Density Estimation with Replicate Heteroscedastic Measurements

Julie McIntyre · Leonard A. Stefanski

Abstract   We present a deconvolution estimator for the density function of a random variable from a set of independent replicate measurements. We assume that measurements are made with normally distributed errors having unknown and possibly heterogeneous variances. The estimator generalizes well-known deconvoluting kernel density estimators, with error variances estimated from the replicate observations. We derive expressions for the integrated mean squared error and examine its rate of convergence as $n \to \infty$ and the number of replicates is fixed. We investigate the finite-sample performance of the estimator through a simulation study and an application to real data.

Keywords   Bandwidth · Bootstrap · Deconvolution · Hypergeometric series · Measurement error

1 Introduction

We consider estimating the density function of an unobservable random variable from a set of replicate measurements. Let X have unknown density function f_x. Suppose that $X_1, \ldots, X_n$ are measured independently and repeatedly as $\{W_{r,j}\}_{r=1, j=1}^{n, \ m_r}$, where $W_{r,j} = X_r + U_{r,j}$ and $m_r \geq 2$. Then the density of the observed $W_{r,j}$ is related to f_x by the convolution, $f_{W_{r,j}} = f_x * f_{U_{r,j}}$. We present an estimator for f_x under the assumption that $U_{r,j} \sim N(0, \ \sigma_r^2)$, independent of X_r, $r = 1, \ldots, n$ and $j = 1, \ldots, m_r$.

J. McIntyre (✉)
Department of Mathematics and Statistics, University of Alaska Fairbanks, Fairbanks, AK 99775, USA
e-mail: ffjpm@uaf.edu

L. A. Stefanski
Department of Statistics, North Carolina State University, Raleigh, NC 27695-8203, USA
e-mail: stefansk@stat.ncsu.edu

We do not model the error variances σ_r^2, and thus our methods are completely general. We note that in applications where reliable models for variances (say, as functions of the means) can be specified, more efficient estimators can likely be obtained.

Several authors have studied estimation of f_x when measurement errors are identically distributed with a known density function. The approach we take is closely related to that of Stefanski and Carroll (1990), who presented deconvoluting kernel density estimators appropriate for a wide class of error distributions. The properties of deconvoluting estimators have been studied in detail (see, for example, Carroll and Hall 1988; Devroye 1989; Stefanski 1990; Fan 1991a,b, 1992; Wand et al. 1998). Recently, authors have studied deconvolution under less restrictive assumptions. Estimating f_x when errors are identically distributed with an unknown density function has been addressed by Diggle and Hall (1993), Patil (1996), Li and Vuong (1998), Meister (2006), Neumann (2007) and Delaigle et al. (2007, 2008), among others.

Fewer authors have addressed deconvolution in the presence of heteroscedastic measurement error. Staudenmayer et al. (2008) assume the availability of replicate measurements and model measurement error variances as functions of the unobserved data. The density function f_x is modeled as a convex mixture of B-spline densities. Delaigle and Meister (2007, 2008) present a generalization to the deconvoluting kernel estimator for the case where measurement errors are heteroscedastic and have known distributions. The characteristic function for the measurement error density is described by a weighted average of characteristic functions. Additionally they discuss estimation when error densities are imperfectly known but estimable, for example, from replicate measurements. Under certain conditions, the known error densities may be replaced with their estimates without affecting convergence properties.

We assume a fixed number of replicate measurements and approach the problem via the conditional distributions of the sample means and variances of these measurements. A natural estimator arises from the results of Stefanski et al. (2005), and generalizes the deconvoluting estimator of Stefanski and Carroll (1990). Measurement errors are assumed normal, but not identically distributed. The estimator accommodates heteroscedastic and homoscedastic measurement errors. However as the main focus of this paper is estimation in the presence of heteroscedastic error, a detailed examination of our estimator's properties is restricted to this case.

We start with a brief review of the results of Stefanski et al. (2005). In Sect. 2, we define our estimator and show its connection to the Stefanski and Carroll (1990) estimator. We derive and examine the asymptotic mean integrated squared error in Sect. 3, and in Sect. 4 we investigate finite-sample properties via simulation. We address bandwidth estimation in Sect. 5, describing a bootstrap bandwidth selection procedures. In Sect. 6 we present an application to real data.

1.1 Background

Our estimator of f_x uses results of Stefanski et al. (2005), who presented a method for constructing unbiased estimators of $g(\mu)$ where μ is the mean of a normally distributed random variable and $g()$ is an entire function over the complex plane. For reference, we present the following theorem, without proof, from Stefanski et al. (2005).

Theorem 1.1.1 *Suppose $\widehat{\mu}$ and $\widehat{\sigma}^2$ are independent random variables such that $\widehat{\mu} \sim N\left(\mu,\ \tau\sigma^2\right)$ and $(d\widehat{\sigma}^2/\sigma^2) \sim$ Chi-Squared(d). Define $T = Z_1/\sqrt{Z_1^2 + \cdots + Z_d^2}$, where $Z_1, \ldots, Z_d$ are independent and identically distributed $N(0,\ 1)$ random variables, and let $i = \sqrt{-1}$. Let $g()$ be an entire function, so that $g()$ has a series expansion defined at each point in the complex plane, and suppose that the interchange of expectation and summation in this expansion is justified. Then the estimator,*

$$\widehat{\theta} = E\left\{ g\left(\widehat{\mu} + i(\tau d)^{1/2}\widehat{\sigma}T\right) \middle| \widehat{\mu}, \widehat{\sigma}^2 \right\} \tag{1}$$

is uniformly minimum variance unbiased for $g(\mu)$ provided Var$\left(\widehat{\theta}\right) < \infty$.

A proof and discussion may be found in Stefanski et al. (2005) and Stefanski (1989). However, a few comments about the estimator $\widehat{\theta}$ are in order. Although $g()$ is complex-valued, $\widehat{\theta}$ is real-valued, because the imaginary part of $g()$ has expectation 0. For most functions $g()$, however, finding a closed-form expression for the estimator in (1) is difficult. A Monte Carlo approximation to $\widehat{\theta}$ is given by

$$\widehat{\theta}_B = \frac{1}{B}\sum_{b=1}^{B}\text{Re}\left\{ g\left(\widehat{\mu} + i(\tau d)^{1/2}\widehat{\sigma}T_b\right) \right\}, \tag{2}$$

where $T_1, \ldots, T_B$ are independent replicates of the variable T. Variances of (1) and (2) are real-valued, but generally do not have closed-form expressions. However, Monte Carlo methods can be used to estimate both as described in Stefanski et al. (2005).

