Abstract

MCINTYRE, JULIE PILAR. Density Deconvolution with Replicate Measurements and Auxiliary Data. (Under the supervision of Dr. Leonard A. Stefanski)

We present two deconvolution estimators for the density function of a random variable X that is measured with error. The first estimates the density of X from the set of independent replicate measurements $\{W_{r,j}\}_{r=1,j=1}^{n,\ m_r}$, where $W_{r,j} = X_r + U_{r,j}$. We derive an estimator assuming that $\{U_{r,j}\}_{r=1,j=1}^{n,\ m_r}$ are normally distributed measurement errors having unknown and possibly nonconstant variances σ_r^2. The estimator generalizes the deconvolution estimator of Stefanski and Carroll (1990), with the measurement error variances estimated from replicate observations. We derive the integrated mean squared error and examine the rate of convergence as $n \to \infty$ and the number of replicates is fixed. The finite-sample performance of the estimator is illustrated through a simulation study and an example.

The second is a semi-parametric deconvolution estimator that assumes the availability of a covariate vector $\boldsymbol{Z}$ statistically related to $\boldsymbol{X}$, but independent of the error in measuring X, and such that the regression error $X - E(X|\boldsymbol{Z})$ is normally distributed. The estimator combines parametric modeling of the regression residuals with nonparametric estimation of the mean function. The asymptotic properties of the estimator are discussed. The reliance of the estimator on assumptions of the regression model and normality of model errors is examined via simulation, and an application to real data is presented.

DENSITY DECONVOLUTION WITH REPLICATE MEASUREMENTS AND AUXILIARY DATA

by

JULIE P. MCINTYRE

A dissertation submitted to the Graduate Faculty of
North Carolina State University
in partial fulfillment of the
requirements for the Degree of
Doctor of Philosophy

STATISTICS

Raleigh

2003

APPROVED BY:

Leonard A. Stefanski Chair of Advisory Committee	Marie Davidian
Jason A. Osborne	Sastry G. Pantula

Biography

Julie McIntyre was born in Mineral Wells, Texas, and grew up in St. Louis, Missouri. She earned a B.A. degree in mathematics from Northwestern University in 1992, and a M.S. degree in statistics from the University of Alaska Fairbanks in 1998. She began her doctoral studies in August of 1999 in the Department of Statistics at North Carolina State University, and hopes to complete her degree in August of 2003. She has accepted a position of Visiting Assistant Professor in the Department of Statistics at Carnegie Mellon University beginning in September, 2003.

Acknowledgements

I thank my advisor, Len Stefanski, for his guidance during the completion of this dissertation. It was a privilege and a pleasure to learn from him. I also thank my committee members, Marie Davidian, Jason Osborne, and Sastry Pantula, for many thoughtful comments on my research.

This work would not have been possible without the support of the faculty, staff, and graduate students in the statistics department. Their dedication and enthusiasm create an excellent atmosphere for learning.

My research benefitted from several people outside the university. I thank Dr. Don Richards for pointing out mathematical results involving special functions that greatly advanced my research. The first two years of my doctoral studies were supported in part by the U.S. Geological Survey, and I thank the scientists there, particularly Mike Meador and Jerry McMahon, for giving me the opportunity to collaborate on several projects. I also received funding throughout my studies from an NSF-VIGRE grant.

Finally, I am grateful to my family and friends for helping me pursue my goals. Above all, I thank Josh McIntyre, for friendship, encouragement, and humor throughout it all.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

1.1 Introduction

This thesis presents strategies for estimating the density function of a random variable that is measured with error. A variable X is said to be measured with error when its true value is obscured by random noise. If W denotes an observation of X, then W contains characteristics of both X and the error in its measurement. The assumption that certain variables are error-free is implicit in most data analysis procedures, and naively replacing them with their observed values can lead to imprecise or even incorrect inferences.

This is true when the objective of an analysis is estimation of $f_x(x)$, the density function of X. Traditional estimators of $f_x(x)$ suppose a random sample, $X_1, \ldots, X_n$, is available from this density, and substituting the observed sample $W_1, \ldots, W_n$ results in an inconsistent estimate.

The need for methods that consistently estimate $f_x(x)$ from such an observed sample arises in numerous applications. Density estimates of pollutant levels in automobile exhaust help researchers at the U.S. EPA develop effective regulatory strategies. Pollutant levels are measured from the exhaust of travelling cars, and are known to contain a considerable amount of error. Density estimates of nutrient consumptions are used by scientists at the USDA to produce nutritional guidelines. Food-recall surveys, often used to measure nutrient consumption, are subject to error from the inability of respondents to accurately recall the types and amounts of foods they ate on a given day. Wildlife biologists interested in the size of animal populations estimate the density function of the distance at which animals are sighted from a transect line. Because distances cannot be measured accurately without disturbing animals, they are roughly estimated by eye.

This chapter contains an overview of measurement error problems. To provide background, we first discuss measurement error in the context of some familiar statistical procedures. This provides a foundation for considering the effects of measurement error in density estimation, and we discuss this problem in detail and review relevant research. The chapter concludes with a statement of the specific density estimation problems that are the focus of this dissertation.

1.2 Measurement Error and Statistical Inference

Many important concepts in the analysis of data that are measured with error can be illustrated with simple examples. As we will show, measurement error problems have several common features. Ignoring measurement error can induce bias and increase variance in estimated parameters. In this section, we examine the consequences of ignoring measurement error on certain summary statistics and linear regression models. We compare inferences made from the observed data, $W_1, \ldots, W_n$, to those made were the true data, $X_1, \ldots, X_n$, available. Although more complicated error structures can be considered, it is common to assume that the error in measuring X is additive, unbiased, and independent, and we suppose that for $r = 1, \ldots, n$,

$$W_r = X_r + U_r, \tag{1.2.1}$$

where U_r is a measurement error that has mean 0, variance σ_u^2, and is independent of X_r. This is known as the classical error model in the measurement error modeling literature.

As our discussion will make clear, correcting for the effects of measurement error requires additional information about the error variable. At a minimum, the measurement error variance, σ_u^2, must be known or estimable. Replicate measurements of the true data, for instance, can provide a means of estimating σ_u^2.

1.2.1 The Effects of Measurement Error on Some Summary Statistics

First, consider estimating μ_x, the mean of X. The true-data estimator,

$$\overline{X} = \frac{1}{n}\sum_{r=1}^{n} X_r,$$

is an unbiased estimator of μ_x. Naively replacing the true data with observed values in this estimator produces the estimate $\overline{W}$,

$$\overline{W} = \frac{1}{n}\sum_{r=1}^{n} W_r = \frac{1}{n}\sum_{r=1}^{n}(X_r + U_r) = \overline{X} + \overline{U}.$$

$\overline{W}$ is also an unbiased estimator of μ_x. However, consider the variances of these two statistics,

$$\mathrm{Var}(\overline{X}) = \frac{\sigma_x^2}{n},$$

while

$$\mathrm{Var}(\overline{W}) = \frac{\sigma_x^2}{n} + \frac{\sigma_u^2}{n}.$$

Thus, when estimating a mean, ignoring measurement error does not affect bias but does increase variability.

