of daily intake. In NHANES, participants are asked to report their food and beverage consumption during the previous 24 hours. Because a person's intake varies from day to day, daily intakes are noisy measurements of the person's usual or long-run average intake of a nutrient. The data are in a format similar to the panel data structure described in Eq. 2.5.

One approach to estimate the density of the usual log-ratio of sodium and potassium intake is to first obtain the joint density of usual log-sodium and log-potassium and then numerically derive the density of the difference. We applied KDDE to achieve this, as explained in the following.

Let $\mathbf{Y}_{ij}$ denote the bivariate observation of the i -th subject in the j -th day ($j = 1$ or 2). The entries are log-sodium (first) and log-potassium (last).

Figure 8: Scatterplot of the difference of log-sodium (day 2 - day 1) and the difference of log-potassium (day 2 - day 1). A regular kernel density estimator has been overlaid

Since the number of observations varied by subject, we used a weighted empirical characteristic function of the form

$$\hat{\phi}_{\mathbf{Y}}(\mathbf{t}) = \frac{1}{n} \sum_{i=1}^n \frac{1}{m_i} \sum_{j=1}^{m_i} \exp(it\mathbf{Y}_{ij}), \quad (5.1)$$

where $m_i = 1, 2$. In order to model the measurement error, we took the difference of the observations for those subjects with two days of measurements. As it can be appreciated in Fig. 8, the errors are highly correlated. We therefore, opted for using the empirical characteristic function described in Eq. 3.6.

We used the flat-top kernel (see Table 1) to estimate the distribution of the errors, and approximated the Riemann sums in Eq. 3.12 using a regular grid with resolution of 256×256 via the FFT method. We then applied a bootstrap approach to select the bandwidth h . Specifically, for every bandwidth value, we generated 100 bootstrap samples from the KDDE with that bandwidth. Then, we computed the KDDE for each of the bootstrap samples and used these to approximate the mean integrated squared error (MISE),

Figure 9: KDDE of log-sodium and log-potassium

based on those 100 samples. We selected the bandwidth that minimized the bootstrap-approximated MISE. The resulting value for the bandwidth was $h = 0.135$.

In Fig. 9, we present the estimated joint density of log-sodium and log-potassium.

If Y_1 and Y_2 are two random variables with joint density f , then the density of $Y_1 - Y_2$, say, g , is equal to

$$g(z) = \int_{-\infty}^{\infty} f(z + y, y) dy, \quad (5.2)$$

which is derived via an application of the Leibniz integral rule. Consequently, we can obtain an estimate of the log-ratio difference of the two random variables if we have an estimate of their joint density. Therefore, after computing the KDDE of the joint distribution of log sodium and log potassium, we applied (5.2) to obtain an estimated density of the log-ratio of these variables. The integral (5.2) was approximated with Riemann sums, and the resulting estimator is shown in Fig. 10.

For adults, the recommended intake of sodium is 2,300 mg/day and for potassium it is 4,800 mg/day (see, e.g., NRC, 2019), for a maximum recommended ratio of about 0.48, or a log-ratio of about -0.32. From Fig. 10, we see that the proportion of adults in the United States whose long-run average ratio of sodium to potassium intake meets recommendations is about 4%.

Figure 10: Density of the log-ratio of sodium and potassium

This is computed as the area under the curve in Fig. 10 to the left of -0.32. This estimated prevalence is consistent with what is known about excessive consumption of sodium among adults in the United States.

6 Conclusions and Promising Research Avenues

We examined the performance of several different methods for deconvolution density estimation with bivariate panel data (repeated measurements for each individual contaminated by measurement error). While deconvolution estimation has received much attention for univariate data, there has been less consideration for multivariate data and, in particular, for panel data structures which arise in scientific studies. For such data, we examined the impact of several choices available to the practitioner when implementing a density estimator for the underlying variable of interest. These include the type of kernel, the approach for processing replicate measurements in order to estimate error distributions, and the type of estimator of the error characteristic function. One general observation is that the performance of deconvolution density estimators can be improved by using kernel density estimators (and their characteristic functions) formed from error terms rather than the empirical characteristic functions from errors; yet the latter is most commonly proposed in the literature. However, among possibilities considered, we found that the most important and substantial feature impacting bivariate kernel deconvolution is the choice of kernel function. Traditionally, the sinc kernel has been popular for univariate deconvolution, because its Fourier transform is very simple and exhibits good properties in terms of integrated squared error; we show that those features continue to be important

for bivariate panel data. However, we also found that flat-top kernels result in similar performance to the sinc kernel in terms of estimation error, while producing bivariate density estimates with more attractive geometries and fewer wavy artifacts. We have also described numerical integration recipes based on FFTs in order to quickly and practically implement deconvolution estimators in the multivariate data setting; R software and illustrations of implementation are available in a Github repository in Basulto-Elias (2016).

Theoretically, the methods we discussed in this paper apply for performing deconvolution in higher dimensions. In practice, however, such methods require numerical integration which becomes computationally difficult in high dimensions. For example, when using Riemann sums on a grid to numerically approximate these types of integrals, the number of points required in the grid increases exponentially with the sampling dimension (see Basulto-Elias et al., 2017 for a simple time comparison between univariate and bivariate Fourier integrals). Furthermore, such numerical Riemann sums are prone to error, even more so when the number of grid points is not large enough. Therefore in practice, approximating integrals in more than two or three dimensions is challenging.