2 The deconvolution estimator

2.1 Heteroscedastic measurement errors

The estimators in (1) and (2) can be used to estimate the density function of a random variable X that is measured with normally distributed error having unknown, nonconstant variance, provided each value of X is measured at least twice, and that the measurement error variance is constant among replicates but may differ between replicates. Suppose that the variables $X_1, \ldots, X_n$ are measured independently and repeatedly as $\{W_{r,j}\}_{r=1, j=1}^{n, m_r}$, $m_r \geq 2$, where for each $r = 1, \ldots, n$ and $j = 1, \ldots, m_r$,

$$W_{r,j} = X_r + U_{r,j}, \qquad U_{r,j} \sim \text{iid N}(0, \sigma_r^2), \tag{3}$$

$U_{r,j}$ is independent of X_r and σ_r^2 is unknown. Define the sample moments $\overline{W}_r = m_r^{-1}\sum_{j=1}^{m_r} W_{r,j}$ and $\widehat{\sigma}_r^2 = (m_r - 1)^{-1}\sum_{j=1}^{m_r}(W_{r,j} - \overline{W}_r)^2$. Then conditioned on X_r, the following are true: (1) $\overline{W}_r$ and $\widehat{\sigma}_r^2$ are independent random variables, (2) $\overline{W}_r \sim \text{N}\left(X_r,\ \sigma_r^2/m_r\right)$, and (3) $(m_r - 1)\widehat{\sigma}_r^2/\sigma_r^2 \sim$ Chi-squared$(m_r - 1)$. Let $Q(x)$ be a probability density function that is entire over the complex plane, and for each

$r = 1, \ldots, n$ define $T_{r,m_r-1} = Z_{r,1} / \sqrt{Z_{r,1}^2 + \cdots + Z_{r,m_r-1}^2}$, where the $Z_{r,j}$ are independent $N(0, 1)$ random variables. It follows from Theorem 1.1.1 that the estimator

$$\widehat{f}_{\text{het}}(x) = \frac{1}{n\lambda} \sum_{r=1}^{n} E \left\{ Q \left(\frac{x - \left[\overline{W}_r + i(\frac{m_r-1}{m_r})^{1/2} \widehat{\sigma}_r T_{r,m_r-1} \right]}{\lambda} \right) \Bigg| \overline{W}_r, \widehat{\sigma}_r^2 \right\} \quad (4)$$

is such that

$$E \left\{ \widehat{f}_{\text{het}}(x) | X_1, \ldots, X_n \right\} = \frac{1}{n\lambda} \sum_{r=1}^{n} Q \left(\frac{x - X_r}{\lambda} \right). \quad (5)$$

The right hand side of (5) is a kernel density estimator of $f_x(x)$, with bandwidth $\lambda > 0$. Thus it follows that unconditionally, $\widehat{f}_{\text{het}}(x)$ has the same expectation and same bias as the kernel density estimator constructed from the true, unobserved data.

Note that Theorem 1.1.1 also applies to the case where the measurement error variance in (3) is constant and data are pooled to estimate the common variance σ^2. Let $\widehat{\sigma}^2 = d^{-1} \sum_{r=1}^{n} (m_r - 1) \widehat{\sigma}_r^2$ be the pooled estimator of the measurement error variance based on $d = \sum_{r=1}^{n} (m_r - 1)$ degrees of freedom, and let $T_{r,d} = Z_{r,1} / \sqrt{Z_{r,1}^2 + \cdots + Z_{r,d}^2}$ where the $Z_{r,j}$ are independent $N(0, 1)$ random variables. It follows from Theorem 1.1.1 that the estimator

$$\widehat{f}_{\text{hom}}(x) = \frac{1}{n\lambda} \sum_{r=1}^{n} E \left\{ Q \left(\frac{x - \left[\overline{W}_r + i(\frac{d}{m_r})^{1/2} \widehat{\sigma} T_{r,d} \right]}{\lambda} \right) \Bigg| \overline{W}_r, \widehat{\sigma}^2 \right\} \quad (6)$$

possesses the key property in (5), i.e., that conditioned on the true data, it is an unbiased estimator of a true-data, kernel density estimator.

2.1.1 Connection to the deconvoluting kernel density estimator

The estimators in (4) and (6) generalize the deconvoluting kernel density estimator of Stefanski and Carroll (1990). They considered the case with data $\{W_r\}_{r=1}^{n}$, such that $W_r = X_r + U_r$, and $\{U_r\}_{r=1}^{n}$ are independent and identically distributed with known characteristic function $\Phi_u(t) = E(e^{itu})$, and are independent of $\{X_r\}_{r=1}^{n}$. Their estimator is based on Fourier inversion of the empirical characteristic function of the kernel density estimator of the observed-data density, $f_w(w)$. Let $Q(x)$ be a probability density function whose extension to the complex plane is entire throughout the complex plane, and denote the characteristic function of $Q(x)$ by $\Phi_Q(t)$. For the case of $N(0, \sigma^2)$ measurement errors, the deconvoluting estimator is given by

$$\widehat{f}_{\text{sc}}(x) = \frac{1}{n\lambda} \sum_{r=1}^{n} Q_d \left(\frac{x - W_r}{\lambda}, \lambda, \sigma \right), \quad (7)$$

where $Q_d()$ is the deconvoluting kernel defined as

$$Q_d(z, \lambda, \sigma) = \frac{1}{2\pi} \int e^{-itz} e^{t^2\sigma^2/2\lambda^2} \Phi_Q(t) \mathrm{d}t. \tag{8}$$

We now establish the connection between $\widehat{f}_{sc}(x)$ and our estimator under heteroscedastic errors, $\widehat{f}_{het}(x)$. Arguments similar to those presented below also apply to the homoscedastic-error estimator $\widehat{f}_{hom}(x)$. Consider again $\widehat{f}_{het}(x)$ in (4). Define

$$Q_r^*(z, \lambda, m_r, \widehat{\sigma}_r) = E\left\{ Q\left(z - \frac{i \left(\frac{m_r-1}{m_r}\right)^{1/2} \widehat{\sigma}_r T_{r,m_r-1}}{\lambda} \right) \middle| \overline{W}_r, \widehat{\sigma}_r^2 \right\}, \tag{9}$$

and note that $Q_r^*(z, \lambda, m_r, \widehat{\sigma}_r)$ is real-valued. With this definition we write

$$\widehat{f}_{het}(x) = \frac{1}{n\lambda} \sum_{r=1}^n Q_r^*\left(\frac{x - \overline{W}_r}{\lambda}, \lambda, m_r, \widehat{\sigma}_r \right). \tag{10}$$

The connection between the estimators $\widehat{f}_{het}(x)$ in (10) and $\widehat{f}_{sc}(x)$ in (7) is made clear through examination of the functions $Q_r^*()$ and $Q_d()$. For each $r = 1, \ldots, n$, the sample mean $\overline{W}_r$ measures X_r with a normally distributed measurement error that has variance σ_r^2/m_r. The deconvoluting kernel in Eq. 8 depends on the *assumed known* inverse of the measurement-error characteristic function $\Phi_{u_r}^{-1}(t) = e^{t^2\sigma_r^2/2m_r\lambda^2}$. However, when σ_r^2 is unknown, so too is $\Phi_{u_r}^{-1}(t)$. As we show next, $Q_r^*(z, \lambda, m_r, \widehat{\sigma}_r)$ unbiasedly estimates $Q_d(z, \lambda, \sigma_r^2/m_r)$ through unbiased estimation of $\Phi_{u_r}^{-1}(t)$.