Now consider estimating the variance of X. The true-data estimator is the sample variance,

$$s_x^2 = \frac{1}{n-1}\sum_{r=1}^{n}(X_r - \overline{X})^2,$$

and is unbiased for σ_x^2. The estimator constructed with the observed-data sample is

$$s_w^2 = \frac{1}{n-1}\sum_{r=1}^{n}(W_r - \overline{W})^2 = s_x^2 + s_u^2 + 2s_{xu},$$

where s_{xu} is the sample covariance between X and U. When X and U are independent,

$$E(s_w^2) = \sigma_x^2 + \sigma_u^2. \tag{1.2.2}$$

Thus, s_w^2 is a biased estimator of σ_x^2. It also has a larger variance, because

$$\begin{aligned}
\mathrm{Var}(s_w^2) &= \mathrm{Var}\{E(s_w^2|\ X_1, \ldots, X_n)\} + E\{\mathrm{Var}(s_w^2|\ X_1, \ldots, X_n)\} \\
&= \mathrm{Var}(s_x^2 + \sigma_u^2) + E\{\mathrm{Var}(s_w^2|\ X_1, \ldots, X_n)\} \\
&= \mathrm{Var}(s_x^2) + E\{\mathrm{Var}(s_w^2|\ X_1, \ldots, X_n)\},
\end{aligned}$$

and so provided $E\{\mathrm{Var}(s_w^2|\ X_1, \ldots, X_n)\} > 0$, $\mathrm{Var}(s_w^2) > \mathrm{Var}(s_x^2)$.

Notice that measurement error induces bias in the sample estimate of the variance, s_w^2, but not in the sample estimate of the mean, $\overline{W}$. In general, measurement error induces bias in estimators that are nonlinear functions of the data. The sample mean $\overline{W}$ is linear in $W_1, \ldots, W_n$, while s_w^2 is not.

Given knowledge of the measurement error variance, an unbiased estimator of σ_x^2 can be constructed from the observed data. Let

$$s_w^{*2} = s_w^2 - \sigma_u^2.$$

It follows from equation (1.2.2) that $E(s_w^{*2}) = \sigma_x^2$. Note that replacing σ_u^2 by a consistent estimator of the measurement error variance does not affect the consistency of s_w^{*2}.

These examples illustrate some fundamental problems with basing statistical inferences on data that contain measurement error. Biases result in inconsistent parameter

estimates, and increased variance leads to wider confidence intervals and less powerful hypothesis tests. These ideas are illustrated further in the next section, through an examination of linear regression models.

1.2.2 Measurement Error in Linear Regression Models

Linear regression analysis assumes that model covariates are measured precisely, and in this section we discuss the consequences of ignoring measurement error in these variables. Covariate measurement error introduces bias and increases variability in the estimates of model coefficients, which in turn adversely affect the performance (size and power) of hypothesis tests. The effects of measurement error on parameter estimation in linear regression models is complicated when a model contains multiple predictors. We concentrate our discussion on simple linear regression, and describe selected results for multiple linear regression. A comprehensive examination of the effects of covariate measurement error in linear models is given by Fuller (1987). The complexity of the problem is magnified even more when models are nonlinear in their regression parameters, and we do not discuss this subject here. An extensive treatment of the effects of covariate measurement error in nonlinear models is given by Carroll, Ruppert, and Stefanski (1995).

Simple Linear Regression

Consider the model where for $r = 1, \ldots, n$,

$$Y_r = \beta_0 + \beta_x X_r + \epsilon_r,$$

and the model errors, $\epsilon_1, \ldots, \epsilon_n$ are independent and identically distributed random variables with mean 0 and variance σ_ϵ^2. Given the true-data sample, $X_1, \ldots, X_n$, the least squares estimator of β_x is

$$\widehat{\beta}_x = \frac{s_{xy}}{s_x^2}, \tag{1.2.3}$$

and because

$$\frac{s_{xy}}{s_x^2} \xrightarrow{P} \frac{\sigma_{xy}}{\sigma_x^2} = \beta_x,$$

$\widehat{\beta}_x$ is a consistent estimator of the true slope.

We now consider the consequences of naively estimating β_x with the observed-data sample, $W_1, \ldots, W_n$. Suppose that data are observed according to equation (1.2.1), where for $r = 1, \ldots, n$, U_r has mean 0, variance σ_u^2, and is independent of both X_r and ϵ_r.

Calculating equation (1.2.3) with the observed data results in the naive estimate of β_x,

$$\widehat{\beta}_w = \frac{s_{wy}}{s_w^2}. \tag{1.2.4}$$

The sample covariance, s_{wy}, is a linear function of $W_1, \ldots, W_n$, and is an unbiased estimator of σ_{xy}. With equation (1.2.2), it follows that

$$\widehat{\beta}_w \xrightarrow{P} \frac{\sigma_{xy}}{(\sigma_x^2 + \sigma_u^2)} = \xi\beta_x,$$

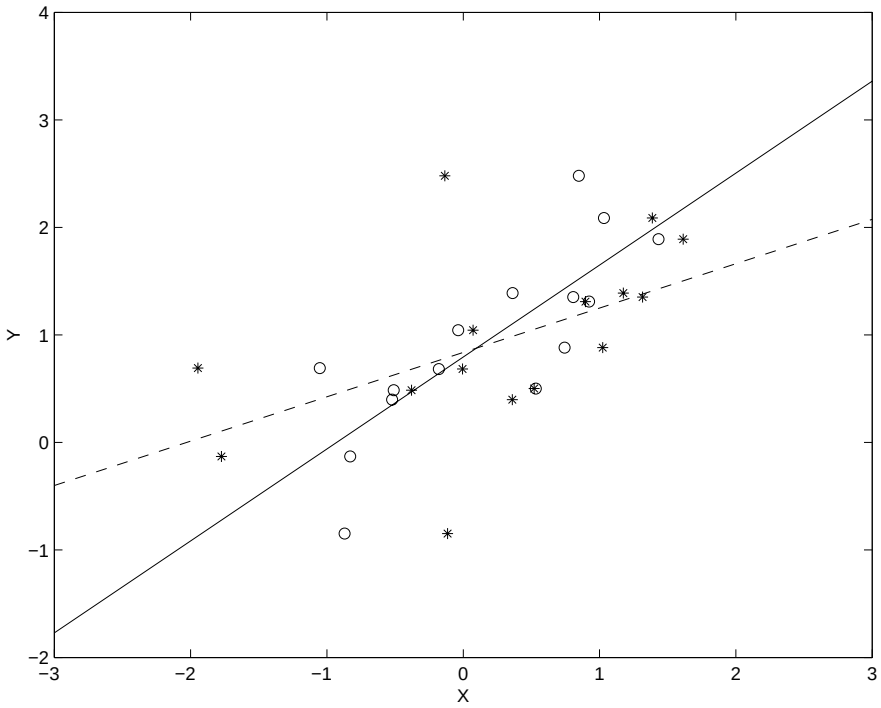

Figure 1.1: Attenuation in the slope of a simple linear regression model caused by measurement error. Circle, solid line: True data and estimated regression line; Star, dashed line: Observed data and estimated regression line; Reliability ratio is 1/2.