Open research issues remain mostly associated with the development of theory for multivariate deconvolution estimation with panel data. Two specific open problems are determining rates of convergence of the kernel density estimators and defining optimal bandwidth choices. A related and important issue for investigation concerns data-driven approaches for selecting bandwidths that are suitable for panel data in the multivariate case.

Acknowledgments. The authors are grateful to two reviewers and the editors for thoughtful comments and suggestions which improved the manuscript.

Open Access. This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- BAILEY, D. H. and SWARZTRAUBER, P. N. (1994). A fast method for the numerical evaluation of continuous fourier and laplace transforms. *SIAM J. Sci. Comput.* **15**, 1105–1110.
- BASULTO-ELIAS, G. (2016). kerdec: an R package for deconvolution methods. Available at <https://github.com/gbasulto/kerdec>.
- BASULTO-ELIAS, G., CARRIQUIRY, A., BRABANTER, K. D. and NORDMAN, D. J. (2017). Fourierin: An R package to compute Fourier integrals. *The R Journal* **9**, 72–83.
- COMTE, F. and LACOUR, C. (2010). Pointwise deconvolution with unknown error distribution. *Comptes Rendus Mathématique* **348**, 323–326.
- COMTE, F. and LACOUR, C. (2011). Data-driven density estimation in the presence of additive noise with unknown distribution. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)* **73**, 601–627.
- COMTE, F., SAMSON, A. and STIRNEMANN, J. J. (2013). Deconvolution estimation of onset of pregnancy with replicate observations. *Scandinavian Journal of Statistics*.
- COMTE, F., LACOUR, C. et al. (2013). Anisotropic adaptive kernel deconvolution. In: *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques*, 49, Institut Henri Poincaré, 569–609.
- DELAIGLE, A. and GIJBELS, I. (2004a). Practical bandwidth selection in deconvolution kernel density estimation. *Computational Statistics & Data Analysis* **45**, 249–267.
- DELAIGLE, A. and GIJBELS, I. (2004b). Bootstrap bandwidth selection in kernel density estimation from a contaminated sample. *Ann. Inst. Stat. Math.* **56**, 19–47.
- DELAIGLE, A. and HALL, P. (2006). On optimal kernel choice for deconvolution. *Statistics & Probability Letters* **76**, 1594–1602.
- DELAIGLE, A. and MEISTER, A. (2008). Density estimation with heteroscedastic error. *Bernoulli* **14**, 562–579.
- DEVROYE, L. (1989). Consistent deconvolution in density estimation. *Canadian Journal of Statistics* **17**, 235–239.
- DUONG, T. and ET AL. (2007). ks: Kernel density estimation and kernel discriminant analysis for multivariate data in r. *J. Stat. Softw.* **21**, 1–16.
- FAN, J. (1991). On the optimal rates of convergence for nonparametric deconvolution problems. *Ann. Stat.*, 1257–1272.
- HAHN, T. (2005). Cuba-a library for multidimensional numerical integration. *Comput. Phys. Commun.* **168**, 78–95.
- INVERARITY, G. (2002). Fast computation of multidimensional fourier integrals. *SIAM J. Sci. Comput.* **24**, 645–651.
- JOHANNES, J. (2009). Deconvolution with unknown error distribution. *Ann. Stat.* **37**, 2301–2323.
- MASRY, E. (1991). Multivariate probability density deconvolution for stationary random processes. *IEEE Transactions on Information Theory* **37**, 1105–1115.
- MASRY, E. (1993a). Asymptotic normality for deconvolution estimators of multivariate densities of stationary processes. *Journal of Multivariate Analysis* **44**, 47–68.
- MASRY, E. (1993b). Strong consistency and rates for deconvolution of multivariate densities of stationary processes. *Stochastic Processes and their Applications* **47**, 53–74.
- MEISTER, A. (2009). Deconvolution problems in nonparametric statistics, volume 193. Springer.
- NEUMANN, M. H. and HÖSSJER, O. (1997). On the effect of estimating the error density in nonparametric deconvolution. *Journal of Nonparametric Statistics* **7**, 307–330.

- NRC (2019). *Dietary Reference Intakes for Sodium and Potassium*. National Academy Press, Washington.
- POLITIS, D. N. (2003). Adaptive bandwidth choice. *Journal of Nonparametric Statistics* **15**, 517–533.
- POLITIS, D. N. and ROMANO, J. P. (1995). Bias-corrected nonparametric spectral estimation. *Journal of Time Series Analysis* **16**, 67–103.
- POLITIS, D. N. and ROMANO, J. P. (1999). Multivariate density estimation with general flat-top kernels of infinite order. *J. Multivar. Anal.* **68**, 1–25.
- STIRNEMANN, J. J., COMTE, F. and SAMSON, A. (2012). Density estimation of a biomedical variable subject to measurement error using an auxiliary set of replicate observations. *Statistics in Medicine* **31**, 4154–4163.
- WAND, M. P. and JONES, M. C. (1995). *Kernel Smoothing*, 60. Chapman & hall/CRC, London.
- YOUNDJÉ, É. and WELLS, M. T. (2008). Optimal bandwidth selection for multivariate kernel deconvolution density estimation. *Test* **17**, 138–162.

Publisher's Note. Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

GUILLERMO BASULTO-ELIAS
ALICIA L. CARRIQUIRY
KRIS DE BRABANTER
DANIEL J. NORDMAN
IOWA STATE UNIVERSITY,
AMES, IA, USA
E-mail: basulto@iastate.edu

Paper received: 6 December 2019.