First, applying the Fourier inversion formula and interchanging the operations of integration and expectation in (9) yields

$$Q_r^*(z, \lambda, m_r, \widehat{\sigma}_r)$$
$$= \frac{1}{2\pi} \int e^{-itz} E\left\{ \exp\left(\frac{-t}{\lambda} \left(\frac{m_r-1}{m_r}\right)^{1/2} \widehat{\sigma}_r T_{r,m_r-1} \right) \middle| \overline{W}_r, \widehat{\sigma}_r^2 \right\} \Phi_Q(t) \mathrm{d}t. \tag{11}$$

Because T_{r,m_r-1} is independent of the data, the conditional expectation in (11) is the characteristic function of T_{r,m_r-1} evaluated at the argument $\theta(t) = it\{(m_r - 1)/m_r\}^{1/2}\widehat{\sigma}_r/\lambda$. By construction, $T_{r,m_r-1} = \boldsymbol{T}^T \boldsymbol{e}_1$, where $\boldsymbol{T}^T$ is a random vector uniformly distributed on the $(m_r - 1)$ dimensional unit sphere, and $\boldsymbol{e}_1$ is the $(m_r - 1) \times 1$ dimensional unit vector having a one in the first position. The characteristic function of $\boldsymbol{T}^T \boldsymbol{e}_1$ is given by (Watson 1983)

$$\Phi_T(\theta(t)) = \Gamma\left(\frac{m_r-1}{2} \right) J_{\frac{m_r-1}{2}-1}(\theta(t)) \left(\frac{\theta(t)}{2} \right)^{1-\frac{m_r-1}{2}}. \tag{12}$$

Here $J_v(z)$ denotes a Bessel function of the first kind,

$$J_v(z) = \sum_{k=0}^{\infty} \frac{(-1)^k}{k!\,\Gamma(v+k+1)} \left(\frac{z}{2}\right)^{v+2k}.$$

Substituting Eq. 12 into Eq. 11 yields

$$
\begin{aligned}
Q_r^*(z, \lambda, m_r, \widehat{\sigma}_r) &= \frac{1}{2\pi}\int e^{-itz}\Gamma\left(\frac{m_r-1}{2}\right) J_{\frac{m_r-1}{2}-1}(\theta(t)) \left\{\frac{\theta(t)}{2}\right\}^{1-\frac{m_r-1}{2}} \Phi_Q(t)dt \\
&= \frac{1}{2\pi}\int e^{-itz}\Phi_Q(t)\widehat{\Phi}_{u_r}^{-1}(t/\lambda)dt,
\end{aligned}
\tag{13}
$$

where

$$
\begin{aligned}
\widehat{\Phi}_{u_r}^{-1}(t/\lambda) &= \Gamma\left(\frac{m_r-1}{2}\right) J_{\frac{m_r-1}{2}-1}(\theta(t)) \left\{\frac{\theta(t)}{2}\right\}^{1-\frac{m_r-1}{2}} \\
&= \sum_{k=0}^{\infty} \frac{\left(\frac{t^2}{4\lambda^2}\frac{m_r-1}{m_r}\widehat{\sigma}_r^2\right)^k \Gamma\left(\frac{m_r-1}{2}\right)}{k!\,\Gamma\left(k+\frac{m_r-1}{2}\right)}.
\end{aligned}
\tag{14}
$$

The only random quantity in $Q_r^*(z, \lambda, m_r, \widehat{\sigma}_r)$ is the sample variance, $\widehat{\sigma}_r^2$, which appears in $\widehat{\Phi}_{u_r}^{-1}(t/\lambda)$. From the relationship

$$E\{\widehat{\sigma}_r^{2k}\} = \frac{\Gamma\left(k+\frac{m_r-1}{2}\right)\left(\frac{2\sigma_r^2}{m_r-1}\right)^k}{\Gamma\left(\frac{m_r-1}{2}\right)},\tag{15}$$

it follows that $E\left\{\widehat{\Phi}_{u_r}^{-1}(t/\lambda)\right\} = e^{t^2\sigma_r^2/2m_r\lambda^2}$, the inverse of the characteristic function of a N$\left(0,\ \sigma_r^2/m_r\right)$ random variable evaluated at t/λ. Substitution into (13) yields

$$E\{Q_r^*(z, \lambda, m_r, \widehat{\sigma}_r)\} = \frac{1}{2\pi}\int e^{-itz}e^{t^2\sigma_r^2/2m_r\lambda^2}\Phi_Q(t)dt.$$

Comparison with (8) reveals that $Q_r^*(z, \lambda, m_r, \widehat{\sigma}_r)$ is unbiased for $Q_d(z, \lambda, \sigma_r^2/m_r)$.

2.2 Monte Carlo estimation

In most cases, the estimators in (4) and (6) do not have simple closed-form expressions. However, Monte Carlo approximations to the conditional expectations in Eqs. 4 and 6 follow directly from (2). Consider the expression for $\widehat{f}_{het}(x)$ in (4). For each $r = 1, \ldots, n$, generating $T_{r,1}, \ldots, T_{r,B}$ as independent replicates of $T_r = Z_{r,1}/\sqrt{Z_{r,1}^2 + \cdots + Z_{r,m_r-1}^2}$ yields an approximation to $\widehat{f}_{het}(x)$,

$$\widehat{f}_{B,\text{het}}(x) = \frac{1}{n\lambda} \sum_{r=1}^{n} \frac{1}{B} \sum_{b=1}^{B} \text{Re} \left\{ Q\left(\frac{x - \left\{ \overline{W}_r + \text{i}\left(\frac{m_r-1}{m_r}\right)^{1/2} \widehat{\sigma}_r T_{r,b} \right\}}{\lambda} \right) \right\}.$$

Note that when $m_r = 2$, T_{r,m_r-1} in (4) is equal to either 1 or -1, each with probability 1/2. Thus in the special case of two replicate measurements, the conditional expectation in (4) can be evaluated without Monte Carlo approximation, and is

$$\widehat{f}_{\text{het}}(x) = \frac{1}{2n\lambda} \sum_{r=1}^{n} \sum_{k=1}^{2} \text{Re} \left\{ Q\left(\frac{x - \left\{ \overline{W}_r + (-1)^k \text{i} | W_{r,1} - W_{r,2}|/2 \right\}}{\lambda} \right) \right\}.$$

In the case of homoscedastic measurement error, the Monte Carlo version of (6) is

$$\widehat{f}_{B,\text{hom}}(x) = \frac{1}{n\lambda} \sum_{r=1}^{n} \frac{1}{B} \sum_{b=1}^{B} \text{Re} \left\{ Q\left(\frac{x - \left\{ \overline{W}_r + \text{i}(\frac{d}{m_r})^{1/2} \widehat{\sigma} T_{r,b} \right\}}{\lambda} \right) \right\},$$

where $T_{r,1}, \ldots, T_{r,B}$ are independent copies of $T_r = Z_{r,1}/\sqrt{Z_{r,1}^2 + \cdots + Z_{r,d}^2}$ for each r.

3 Mean integrated squared error

We now derive the mean integrated squared error (MISE) of the heteroscedastic-error estimator in (4). An objective is to compare the asymptotic properties of (4) to those of the known-variance deconvolution estimator. It was shown by Carroll and Hall (1988) that when measurement errors are normally distributed with constant variance, the optimal rate at which any deconvolution estimator can converge to $f_x(x)$ is $\{\log(n)\}^{-2}$. The deconvoluting kernel density estimator achieves this rate of convergence (Stefanski and Carroll 1990), and does so with the optimal bandwidth of $\lambda = \sigma\{\log(n)\}^{-1/2}$ (Stefanski 1989).

For each X_r, assume a fixed number $m_r \geq 2$ of replicate measurements are observed and used to estimate σ_r^2. The independence of the summands in (4) simplifies the derivation of the MISE and results in closed-form expressions for certain special cases of m_r. In practice, it is common that the number of replicates is small and we examine in detail the case of two replicate measurements, $m_r = 2$ for all r. In addition, we present a weighted estimator, a natural extension in the case of nonconstant variances, where weights are selected to minimize the asymptotic integrated variance. We derive the MISE of the weighted estimator under the assumption that the weights are known. We note that, although we limit our asymptotic analysis to $n \to \infty$, it is also reasonable to consider asymptotic behavior as the number of replicates increases to infinity. Delaigle (2008) provides convincing arguments for this approach.