where

$$\xi = \frac{\sigma_x^2}{(\sigma_x^2 + \sigma_u^2)} \tag{1.2.5}$$

is known as the reliability ratio. Because $\xi < 1$, naively replacing the true data by observed data results in an estimate of the slope that is is attenuated, or biased toward 0. The attenuation in the naive estimate is illustrated in Figure 1.1.

Inferences about the slope are affected by measurement error. In general, measurement error results in a loss of power for testing the null hypothesis that $\beta_x = 0$. This is straightforward to see when Y, X, and W are jointly normally distributed. In this case, the noncentrality parameter for the test of $H_0 : \beta_x = 0$ is

$$\kappa_x = \frac{n\beta_x^2\sigma_x^2}{\sigma_\epsilon^2}$$

given the true data, and

$$\kappa_w = \frac{n\beta_x^2\sigma_x^2\rho_{xw}^2}{\{\sigma_\epsilon^2 + \beta_x^2\sigma_x^2(1-\rho_{xw}^2)\}}$$

given the observed data, where ρ_{xw} denotes the correlation between X and W. When $\rho_{xw} < 1$, $\kappa_w < \kappa_x$, and thus the power of the observed-data hypothesis test is diminished.

Given knowledge of the measurement error variance, σ_u^2, the bias in the naive estimate can be corrected. Let

$$\widehat{\beta}_w^* = \left(\frac{s_w^2}{s_w^2 - \sigma_u^2}\right)\widehat{\beta}_w.$$

Then it follows from equation (1.2.4) that $\widehat{\beta}_w^* \xrightarrow{P} \beta_x$. The same result holds when σ_u^2 is replaced by a consistent estimate.

Multiple Linear Regression

Consideration of the effects of measurement error on parameter estimation is complicated when a model has multiple covariates, some or all of which may contain measurement error. When ignored, measurement error in a single covariate can compromise inferences about all model parameters. The analysis in these models quickly becomes complex, and most results require strong distributional assumptions on the model components. Still, a few general conclusions can be drawn, and we illustrate these by considering a multiple linear regression model where a single covariate is measured with error. Let

$$Y_r = \beta_0 + \beta_x X_r + \beta_z^T \boldsymbol{Z}_r + \epsilon_r,$$

where $\epsilon_1, \ldots, \epsilon_n$ are independent model errors with mean 0 and variance σ^2_ϵ. We adopt the notation of Carroll, Ruppert, and Stefanski (1995) and assume that the covariate X is measured with error, but the covariates in the vector $\boldsymbol{Z}$ are error free.

Suppose that the random sample $X_1, \ldots, X_n$ is observed as $W_1, \ldots, W_n$ according to equation (1.2.1). We impose the same conditions on the measurement errors as for the case of simple linear regression, i.e., for $r = 1, \ldots, n$, U_r has mean 0, variance σ^2_u, and is independent of X_r and ϵ_r, and we add the additional assumption that U_r is independent of $\boldsymbol{Z}_r$.

In this setting, when true data are replaced by their observed values, not only is the estimate of β_x biased, but when X and $\boldsymbol{Z}$ are correlated, so is the estimate of β_z. First consider estimating β_x. Calculated with the observed data, the least squares estimate of β_x is $\widehat{\beta}_w$ where

$$\widehat{\beta}_w \xrightarrow{P} \left(\frac{\sigma^2_{x|z}}{\sigma^2_{x|z} + \sigma^2_u} \right) \beta_x = \xi_1 \beta_x,$$

and $\sigma^2_{x|z}$ is the conditional variance of X given $\boldsymbol{Z}$ (Carroll, Ruppert & Stefanski 1995). Note that ξ_1 and the reliability ratio, ξ in equation (1.2.5), are equal only if X and $\boldsymbol{Z}$ are uncorrelated. However, because $\sigma^2_{x|z} \leq \sigma^2_x$, it follows that $\xi_1 \leq \xi$, and so correlation between X and $\boldsymbol{Z}$ actually increases the attenuation in $\widehat{\beta}_w$, and therefore exaggerates the consequences for inference about β_x.

When X and $\boldsymbol{Z}$ are correlated, replacing $X_1, \ldots, X_n$ by $W_1, \ldots, W_n$ also leads to bias in the least squares estimate of β_z,

$$\widehat{\beta}_z \xrightarrow{P} \beta_z + \beta_x(1 - \xi_1)\Gamma_z, \tag{1.2.6}$$

where Γ_z^T is the coefficient on Z from its regression on X (Carroll et al. 1995). As can be seen from equation (1.2.6), the direction and magnitude of the bias in this estimate depends on the parameters β_z and Γ_z^T, and the quantity ξ_1. Thus, ignoring measurement error in X can create either positive or negative biases in the coefficients of covariates that are error-free.

1.3 The Effects of Measurement Error on Density Estimation

In this section, we give a detailed description of the problem of estimating the density function of a variable that is measured with error. Because they are nonlinear functions of the observed data, density estimates are biased when calculated from data measured with error.

The problem of correcting for the effects of measurement in density estimation has received considerable attention, and most proposed methods require strong assumptions on the distribution of the measurement error variable. We review a number of these methods in this section.

Difficulties in estimating the density function of an error-prone variable stem from the following. Let X be a random variable with unknown density function $f_x(x)$. Suppose that a random sample from this density, $X_1, \ldots, X_n$, is observed only as $W_1, \ldots, W_n$ according to equation (1.2.1), so that for $r = 1, \ldots, n$, $W_r = X_r + U_r$.