3.1 Heteroscedastic measurement errors

We obtain expressions for the MISE of $\widehat{f}_{\text{het}}(x)$ using results on hypergeometric series. We start with notation and definitions. For all real x define the *rising factorial*

$$(x)_k = \begin{cases} 1 & \text{if } k = 0; \\ x(x+1)(x+2)\cdots(x+k-1) & \text{if } k = 1, 2, 3, \ldots \end{cases}$$

and note that for real $x \neq 0$ and $y \neq 0$,

$$
\begin{aligned}
(1)_k &= k! \quad \text{for } k = 1, 2, 3, \ldots \\
\frac{(xy)_k}{(x)_k} &= y \frac{(xy+1)_{k-1}}{(x+1)_{k-1}} \quad \text{for } k = 1, 2, 3, \ldots \\
\lim_{x \to 0} \frac{(xy)_k}{(x)_k} &= \begin{cases} 1, & \text{if } k = 0; \\ y, & \text{if } k = 1, 2, 3, \ldots \end{cases}
\end{aligned}
\tag{16}
$$

Denote the generalized hypergeometric series (Erdelyi 1953) by

$$
{}_pF_q(a_1, \ldots, a_p; \, b_1, \ldots, b_q; \, x) = \sum_{k=0}^{\infty} \frac{(a_1)_k (a_2)_k \ldots (a_p)_k}{(b_1)_k (b_2)_k \ldots (b_q)_k} \frac{x^k}{k!},
\tag{17}
$$

where it is assumed that the a_j and b_j are such that division by 0 does not arise. We use the relationships (for $b \neq 1/2$)

$$
\begin{aligned}
\{{}_0F_1(; \, b; \, x)\}^2 &= {}_2F_3\left(b, b - \frac{1}{2}; \, b, b, 2b - 1; \, 4x\right) \\
&= {}_1F_2\left(b - \frac{1}{2}; \, b, 2b - 1; \, 4x\right).
\end{aligned}
\tag{18}
$$

A proof of the first equality appears in Bailey (1928); the second follows upon simplification after substitution into (17). Because ${}_0F_1(; \, b; \, x)$ is well-defined and continuous in b at $b = 1/2$, the identity in (18) can be extended via

$$
\begin{aligned}
\{{}_0F_1(; \, .5; \, x)\}^2 &= \lim_{b \to 1/2} {}_1F_2\left(b - \frac{1}{2}; \, b, 2b - 1; \, 4x\right) \\
&= 1 + \frac{{}_0F_1(; \, .5; \, 4x) - 1}{2}.
\end{aligned}
\tag{19}
$$

The limit in (19) is readily verified using the properties in (17). We now state and prove our main result.

Theorem 3.1.1 *Consider data following model (3) and the estimator $\widehat{f}_{het}(x)$ in (4). As $n \to \infty$ and $\lambda \to 0$, with $\mu_{Q,2}^2 = \int z^2 Q(z)dz$,*

$$MISE\{\widehat{f}_{het}(x)\} \sim \frac{1}{2\pi n^2 \lambda} \sum_{r=1}^{n} \int \Phi_Q^2(t) H(m_r, 2t^2\lambda^{-2}\sigma_r^2/m_r)dt$$

$$+ \frac{\lambda^4}{4}\mu_{Q,2}^2 \int \{f_x''(x)\}^2 dx, \tag{20}$$

where

$$H(m, z) = \begin{cases} (e^z + 1)/2 & \text{if } m = 2; \\ {}_1F_1\left(\dfrac{m-2}{2}; \ m-2; \ z\right) & \text{if } m \geq 3. \end{cases} \tag{21}$$

Proof First from (14) we write $\widehat{\Phi}_{u_r}^{-1}(t/\lambda) = {}_0F_1(; (m_r - 1)/2; \widehat{\theta})$ where

$$\widehat{\theta} = \frac{t^2(m_r - 1)\widehat{\sigma}_r^2}{4\lambda^2 m_r}.$$

Now consider the decomposition of $MISE\{\widehat{f}_{het}(x)\}$,

$$MISE\{\widehat{f}_{het}(x)\} = \int Var\{\widehat{f}_{het}(x)\}dx + \int Bias^2\{\widehat{f}_{het}(x)\}dx. \tag{22}$$

It follows from (5) that as $\lambda \to 0$,

$$\int Bias^2\{\widehat{f}_{het}(x)\}dx \sim \frac{\lambda^4}{4}\mu_{Q,2}^2 \int \{f_x''(x)\}^2 dx. \tag{23}$$

The integrated variance can be further decomposed and from (10), $\int Var\{\widehat{f}_{het}(x)\}dx = V_1 - V_2$ where

$$V_1 = \frac{1}{n^2\lambda^2} \sum_{r=1}^{n} \int E\left[\left\{Q^*\left(\frac{x - \overline{W}_r}{\lambda}, \lambda, m_r, \widehat{\sigma}_r\right)\right\}^2\right] dx$$

$$= \frac{1}{n^2\lambda} \sum_{r=1}^{n} E \int \left\{Q^*(z, \lambda, m_r, \widehat{\sigma}_r)\right\}^2 dz$$

$$= \frac{1}{2\pi n\lambda} E \int \Phi_Q^2(t)\widehat{\Phi}_{u_r}^{-2}(t/\lambda)dt$$

$$= \frac{1}{2\pi n\lambda} \int \Phi_Q^2(t) E\left[\{{}_0F_1(; (m_r - 1)/2; \widehat{\theta})\}^2\right] dt \tag{24}$$

using a change of variables and Parseval's Identity, and

$$V_2 = \frac{1}{n^2\lambda^2} \sum_{r=1}^{n} \int \left\{ E\left[Q^*\left(\frac{x - \overline{W}_r}{\lambda}, \lambda, m_r, \widehat{\sigma}_r \right) \right] \right\}^2 dx$$

$$= \frac{1}{n^2\lambda^2} \sum_{r=1}^{n} \int \left\{ E\left[Q\left(\frac{x - X_r}{\lambda} \right) \right] \right\}^2 dx$$

$$= \frac{1}{n} \int \left\{ \int Q(z) f_x(x - \lambda z) dz \right\}^2 dx$$

$$= \frac{1}{2\pi n} \int |\Phi_f(t)|^2 \Phi_Q^2(\lambda t) dt. \tag{25}$$

The second equality is a consequence of Theorem 1.1.1. The last two follow from a change of variables and Parseval's Identity.

It remains to evaluate the expectation of $\{_0F_1(; (m_r - 1)/2; \widehat{\theta})\}^2$ in (24). Using Eqs. 15, 17, and 18 (for $m_r \geq 3$) or 19 (for $m_r = 2$), it is easily verified that

$$E[\{_0F_1(; (m_r - 1)/2; \widehat{\theta})\}^2] = H(m_r, \ 2t^2\lambda^{-2}\sigma_r^2/m_r), \tag{26}$$

so that

$$V_1 = \frac{1}{2\pi n^2\lambda} \sum_{r=1}^{n} \int \Phi_Q^2(t) H(m_r, \ 2t^2\lambda^{-2}\sigma_r^2/m_r). \tag{27}$$

The result follows from (22) and the fact that $V_2 = o\{(n\lambda)^{-1}\}$. □

The infinite sum makes the asymptotic behavior of the MISE difficult to examine for general sequences $m_r, r = 1, 2, \ldots$. This sum converges for $m_r \geq 2$, and closed-form expressions can be obtained for special cases of m_r. We focus the remainder of our analysis on the important case of two replicate measurements.