In addition, suppose that U_r has density function $f_u(u)$ and is independent of X_r. Traditional density estimators constructed from the true-data sample consistently estimate $f_x(x)$. Naively computed with the observed-data sample, however, they estimate not $f_x(x)$, but rather $f_w(w) = f_x * f_u(w)$, where $*$ denotes convolution, i.e.,

$$f_w(w) = \int_{-\infty}^{\infty} f_x(x) f_u(w - x) dx.$$

Typically, the observed-data density, $f_w(w)$, is flatter, more spread out, and smoother than the true-data density. Some examples are plotted in Figure 1.2 for the case where $f_u(u)$ is the N(0, 1/2) density. Figure 2a shows the case where $f_x(x)$ is the standard normal density, so that $f_w(w)$ is the N(0, 3/2) density. Relative to the true-data density, the observed-data density has a shorter peak and thicker tails. A case where $f_x(x)$ is bimodal is shown in Figure 2b. Here, $f_x(x)$ is an equal mixture of normals having means $\pm 1/2$ and common variances 1/2. It follows that $f_w(w)$ is an equal mixture of normals with means $\pm 1/2$ and common variances 1. The observed-data density shows none of the bimodal features seen in the true-data density. Figure 2c illustrates the effect of measurement error on a skewed density function. Here, $f_x(x)$ is the Chi-square(4) density, standardized to have mean 0 and variance 1. The observed-data density is the convolution of $f_x(x)$ with the N(0, 1/2) density, and, unlike the first two examples, does not have a familiar form. Skewness is much less apparent in $f_w(w)$ than in $f_x(x)$, and a particularly large difference is seen in the left tails of the two densities, due to the fact that the support of $f_x(x)$ is bounded below, while $f_w(w)$ has support on the whole real line.

These examples illustrate some of the dangers of making inferences from density estimates that are based on error-prone data. Important characteristics such as bimodality and skewness are often obscured. The thicker tails associated with observed-data densities can result in overestimates of tail probabilities, implying a loss of power for certain hypothesis tests when p-values are estimated from observed data.

Methods that seek to recover $f_x(x)$ from the observed data are known as deconvolution methods. Although a variety of approaches to the deconvolution problem have been proposed, our research focuses on kernel-based estimators of $f_x(x)$. Next, we give a brief overview of kernel density estimation, followed by a review of kernel-based deconvolution estimators that have appeared in the literature.

1.3.1 Kernel Density Estimation

The kernel density estimator of $f_x(x)$ is

$$\widehat{f}_x(x) = \frac{1}{n\lambda} \sum_{r=1}^{n} Q\left(\frac{x - X_r}{\lambda}\right), \tag{1.3.7}$$

where λ is the bandwidth, controlling the smoothness of the estimate, and $Q(x)$ is a kernel function. Typically, $Q(x)$ is assumed to be a symmetric function that satisfies

$$\int Q(x)dx = 1, \quad \int xQ(x)dx = 0, \quad \text{and} \quad \int x^2 Q(x)dx = \mu_{q,2} > 0.$$

These conditions are satisfied when $Q(x)$ is a symmetric probability density function.

The estimator in equation (1.3.7) is a consistent estimator of $f_x(x)$ for certain

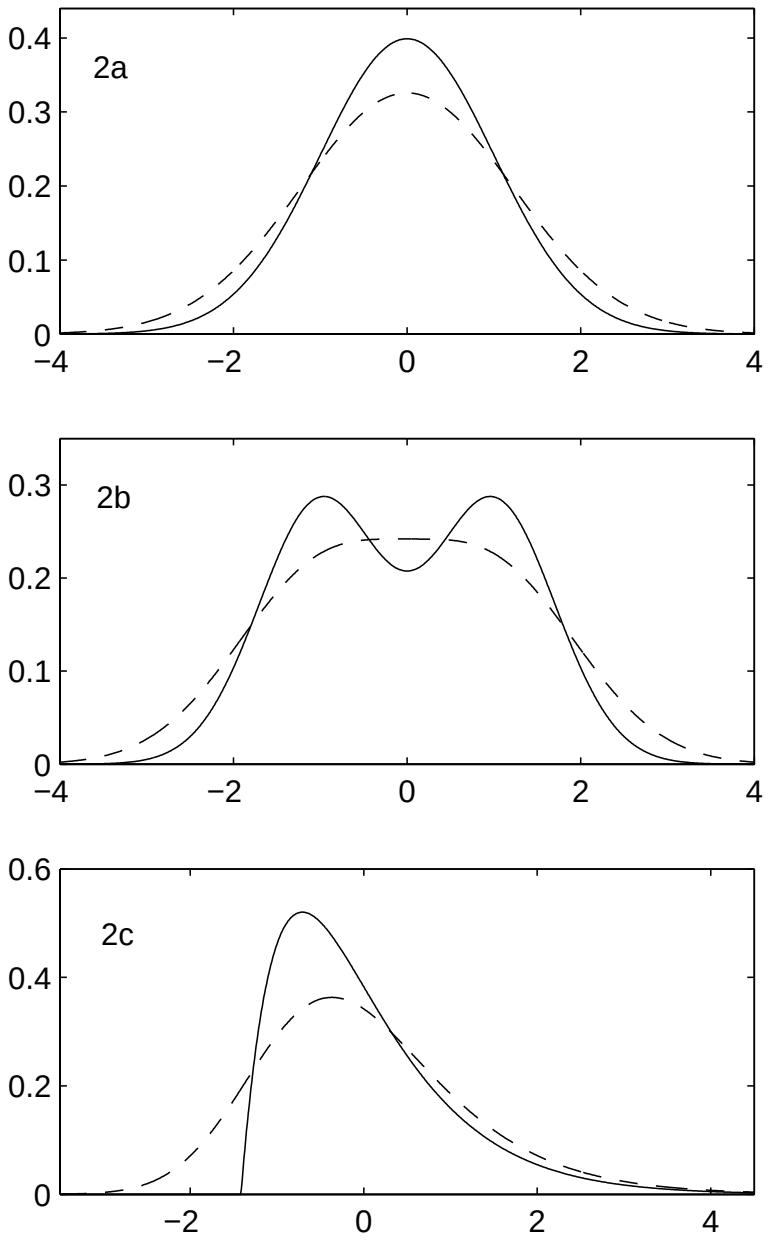

Figure 1.2: True-data and observed-data density functions. Solid line: True-data density; Dashed line: Observed-data density, the convolution of the true-data density with a N(0, 1/2) measurement error density. 2a: True-data density is N(0, 1); 2b: True-data density is a normal mixture with means ± 1 and common variances 1/2; 2c: True-data density is a standardized Chi-squared(4).

sequences λ converging to zero as $n \to \infty$. A popular measure of the overall performance of $\widehat{f}_x(x)$ is the mean integrated squared error (MISE), defined as

$$\text{MISE}\{\widehat{f}_x(x)\} = E\left[\int \{\widehat{f}_x(x) - f_x(x)\}^2 dx\right] = \int E\{\widehat{f}_x(x) - f_x(x)\}^2 dx.$$

The MISE admits the decomposition

$$\text{MISE}\{\widehat{f}_x(x)\} = \int \text{Var}\{\widehat{f}_x(x)\}dx + \int \text{Bias}^2\{\widehat{f}_x(x)\}dx.$$

Asymptotic analysis of the MISE provides insight into the behavior of $\widehat{f}_x(x)$. Assuming that $f_x(x)$ is twice-differentiable, as $n \to \infty$ and $\lambda \to 0$, the MISE of the kernel density estimator in equation (1.3.7) can be approximated by (Silverman 1986)

$$\text{MISE}\{\widehat{f}_x(x)\} \sim \frac{1}{n\lambda}\int Q(x)^2 dx + \frac{\lambda^4}{4}\mu_{Q,2}^2 \int \{f_x''(x)\}^2 dx. \tag{1.3.8}$$

The familiar trade-off between variance and bias is evident in this expression. Large values of λ reduce the variance, but increase the bias. Conversely, small values of λ reduce the bias at the expense of the variance. A challenging problem in density estimation is to find a sequence, $\{\lambda_n\}$, that minimizes the approximate MISE. The bandwidth defined by such a sequence is referred to as the optimal bandwidth, and as can be seen from equation (1.3.8), will depend upon the unknown functional $\int \{f_x''(x)\}^2 dx$. Parzen (1962) showed that the optimal bandwidth is given by

$$\lambda_{opt} = \mu_{q,2}^{-2/5}\left\{\int Q(x)^2 dx\right\}^{1/5}\left\{\int \{f_x''(x)\}^2 dx\right\}^{-1/5} n^{-1/5}.$$

Substitution into equation (1.3.8) shows that when the bandwidth is chosen optimally, the approximate MISE is (Silverman 1986)

$$\text{MISE}\{\widehat{f}_x(x)\} \approx \frac{5}{4}\mu_{q,2}^{2/5}\left\{\int Q(x)^2 dx\right\}^{4/5}\left\{\int \{f_x''(x)\}^2 dx\right\}^{1/5} n^{-4/5}. \tag{1.3.9}$$

It follows from equation (1.3.9) that the MISE of the kernel density estimator converges to 0 at the optimal rate of $n^{-4/5}$.

1.3.2 Density Deconvolution

Calculated with the observed-data sample, the kernel density estimator in equation (1.3.7) is inconsistent for $f_x(x)$. In this section we discuss deconvolution methods that consistently estimate $f_x(x)$ from the observed-data sample. In particular, we restrict our attention to those methods that derive from the kernel density estimator discussed in the previous section.

Deconvolution estimators require assumptions about the distribution of the measurement error variable, and are differentiated by the restrictiveness of the assumptions they make. Most methods assume that measurement errors are independent and identically distributed, and have a known density function, $f_u(u)$. Fewer methods exist that allow $f_u(u)$ to be unknown, or allow measurement errors to be non-identically distributed.

Deconvolution with Measurement Error Density Known

Stefanski and Carroll (1990) developed deconvoluting kernel density estimators for a large class of error distributions. These estimators assume that the measurement errors in equation (1.2.1) are independent and identically distributed and have a known density function.

Let $\Phi_w(t)$, $\Phi_x(t)$, and $\Phi_u(t)$ denote the characteristic functions of W, X, and

U respectively. When measurement errors are additive and independent, $\Phi_w(t) = \Phi_x(t)\Phi_u(t)$. It follows that the characteristic function of X is

$$\Phi_x(t) = \Phi_w(t)\Phi_u^{-1}(t).$$

An expression for $f_x(x)$ follows from substitution into the Fourier inversion formula,

$$f_x(x) = \frac{1}{2\pi}\int e^{-itx}\Phi_x(t)dt = \frac{1}{2\pi}\int e^{-itx}\Phi_w(t)\Phi_u^{-1}(t)dt. \tag{1.3.10}$$

Because $f_u(u)$ is known, $\Phi_u^{-1}(t)$ is known. Consider the kernel density estimator of $f_w(w)$, calculated from the observed-data sample,

$$\widehat{f}_w(w) = \frac{1}{n\lambda}\sum_{r=1}^{n} Q\left(\frac{w - W_r}{\lambda}\right).$$

$\widehat{f}_w(w)$ has characteristic function

$$\Phi_{\widehat{f}}(t) = \int e^{itw}\frac{1}{n\lambda}\sum_{r=1}^{n} Q\left(\frac{w - W_r}{\lambda}\right)dw.$$

For general kernel functions, when $\Phi_{\widehat{f}}(t)$ is substituted for $\Phi_w(t)$ in equation (1.3.10), the resulting integral is undefined. Stefanski and Carroll (1990) found kernel functions for which equation (1.3.10) is satisfied. Let $Q(x)$ be a bounded, even probability density function with characteristic function $\Phi_Q(t)$ that satisfies

$$sup\ _t\ |\Phi_Q(t)/\Phi_u(t/\lambda)| < \infty \ \text{ and } \ \int |\Phi_Q(t)/\Phi_u(t/\lambda)|dt < \infty$$

for each $\lambda > 0$. The deconvoluting kernel density estimator of $f_x(x)$ is given by

$$\widehat{f}_x(x) = \frac{1}{2\pi}\int e^{-itx}\Phi_{\widehat{f}}(t)\Phi_u^{-1}(t)dt.$$

This can be rewritten as

$$\widehat{f}_x(x) = \frac{1}{n\lambda} \sum_{r=1}^{n} Q_d \left(\frac{x - W_r}{\lambda} \right),$$

where

$$Q_d(x) = \frac{1}{2\pi} \int e^{-itx} \Phi_Q(t) \Phi_u^{-1}(t/\lambda) dt.$$

$Q_d(x)$ is called the deconvoluting kernel, and has the property that conditional on the true data,

$$E\{Q_d((x - W_r)/\lambda)|X_r\} = Q((x - X_r)/\lambda).$$

From this it follows that $\widehat{f}_x(x)$ is a conditionally unbiased estimator of the true-data kernel density estimator,

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

Therefore unconditionally, $\widehat{f}_x(x)$ has the same expectation and same bias as the true-data kernel density estimator.

Stefanski and Carroll (1990) showed that $\widehat{f}_x(x)$ is pointwise consistent for $f_x(x)$ and derived the following approximation to its integrated mean squared error,

$$\text{MISE}\{\widehat{f}_x(x)\} \sim (2\pi n\lambda)^{-1} \int \Phi_Q^2(t) |\Phi_u(t/\lambda)|^{-2} dt + (\lambda^4/4) \mu_{q,2}^2 \int \{f_x''(x)\}^2 dx$$

where $\mu_{q,2} = \int x^2 Q(x) dx < \infty$.