3.1.1 Heteroscedastic measurement errors: $m_r = 2$

When $m_r = 2$ for all r, the expectation in Eq. 26 simplifies to

$$E[\{_0F_1(; (m_r - 1)/2; \widehat{\theta})\}^2] = \frac{1}{2} e^{t^2\sigma_r^2/\lambda^2} + \frac{1}{2},$$

and V_1 in Eq. 27 becomes

$$V_1 = \frac{1}{4\pi n^2\lambda} \sum_{r=1}^{n} \int \Phi_Q^2(t) e^{t^2\sigma_r^2/\lambda^2} dt + \frac{1}{4\pi n\lambda} \int \Phi_Q^2(t) dt$$

$$= \frac{1}{4\pi n^2\lambda} \sum_{r=1}^{n} \int \Phi_Q^2(t) e^{t^2\sigma_r^2/\lambda^2} dt + o(\{n\lambda^2\}^{-1}),$$

so that as $n \to \infty$ and $\lambda \to 0$,

$$\text{MISE}\{\widehat{f}_{\text{het}}(x)\} \sim \frac{1}{4\pi n^2 \lambda} \sum_{r=1}^{n} \int \Phi_Q^2(t) e^{t^2 \sigma_r^2 / \lambda^2} dt + \frac{\lambda^4}{4} \mu_{Q,2}^2 \int \{f_x''(x)\}^2 dx. \quad (28)$$

Asymptotic analysis of the MISE ($n \to \infty$, $\lambda \to 0$) in Eq. 28 is difficult because of the dependence of the MISE on the particular sequence of variances $\sigma_1^2, \sigma_2^2, \ldots$. However useful insights can be gained under the assumption that the empirical distribution function of $\sigma_1^2, \ldots, \sigma_n^2$ converges to an absolutely continuous distribution. McIntyre et al. (2003) shows that when measurement error variances are distributed uniformly over a finite interval, the rate of convergence of $\widehat{f}_{\text{het}}$ is proportional to $\{\log(n)\}^{-2}$, the same rate found by Stefanski and Carroll (1990) for the deconvoluting estimator when measurement errors are normally distributed with known, constant variance. Furthermore, both the optimal bandwidth and the optimal rate of convergence depend on the upper support boundary of the variance distribution. The indicated conclusions are that estimating heteroscedastic error variances from just two replicates has no effect on the asymptotic rate of convergence, but that the constant multiplying this rate, and also the bandwidth needed to achieve the rate, are dependent on the larger error variances.

3.2 Heteroscedastic measurement errors: weighting

We next investigate the use of weights to reduce the variability of $\widehat{f}_{\text{het}}(x)$ in (4). As $\widehat{f}_{\text{het}}(x)$ is the sum of independent components, having a common mean but different variances, it is reasonable to expect that weighting will reduce variability.

Optimal weights for the weighted estimator are derived under the assumption that the measurement error variances are known, and weights are selected to minimize the asymptotic integrated variance. We examine the asymptotic properties of the estimator calculated with these optimal weights, and our results provide intuition for the properties of $\widehat{f}_{\text{wt}}(x)$, which uses the set of estimated weights calculated with the unknown measurement error variances replaced by their sample estimates.

The optimal weighted estimator has the general form

$$\widetilde{f}_{\text{wt}}(x) = \frac{1}{\lambda} \sum_{r=1}^{n} w_r E \left\{ Q \left(\frac{x - \left\{ \overline{W}_r + i \left(\frac{m_r - 1}{m_r} \right)^{1/2} \widehat{\sigma}_r T_r \right\}}{\lambda} \right) \,\middle|\, \overline{W}_r, \widehat{\sigma}_r^2 \right\}, \quad (29)$$

where $w_1, \ldots, w_n$ are known constants, $w_r \geq 0$ for all r and $\sum_{r=1}^{n} w_r = 1$. The integrated variance of $\widetilde{f}_{\text{wt}}(x)$ is $\int \text{Var}\{\widetilde{f}(x)\} dx = \widetilde{V}_1 - \widetilde{V}_2$ where

$$\widetilde{V}_1 = \frac{1}{\lambda^2} \sum_{r=1}^{n} w_r^2 \int E \left\{ \left[Q^* \left(\frac{x - \overline{W}_r}{\lambda}, \lambda, m_r, \widehat{\sigma}_r \right) \right]^2 \right\} dx, \quad (30)$$

and

$$\widetilde{V}_2 = \frac{1}{\lambda^2} \sum_{r=1}^{n} w_r^2 \int \left\{ E\left[Q^*\left(\frac{x - \overline{W}_r}{\lambda}, \lambda, m_r, \widehat{\sigma}_r \right) \right] \right\}^2 dx,$$

and $Q^*(z, \lambda, m_r, \widehat{\sigma}_r)$ is defined as in Eq. 9.

It follows from Eq. 25 that, provided $w_r \leq B/n$ for $r = 1, \ldots, n$ and some $0 < B < \infty$, $\widetilde{V}_2 = o\{(n\lambda)^{-1}\}$. Thus asymptotically, $\widetilde{V}_1$ is the dominant term in the integrated variance, and we select weights to minimize this quantity. Using (21), let

$$h_r(\sigma_r^2, m_r, \lambda) = \int \Phi_Q^2(t) H(m_r, 2t^2\lambda^{-2}\sigma_r^2/m_r) dt$$

and note from Eq. 27 that $h_r(\sigma_r^2, m_r, \lambda)$ is proportional to the contribution to the asymptotic integrated variance from the rth component of $\widehat{f}_{het}(x)$. Define

$$w_r = \frac{\{h_r(\sigma_r^2, m_r, \lambda)\}^{-1}}{\sum_{r=1}^{n} \{h_r(\sigma_r^2, m_r, \lambda)\}^{-1}} \tag{31}$$

for $r = 1, \ldots, n$. A straightforward application of the Cauchy-Schwartz inequality shows that these weights minimize (30). Substitution into (30) yields

$$\widetilde{V}_1 = \left\{ 2\pi\lambda \sum_{r=1}^{n} \left\{ \int \Phi_Q^2(t) H(m_r, 2t^2\lambda^{-2}\sigma_r^2/m_r) dt \right\}^{-1} \right\}^{-1}. \tag{32}$$

Finally, substituting the sample variances, $\widehat{\sigma}_1^2, \ldots, \widehat{\sigma}_n^2$ for the true variances in Eq. 31 forms the set of estimated weights, $\widehat{w}_1, \ldots, \widehat{w}_n$, and the estimator

$$\widehat{f}_{wt}(x) = \frac{1}{\lambda} \sum_{r=1}^{n} \widehat{w}_r E \left\{ Q\left(\frac{x - \left\{ \overline{W}_r + i\left(\frac{m_r-1}{m_r}\right)^{1/2} \widehat{\sigma}_r T_r \right\}}{\lambda} \right) \middle| \overline{W}_r, \widehat{\sigma}_r^2 \right\}. \tag{33}$$

3.2.1 Mean integrated squared error of $\widetilde{f}_{wt}(x)$

We now consider the asymptotic properties of $\widetilde{f}_{wt}(x)$, calculated with the optimal weights in Eq. 31. We derive an expression for the MISE of $\widetilde{f}_{wt}(x)$ under the assumption that the sequence of measurement error variances, $\sigma_1^2, \ldots, \sigma_n^2$, is known. This analysis provides guidelines for the asymptotic behavior of $\widehat{f}_{wt}(x)$ in Eq. 33. Because the estimated weights are not consistent for the true weights, the resulting estimator is not guaranteed to perform as well as the true-weight estimator asymptotically. Nevertheless, the simulation results reported in Sect. 4 indicate the approximate weighting has a significant beneficial effect in finite samples.