The asymptotic properties of deconvoluting kernel density estimators have been studied extensively. Stefanski and Carroll (1990) derived bounds on the integrated variance of $\widehat{f}_x(x)$ for several error distributions. Stefanski (1989) found optimal sequences of bandwidths and derived rates of convergence for the estimator. It was

shown by Carroll and Hall (1988) that the rates achieved by deconvoluting kernel density estimators are optimal among all deconvolution estimators. For more discussion of the properties of deconvoluting kernel estimators see Wand (1998), Fan (1991a, 1991b, 1992), Devroye (1989), and Liu and Taylor (1989).

Deconvolution when measurement errors are normally distributed is particularly important in practice. This case will be discussed in detail in Chapter 2.

Deconvolution with Measurement Error Density Unknown

Fewer methods exist for density deconvolution when the measurement error distribution is unknown. Diggle and Hall (1993) presented an estimator for $f_x(x)$ given independent observations of both W and U. Let $W_1, \ldots, W_{n_w}$ and $U_1, \ldots, U_{n_u}$ be independent samples from $f_w(w)$ and $f_u(u)$. Define their respective empirical characteristic functions as

$$\widehat{\Phi}_w(t) = \frac{1}{n_w} \sum_{r=1}^{n_w} e^{itW_r} \quad \text{and} \quad \widehat{\Phi}_u(t) = \frac{1}{n_u} \sum_{r=1}^{n_u} e^{itU_r}.$$

The following estimator is defined based on the Fourier inversion formula,

$$\widehat{f}_x(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{itx} d(x) \widehat{\Phi}_w(x) \widehat{\Phi}_u(x) dx, \tag{1.3.11}$$

where $d(x)$ is a damping function and controls the smoothness of the estimate. A special case of (1.3.11) is the truncated estimator,

$$\widehat{f}_x(x) = \frac{1}{2\pi} \int_{-p}^{p} e^{itx} \widehat{\Phi}_w(x) \widehat{\Phi}_u(x) dx. \tag{1.3.12}$$

The authors discuss selection of the truncation parameter, p, and derive asymptotic properties of (1.3.12).

Patil (1996) considered the case where independent replicate measurements are observed of each $X_1, \ldots, X_n$. For $r = 1, \ldots, n$ and $j = 1, \ldots, m$ suppose that X_r is measured as

$$W_{r,j} = X_r + U_{r,j},$$

where $U_{r,j}$ has constant variance σ^2.

Patil (1996) suggested substituting the means of the replicate measurements, $\overline{W}_1, \ldots, \overline{W}_n$ for the true data in the usual kernel density estimator,

$$\widehat{f}_x(x) = \frac{1}{n\lambda} \sum_{r=1}^{n} Q\left(\frac{x - \overline{W}_r}{\lambda}\right). \tag{1.3.13}$$

He derived asymptotic properties of this estimator, letting both $n \to \infty$ and $m \to \infty$. Under these conditions, the estimator in (1.3.13) converges pointwise to $f_x(x)$ at the rate of $n^{-4/5}$, the same rate as the true-data kernel density estimator.

1.4 Problem Description

In this dissertation, we present two new deconvolution estimators of $f_x(x)$. The first is developed for the case where measurement errors are known to be normally distributed, but have unknown and possibly nonconstant variances. Similar to Patil (1996), we consider the case where replicate measurements are observed of each $X_1, \ldots, X_n$. For $r = 1, \ldots, n$ and $j = 1, \ldots, m_r$, we suppose that X_r is observed as

$$W_{r,j} = X_r + U_{r,j},$$

where $U_{r,j}$ is a N(0, σ_r^2) measurement error. Replicate measurements are used to estimate the set of measurement error variances, $\sigma_1^2, \ldots, \sigma_n^2$, and the estimator that follows is a generalization of the Stefanski-Carroll deconvoluting kernel density estimator for normally distributed measurement error. We assume that the number of replicates, m_r, is fixed for all r, and derive asymptotic properties of the estimator as $n \to \infty$.

The second is a parametric deconvolution estimator. We assume a known functional relationship between X and a set of covariates, $\boldsymbol{Z}$. Estimates of the true-data values are derived from the regression of the observed data on the covariates, and are substituted into a kernel density estimator. We examine the asymptotic properties of the estimator under the assumption that both the regression errors and measurement errors are normally distributed, and that the measurement error variance is constant and known.

Chapter 2

Density Estimation with Replicate Measurements

2.1 Introduction

We consider the problem of estimating $f_x(x)$, the density function of a random variable X, from a set of replicate measurements. Suppose that a random sample from this density, $X_1, \ldots, X_n$, is measured independently and repeatedly as $\{W_{r,j}\}_{r=1,j=1}^{n,\ m_r}$, where $W_{r,j} = X_r + U_{r,j}$. We present an estimator for $f_x(x)$ under the assumption that for each $r = 1, \ldots, n$ and $j = 1, \ldots, m_r$, $U_{r,j}$ is a normally distributed measurement error having mean 0 and variance σ_r^2, and is independent of X_r.

Several authors have considered estimation of $f_x(x)$ when measurement errors are identically distributed with a known density function. Stefanski and Carroll (1990) 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, Wand, 1998; Carroll and Hall 1988; Devroye, 1989; Stefanski, 1990; Fan, 1991a, 1991b, 1992).

The problem of estimating $f_x(x)$ when errors are identically distributed with an unknown density function also has received some attention. Diggle and Hall (1993) developed an estimator of $f_x(x)$ given independent observations of both W and U. Patil (1996) replaced the true data in the usual kernel density estimator with the sample means of replicate measurements. When both the sample size and the number of replicates increase to infinity, he showed that the rate of convergence for this estimator is $n^{-4/5}$, the usual rate for kernel density estimators in the absence of measurement error.

We assume a fixed number of replicate measurements, and approach the problem through the conditional distributions of the sample means and sample variances of these measurements. An estimator arises naturally from the results of Stefanski et al. (2003). It can be viewed as a generalization of the deconvoluting estimator of Stefanski and Carroll (1990). Although measurement errors must be assumed normal, they need not be identically distributed. The estimator accommodates both heteroscedastic and homoscedastic measurement errors.

This chapter is organized as follows. In Section 2.2, we define our estimator and show its connection to the Stefanski-Carroll deconvoluting estimator. We derive and examine the estimator's asymptotic mean integrated squared error in Section 2.3, and in Section 2.4 we investigate its finite-sample properties via simulation. We take up

the problem of bandwidth estimation in Section 2.5, describing a bootstrap method to select a bandwidth for the estimator. In Section 2.6 we present an application to real data.

2.1.1 Background

Our estimator for the density function of X follows from the results of Stefanski et al. (2003). They 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 (2003).