The estimator $\widetilde{f}_{wt}(x)$ has the same key property in Eq. 5 as $\widehat{f}_{het}(x)$, and so has the same integrated squared bias as the kernel density estimator of $f_x(x)$ based on the true data. Combining (23) and (32), we have that as $n \to \infty$ and $\lambda \to 0$,

$$\text{MISE}\{\widetilde{f}_{\text{wt}}(x)\} \sim \left\{ 2\pi\lambda \sum_{r=1}^{n} \left[\int \Phi_Q^2(t) H(m_r, 2t^2\lambda^{-2}\sigma_r^2/m_r)\mathrm{d}t \right]^{-1} \right\}^{-1}$$
$$+ \frac{\lambda^4}{4}\mu_{Q,2}^2 \int \{f_x''(x)\}^2\mathrm{d}x. \tag{34}$$

The expression simplifies in the case of two replicate measurements. From Eqs. 32 and 21, when $m_r = 2$ for all r, as $n \to \infty$ and $\lambda \to 0$,

$$\widetilde{V}_1 \sim \left\{ 4\pi\lambda \sum_{r=1}^{n} \left[\int \Phi_Q^2(t) \mathrm{e}^{t^2\sigma_r^2/\lambda^2}\mathrm{d}t \right]^{-1} \right\}^{-1},$$

and the MISE becomes

$$\text{MISE}\{\widetilde{f}_{\text{wt}}(x)\} \sim \left\{ 4\pi\lambda \sum_{r=1}^{n} \left[\int \Phi_Q^2(t) \mathrm{e}^{t^2\sigma_r^2/\lambda^2}\mathrm{d}t \right]^{-1} \right\}^{-1} + \frac{\lambda^4}{4}\mu_{Q,2}^2 \int \{f_x''(x)\}^2\mathrm{d}x.$$

The asymptotic behavior of MISE again depends on the sequence of variances $\sigma_1^2, \ldots,$ σ_n^2. McIntyre et al. (2003) shows that when the variances are uniformly distributed over a finite interval, the MISE of $\widehat{f}_{\text{wt}}(x)$ converges to 0 at a rate proportional to $\{\log(n)\}^{-2}$, the same rate achieved by the unweighted estimator under that assumption. The result implies that when measurement error variances are known and uniformly distributed, there is no advantage to weighting in terms of the rate of convergence of the MISE. However, the constant multiplying the rate is reduced. Furthermore, the primary advantage of weighting is in finite samples as we study via simulation in Sect. 4.

4 Simulation study

We performed a simulation study to investigate the finite-sample properties of the proposed estimators in Eqs. 4, 29, and 6. These were calculated using the kernel $Q(x) \propto \{\sin(x)/x\}^4$, where $Q(x)$ was scaled to have mean 0 and variance 1. We also included in the study the so-called naive estimator which ignores measurement error. Specifically, we studied the naive estimator to be the kernel density estimator calculated with the sample means of the replicate measurements,

$$\widehat{f}_{\text{naive}}(x) = \frac{1}{n\lambda} \sum_{r=1}^{n} \phi\left(\frac{x - \overline{W}_r}{\lambda} \right),$$

where $\phi(x)$ is the standard normal density.

Estimators were compared on the basis of average integrated squared error (ISE). For each simulated data set, estimators were computed at their optimal bandwidths, found by minimizing the integrated squared error. As this requires knowledge of the unknown density $f_x(x)$, the estimators in this study are not true estimators.

However our results provide insight into their relative optimal performances, independent of the problem of estimating a bandwidth. We defer a discussion of bandwidth estimation to Sect. 5. For additional comparison, however, we also calculated the naive estimator using the popular normal reference bandwidth estimator, $\widehat{\tau} = 1.06 \, \widehat{\sigma}_{\overline{W}} \, n^{-1/5}$ where $\widehat{\sigma}^2_{\overline{W}}$ is the sample variance of the means $\overline{W}_1, \ldots, \overline{W}_n$ (Silverman 1986).

We examined three factors in this study. First, six different sample sizes were considered, $n = 100, 500, 1000, 2000$ and 2500. Second, we examined the effect of the true-data density, generating data, $X_1, \ldots, X_n$, from the N(0, 1) density and the Chi-squared(4) density, standardized to have mean 0 and variance 1. Finally, we examined the effect of the homogeneity of measurement error variances on the performance of each estimator. Observed data were generated with normal measurement errors having variances $\sigma^2_1 = \cdots = \sigma^2_n = 1$ for the case of constant variances, and $\sigma^2_1, \ldots, \sigma^2_n$ chosen uniformly over the interval (0, 2) for the case of nonconstant variances. All estimators were computed on each set of observed data.

In practice, often only a small number of replicate measurements is observed, with $m_r = 2$ common. Thus in our simulations, we considered only the case of $m_r = 2$ replicate measurements. All results are based on fifty simulated data sets.

Average ISEs are plotted by sample size in Fig. 1 for $X_r \sim$ N(0, 1) and Fig. 2 for $X_r \sim$ Chi-squared(4). For simulations that considered homoscedastic measurement error, the average ISE for $\widehat{f}_{\text{hom}}(x)$ was significantly smaller ($\alpha = 0.05$) than that of the other estimators for each sample size except $n = 100$, where all estimators were most variable. For simulations that considered heteroscedastic measurement errors, $\widehat{f}_{\text{het}}(x)$ had a significantly higher average ISE than all other estimators for sample sizes of $n \geq 500$, and was also much more variable. The estimators $\widehat{f}_{\text{hom}}(x)$ and $\widehat{f}_{\text{wt}}(x)$ had significantly lower average integrated squared errors than all
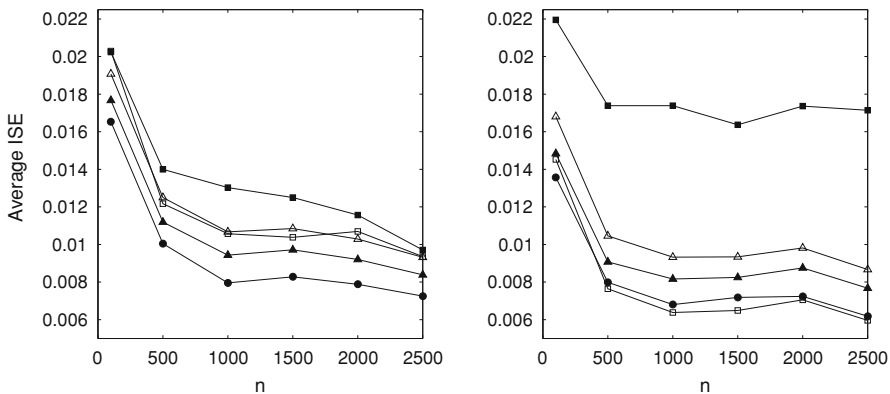
Fig. 1 Average ISE by sample size for $X \sim$ N(0, 1). *Left* homoscedastic measurement errors, N(0, 1). *Right* heteroscedastic measurement errors, N(0, σ^2_r) with σ^2_r uniform on (0, 2). *Open triangle* $\widehat{f}_{\text{naive}}$ with plug-in bandwidth; *Closed triangle* $\widehat{f}_{\text{naive}}$ with optimal bandwidth; *Closed square* $\widehat{f}_{\text{het}}$; *Open square* $\widehat{f}_{\text{wt}}$; *Closed circle* $\widehat{f}_{\text{hom}}$
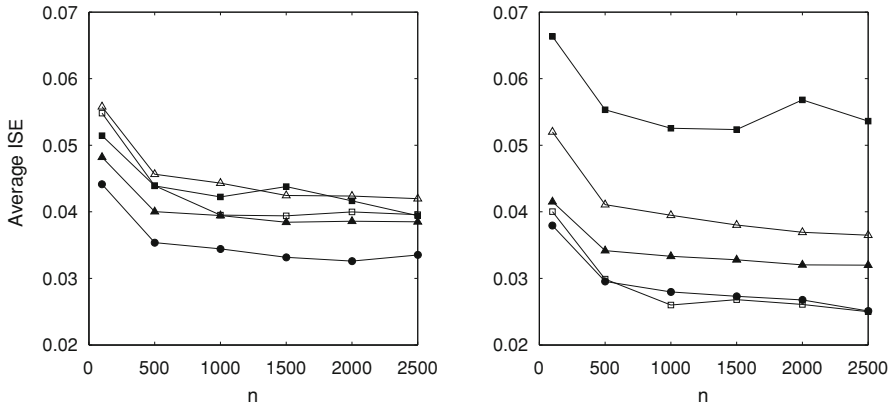
Fig. 2 Average ISE by sample size for $X \sim \chi_4^2$. *Left* homoscedastic measurement errors, N(0, 1). *Right* heteroscedastic measurement errors, N(0, σ_r^2) with σ_r^2 uniform on (0, 2). *Open triangle* $\widehat{f}_{naive}$ with plug-in bandwidth; *Closed triangle* $\widehat{f}_{naive}$ with optimal bandwidth; *Closed square* $\widehat{f}_{het}$; *Open square* $\widehat{f}_{wt}$; *Closed circle* $\widehat{f}_{hom}$

other estimators for $n \geq 500$, but differences between the two were generally not significant.

It is evident from the simulations that weighting is effective in reducing the ISE of $\widehat{f}_{het}(x)$, and $\widehat{f}_{wt}(x)$ is preferred. In almost every case, the average ISE of $\widehat{f}_{wt}(x)$ was significantly less than that of $\widehat{f}_{het}(x)$. However, even the weighted estimator performed relatively poorly when measurement error variances were constant, not surprising considering that here $\widehat{f}_{wt}(x)$ relies on n estimates of the constant variance each obtained from only two replicate measurements. The estimator for homoscedastic errors performed well for both types of measurement errors, suggesting it is a good choice even when there is doubt about the homogeneity of the error variances, provided the heterogeneity is not too great. It is encouraging that for the sample sizes considered here, the measurement error-corrected estimators perform as well or better than the naive estimator. This suggests that given a reliable rule for selecting a bandwidth, correcting for measurement errors is worthwhile in many situations.

5 A bootstrap method for bandwidth election

We now describe a bootstrap bandwidth selection procedure, and illustrate it with data in Sect. 6. The bootstrap has been used for bandwidth selection in traditional kernel density estimation (Marron 1992). Delaigle and Gijbels (2004) extended it to the deconvolution estimator (7). With the bootstrap method the bandwidth estimate can be obtained without resampling. We note that the bandwidth-selection method proposed by Delaigle and Hall (2008) based on SIMEX (Cook and Stefanski 1994; Stefanski and Cook 1995) could also be applied to our estimators.

We obtain an initial estimate of bandwidth, $\widehat{\tau}$, used to construct an initial density estimate for generating bootstrap samples. For our deconvolution problem, we use the naive estimator,

$$\widehat{f}_{\text{naive}}(x; \widehat{\tau}) = \frac{1}{n\widehat{\tau}} \sum_{r=1}^{n} \phi\left(\frac{x - \overline{W}_r}{\widehat{\tau}}\right),$$

where $\widehat{\tau}$ is the normal reference bandwidth (Silverman 1986). This provides a reasonable, though overly smoothed, estimate of $f_x(x)$.

For heteroscedastic measurement errors, bootstrap data are generated for $r = 1, \ldots, n$ and $j = 1, \ldots, m_r$ as $W_{r,j}^* = X_r^* + \widehat{\sigma}_r Z_{r,j}$, where X_r^* is generated from the density $\widehat{f}_{\text{naive}}(x; \widehat{\tau})$, $\widehat{\sigma}_r$ is the estimate of σ_r from the original data and $Z_{r,j} \sim N(0, 1)$. In the case of homoscedastic measurement errors, $\widehat{\sigma}_r$ is replaced with the pooled estimate of variance, $\widehat{\sigma}$.

Let $\widehat{f}_.^*(x; h)$ denote any of the three deconvolution estimators from (4), (6) or (29) computed with the bootstrap sample. The bootstrap estimate of the optimal bandwidth for a measurement error-corrected estimator is the value h that minimizes

$$\text{MISE}^*(h) = E_{\text{BS}}\left[\int \left\{\widehat{f}_.^*(x; h) - \widehat{f}_{\text{naive}}(x; \widehat{\tau})\right\}^2 \mathrm{d}x\right],$$

where E_{BS} denotes expectation with respect to the bootstrap distribution. The optimal bandwidth can be determined empirically by computing a large number of bootstrap estimates over a dense grid of bandwidths and selecting the value of h that achieves the smallest average ISE. However because $\widehat{f}_{\text{naive}}(z, \widehat{\tau})$ replaces the unknown $f(x)$, the optimal bandwidth can also be determined by direct calculation of $\text{MISE}^*(h)$ over the grid of bandwidths, using analytical expressions for MISE, e.g., Eqs. 20 and 34. For more details see McIntyre et al. (2003).

6 An application

We illustrate the bootstrap bandwidth procedure using data from a U.S. EPA study of automobile emissions. Measurements of carbon monoxide (CO) in automobile emissions were collected with a remote sensing device stationed at a highway entrance ramp in North Carolina. The device measured CO in the exhaust of passing automobiles, and also photographed their license plates. Measurements were taken on several different dates resulting in replicate measurements from cars that passed the location multiple times. Of a total of 3,002 automobiles observed, 1233 had multiple measurements (946 with $m_r = 2$ and 287 with $m_r = 3$).

An objective of the study was to characterize average CO emissions among the population of automobiles. Measurements of CO taken from a stationary point on the roadway are subject to error from several sources including meteorological conditions, variations in engine temperature, speed and acceleration, and instrument error. The multiple sources of variation are expected to result in multiplicative errors. Thus we log-transformed the CO measurements.