Theorem 2.1.1.1 *Suppose $\widehat{\mu}$ and $\widehat{\sigma}^2$ are independent random variables such that $\widehat{\mu} \sim N(\mu,\ \tau\sigma^2)$ 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,*

$$\widehat{\theta} = E\left\{ g\left(\widehat{\mu} + i(\tau d)^{1/2}\widehat{\sigma}T\right) \,\middle|\, \widehat{\mu}, \widehat{\sigma}^2 \right\} \tag{2.1.1}$$

is the uniformly minimum variance unbiased estimator of $g(\mu)$ provided it has finite variance.

A few comments are necessary about the estimator $\widehat{\theta}$. Detailed proofs and discussion may be found in Stefanski et al. (2003) and Stefanski (1989). Although $g()$ is a

complex-valued function, $\widehat{\theta}$ is real-valued. Under the given conditions, the imaginary part of $g()$ has expectation 0. For most functions $g()$, however, finding a closed-form expression for the estimator in equation (2.1.1) remains difficult. A Monte Carlo approximation to $\widehat{\theta}$ is given by

$$\widehat{\theta}_B = \frac{1}{B}\sum_{b=1}^{B} Re\left\{g\left(\widehat{\mu} + i(\tau d)^{1/2}\widehat{\sigma}T_b\right)\right\}, \tag{2.1.2}$$

where $T_1, \ldots, T_B$ are independent replicates of the variable T.

The variance of $\widehat{\theta}$ in equation (2.1.1) is

$$\mathrm{Var}(\widehat{\theta}) = E[\widehat{\theta}^2 - E\{g^2(\widehat{\mu} + i(\tau d)^{1/2}\widehat{\sigma}T|\ \widehat{\mu}, \widehat{\sigma}^2)\}].$$

Because $g()$ is an entire function, $g^2()$ is an entire function. It follows that both $\widehat{\theta}$ and $E\{g^2(\widehat{\mu} + i(\tau d)^{1/2}\widehat{\sigma}T|\ \widehat{\mu}, \widehat{\sigma}^2)\}$ are real-valued.

The variance of the Monte-Carlo estimator in equation (2.1.2) is

$$\mathrm{Var}(\widehat{\theta}_B) = \mathrm{Var}(\widehat{\theta}) + E\{\mathrm{Var}(\widehat{\theta}_B|\ \widehat{\mu}, \widehat{\sigma}^2)\}$$

In general, $\mathrm{Var}(\widehat{\theta})$ and $\mathrm{Var}(\widehat{\theta}_B)$ will not have closed-form expressions. However, Monte Carlo methods can be used to approximate both quantities, and are described in Stefanski et. al (2003).

2.2 The Deconvolution Estimator

2.2.1 Estimation of $f_x(x)$ when Measurement Errors are Heteroscedastic

We now show how the estimator in Section 2.1.1 can be used to estimate the density function of a random variable X that is measured with normally distributed error having unknown, nonconstant variance. We assume that each value of X is measured repeatedly, and that the measurement error variance is constant among replicates but may differ between replicates. First, suppose that m independent, replicate measurements of a particular value of X are observed as $\{W_j\}_{j=1}^m$ where $W_j = X + U_j$. Further assume that $U_1, \ldots, U_m$ are independent of X and are normally distributed with mean 0 and unknown variance σ^2.

Let $\overline{W}$ and $\widehat{\sigma}^2$ be the sample mean and sample variance of the m measurements of X. Then conditional on X, the following are true: 1) $\overline{W}$ and $\widehat{\sigma}^2$ are independent random variables, 2) $\overline{W}$ follows a $\mathrm{N}(X,\ \sigma^2/m)$ distribution, and 3) $(m-1)\widehat{\sigma}^2/\sigma^2$ follows a Chi-squared$(m-1)$ distribution. Let $Q(x)$ be a probability density function that is entire over the complex plane, and define $T = Z_1/\sqrt{Z_1^2 + \cdots + Z_m^2}$, where $Z_1, \ldots, Z_m$ are independent and identically distributed N(0, 1) random variables. It follows from Theorem 2.1.1.1 that

$$\widehat{\theta} = E\left\{ \frac{1}{\lambda}\, Q\left(\frac{x - \left\{ \overline{W} + i(\frac{m-1}{m})^{1/2} \widehat{\sigma} T \right\}}{\lambda} \right) \,\middle|\, \overline{W}, \widehat{\sigma}^2 \right\}$$

is, conditional on X, an unbiased estimator of $\lambda^{-1}Q\{(x-X)/\lambda\}$, i.e.,

$$E(\widehat{\theta}\,|X) = \frac{1}{\lambda}\,Q\left(\frac{x-X}{\lambda}\right).$$

From here it is straightforward to construct the estimator of $f_x(x)$. Suppose that the variables $X_1,\ldots,X_n$ are measured independently and repeatedly as $\{W_{r,j}\}_{r=1,j=1}^{n,\ m_r}$ where $W_{r,j} = X_r + U_{r,j}$. For each $r = 1,\ldots,n$, let the measurement errors $U_{r,j},\ j = 1,\ldots,m_r$, be independent of X_r and normally distributed with mean 0 and possibly unequal variances σ_r^2. Then conditional on $X_1,\ldots,X_n$, the sample estimates $\overline{W}_1,\ldots,\overline{W}_n,\ \widehat{\sigma}_1^2,\ldots,\widehat{\sigma}_n^2$ are mutually independent with $\overline{W}_r \sim \mathrm{N}\,(X_r,\ \sigma_r^2/m_r)$ and $((m_r-1)\widehat{\sigma}_r^2/\sigma_r^2) \sim$ Chi-squared(m_r-1). 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. Then the estimator

$$\widehat{f}_x(x) = \frac{1}{n\lambda}\sum_{r=1}^{n} E\left\{Q\left(\frac{x-\left\{\overline{W}_r + i(\frac{m_r-1}{m_r})^{1/2}\widehat{\sigma}_r T_{r,m_r-1}\right\}}{\lambda}\right)\,\middle|\,\overline{W}_r,\widehat{\sigma}_r^2\right\} \qquad (2.2.3)$$

is such that

$$E\left\{\widehat{f}_x(x)|X_1,\cdots,X_n\right\} = \frac{1}{n\lambda}\sum_{r=1}^{n} Q\left(\frac{x-X_r}{\lambda}\right). \qquad (2.2.4)$$

The quantity in equation (2.2.4) is the familiar kernel density estimator of $f_x(x)$. From this it follows that unconditionally, $\widehat{f}_x(x)$ has the same expectation and same bias as the kernel density estimator constructed from the true, unobserved data.