The log-transformed data showed evidence of heteroscedastic measurement errors. A loess curve fit to the scatter plot of $\widehat{\sigma}_r$ versus $\overline{W}_r$ exhibited substantial nonlinearities over the range of sample means, although globally there was a tendency for larger sample standard deviations to be associated with larger sample means. Testing the
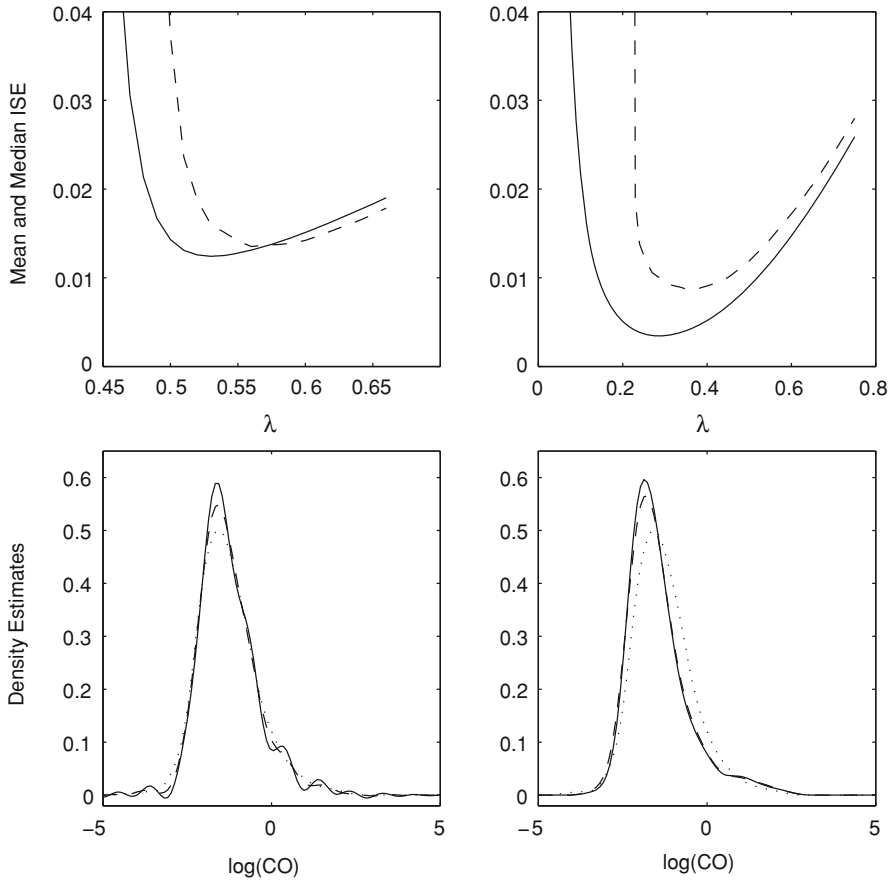
Fig. 3 *Top row* Analytical mean (*Solid line*) and empirical median (*Dashed line*) ISE of bootstrap density estimates as a function of bandwidth. *Left* Homoscedastic estimator; *Right* Weighted estimator; *Bottom row* Density estimates with bandwidth minimizing bootstrap MISE (*solid line*) and median ISE (*dashed line*). *Left* Homoscedastic estimator; *Right* Weighted estimator. *Dotted line* in both panels is the naive estimator with plug-in bandwidth

latter via a simple linear regression of the sample standard deviations on the sample means resulted in a significant slope of 0.0936 ($p < .0001$).

We estimated the density function of the log-transformed CO measurements using the 1,233 automobiles with multiple measurements. For comparison, we fit the weighted estimator for heteroscedastic errors and the estimator for homoscedastic errors. Bandwidths were selected using both analytical and empirical calculation of the bootstrap bandwidth estimate as described in Sect. 5. Each empirical estimate was based on 100 bootstrap data sets and the bandwidth was selected to minimize the median integrated squared error. Results are shown in Fig. 3. Also shown is the naive estimate, calculated using the sample means of replicate observations and the normal reference bandwidth. All measurement error-corrected estimators show a higher peak and slightly thinner tails than the naive estimator. This is expected when the effects of measurement error are removed from the observed-data estimate.

References

Andrews, G., Askey, R., Roy, R. (1999). *Special functions*. Cambridge: Cambridge University Press.

Bailey, W. N. (1928). Products of generalized hypergeometric series. *Proceedings of the London Mathematical Society*, *2*, 242–254.

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

Cook, J. R., Stefanski, L. A. (1994). Simulation–extrapolation estimation in parametric measurement error models. *Journal of the American Statistical Association*, *89*, 1314–1328.

Delaigle, A. (2008). An alternative view of the deconvolution problem. *Statistica Sinica*, *18*, 1025–1045.

Delaigle, A., Gijbels, I. (2004). Bootstrap bandwidth selection in kernel density estimation from a contaminated sample. *Annals of the Institute of Statistical Mathematics*, *56*, 19–47.

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

Delaigle, A., Meister, A. (2007). Nonparametric regression estimation in the heteroscedastic errors-in-variables problem. *Journal of the American Statistical Association*, *102*, 1416–1426.

Delaigle, A., Meister, A. (2008). Density estimation with heteroscedastic error. *Bernoulli*, *14*, 562–569.

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

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

Devroye, L. (1989). Consistent deconvolution in density estimation. *The Canadian Journal of Statistics*, *17*, 235–239.

Diggle, P. J., Hall, P. (1993). A Fourier approach to nonparametric deconvolution of a density estimate. *Journal of the Royal Statistical Society, Series B*, *55*, 523–531.

Erdelyi, A. E. (1953). *Higher transcendental functions* (Vol. 2). New York: McGraw-Hill.

Fan, J. (1991a). Asymptotic normality for deconvolution kernel density estimators. *Sankhya, Series A, Indian Journal of Statistics*, *53*, 97–110.

Fan, J. (1991b). On the optimal rates of convergence for nonparametric deconvolution problems. *Annals of Statistics*, *19*, 1257–1272.

Fan, J. (1992). Deconvolution with supersmooth distributions. *The Canadian Journal of Statistics*, *20*, 155–169.

Li, T., Vuong, Q. (1998). Nonparametric estimation of the measurement error model using multiple indicators. *Journal of Multivariate Analysis*, *65*, 139–165.

Marron, J. S. (1992). Bootstrap bandwidth selection. In: R. LePage & L. Billard (Eds.), *Exploring the limits of bootstrap*. New York: Wiley.

McIntyre, J. (2003). Density deconvolution with replicate measurements and auxiliary data. Ph.D. Thesis, North Carolina State University.

Meister, A. (2006). Density estimation with normal measurement error with unknown variance. *Statistica Sinica*, *16*, 195–211.

Neumann, M. H. (2007). Deconvolution from panel data with unknown error distribution. *Journal of Multivariate Analysis*, *98*, 1955–1968.

Patil, P. (1996). A note on deconvolution density estimation. *Statistics & Probability Letters*, *29*, 79–84.

Silverman, B. W. (1986). *Density estimation for statistics and data analysis*. New York: Chapman & Hall.

Staudenmayer, J., Ruppert, D., Buonaccorsi, J.P. (2008). Density estimation in the presence of heteroskedastic measurement error. *Journal of the American Statistical Association*, *103*, 726–736.

Stefanski, L. A. (1989). Unbiased estimation of a nonlinear function of a normal mean with application to measurement error models. *Communications in Statistics, Series A*, *18*, 4335–4558.

Stefanski, L. A. (1990). Rates of convergence of some estimators in a class of deconvolution problems. *Statistics & Probability Letters*, *9*, 229–235.

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

Stefanski, L. A., Cook, J. R. (1995). Simulation–extrapolation: the measurement error jackknife. *Journal of the American Statistical Association*, *90*, 1247–1256.

Stefanski, L. A., Novick, S. J., Devanarayan, V. (2005). Estimating a nonlinear function of a normal mean. *Biometrika*, *92*, 732–736.

Wand, M. P. (1998). Finite sample performance of deconvolving density estimators. *Statistics & Probability Letters*, *37*, 131–139.

Watson, G. S. (1983). *Statistics on spheres*. New York: Wiley.