Connection to the Deconvoluting Kernel Density Estimator

The estimator in equation (2.2.3) generalizes the deconvoluting kernel density estimator proposed by Stefanski and Carroll (1990). They considered the case where

observations $\{W_r\}_{r=1}^n$ are such that $W_r = X_r + U_r$, and $\{U_r\}_{r=1}^n$ are independent and identically distributed random variables with known characteristic function $\Phi_u(t) = E(e^{itu})$, and are independent of $\{X_r\}_{r=1}^n$. 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 $\Phi_Q(t)$. Define

$$\Phi_{\hat{f}}(t) = \int e^{itw} \frac{1}{n\lambda} \sum_{r=1}^{n} Q\left(\frac{w - W_r}{\lambda}\right) dw.$$

$\Phi_{\hat{f}}(t)$ is the characteristic function of the kernel density estimate of the density of W. The deconvoluting kernel density estimator of $f_x(x)$ is given by

$$\widehat{g}_x(x) = \frac{1}{2\pi} \int e^{-itx} \Phi_{\hat{f}}(t) \Phi_u^{-1}(t) dt. \tag{2.2.5}$$

This may be rewritten as

$$\widehat{g}_x(x) = \frac{1}{n\lambda} \sum_{r=1}^{n} Q_d\left(\frac{x - W_r}{\lambda}, \lambda\right), \tag{2.2.6}$$

where

$$Q_d(z, \lambda) = \frac{1}{2\pi} \int e^{-itz} \Phi_Q(t) \Phi_u^{-1}(t/\lambda) dt. \tag{2.2.7}$$

$Q_d(z, \lambda)$ is called the deconvoluting kernel.

For the case where measurement errors are independent and identically distributed $\mathrm{N}(0,\ \sigma^2)$ random variables, we write the expression in (2.2.7) as

$$Q_d(z, \lambda, \sigma) = \frac{1}{2\pi} \int e^{-itz} e^{t^2\sigma^2/2\lambda^2} \Phi_Q(t) dt. \tag{2.2.8}$$

Now consider again $\widehat{f_x}(x)$ in equation (2.2.3). Define the function

$$Q_r^*(z, \lambda, m_r, \widehat{\sigma}_r) = E\left\{ Q\left(z - \frac{i(\frac{m_r - 1}{m_r})^{1/2} \widehat{\sigma}_r T_{r, m_r - 1}}{\lambda} \right) \middle| \overline{W}_r, \widehat{\sigma}_r^2 \right\}, \tag{2.2.9}$$

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

$$\widehat{f}_x(x) = \frac{1}{n\lambda} \sum_{r=1}^{n} Q_r^* \left(\frac{x - \overline{W}_r}{\lambda}, \lambda, m_r, \widehat{\sigma}_r \right). \tag{2.2.10}$$

The connection between the estimator $\widehat{f}_x(x)$ in equation (2.2.10) and the Stefanski-Carrol estimator in equation (2.2.6) 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 equation (2.2.8) relies upon knowledge of the inverse of the characteristic function of this measurement error, $\Phi_{u_r}^{-1}(t) = e^{t^2\sigma_r^2/2m_r\lambda^2}$. However, when σ_r^2 is unknown, so too is this function. As we show next, for each $r = 1, \ldots, n$, $Q_r^*(z, \lambda, m_r, \widehat{\sigma}_r)$ unbiasedly estimates $Q_d(z, \lambda, \sigma_r^2/m_r)$ through unbiased estimation of $\Phi_{u_r}^{-1}(t)$, i.e.,

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

It will follow from this that evaluated with $\overline{W}_r$,

$$E\left\{ Q_r^* \left(\frac{x - \overline{W}_r}{\lambda}, \lambda, m_r, \widehat{\sigma}_r \right) \right\} = E\left\{ Q_d \left(\frac{x - \overline{W}_r}{\lambda}, \lambda, \frac{\sigma_r^2}{m_r} \right) \right\},$$

for each $r = 1, \ldots, n$. With equations (2.2.6) and (2.2.10), it will follow that $\widehat{g}_x(x)$ and $\widehat{f}_x(x)$ have the same expectation, which from the key property in equation (2.2.4) is the expectation of the true-data kernel density estimator.

First, applying the Fourier inversion formula and interchanging the operations of

integration and expectation in equation (2.2.9) yields

$$Q_r^*\left(z,\lambda,m_r,\widehat{\sigma}_r\right)=\frac{1}{2\pi}\int e^{-itz}E\left\{\exp\left(\frac{-t}{\lambda}\left(\frac{m_r-1}{m_r}\right)^{1/2}\widehat{\sigma}_r T_{r,m_r-1}\right)\,\middle|\,\overline{W}_r,\widehat{\sigma}_r^2\right\}\Phi_Q(t)dt. \tag{2.2.11}$$

Because T_{r,m_r-1} is independent of the data, the conditional expectation in (2.2.11) 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)

$$\Phi_T(\theta(t))=\Gamma\left(\frac{m_r-1}{2}\right)J_{\frac{m_r-1}{2}-1}(\theta(t))\left(\frac{\theta(t)}{2}\right)^{1-\frac{m_r-1}{2}}. \tag{2.2.12}$$

Here $J_\nu(z)$ denotes a Bessel function of the first kind,

$$J_\nu(z)=\sum_{k=0}^{\infty}\frac{(-1)^k}{k!\,\Gamma\left(\nu+k+1\right)}\left(\frac{z}{2}\right)^{\nu+2k}. \tag{2.2.13}$$

Substituting equation (2.2.12) into equation (2.2.11) yields

$$\begin{aligned}Q_r^*\left(z,\lambda,m_r,\widehat{\sigma}_r\right) &= \frac{1}{2\pi}\int e^{-itz}\Gamma\left(\frac{m_r-1}{2}\right)J_{\frac{m_r-1}{2}-1}\left(\theta(t)\right)\left\{\frac{\theta(t)}{2}\right\}^{1-\frac{m_r-1}{2}}\Phi_Q(t)dt\\ &= \frac{1}{2\pi}\int e^{-itz}\Phi_Q(t)\widehat{\Phi}_{u_r}^{-1}(t/\lambda)dt,\end{aligned} \tag{2.2.14}$$

where

$$\begin{aligned}\widehat{\Phi}_{u_r}^{-1}(t/\lambda) &= \Gamma\left(\frac{m_r-1}{2}\right)J_{\frac{m_r-1}{2}-1}\left(\theta(t)\right)\left\{\frac{\theta(t)}{2}\right\}^{1-\frac{m_r-1}{2}}\\ &= \sum_{k=0}^{\infty}\frac{\left(\frac{t^2}{4\lambda^2}\frac{m_r-1}{m_r}\widehat{\sigma}_r^2\right)^k\Gamma\left(\frac{m_r-1}{2}\right)}{k!\,\Gamma\left(k+\frac{m_r-1}{2}\right)}.\end{aligned} \tag{2.2.15}$$

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 $\widehat{\Phi}_{u_r}^{-1}(t/\lambda)$. Finding the expectation of $Q_r^*(z, \lambda, m_r, \widehat{\sigma}_r)$ only requires finding the expectation of $\widehat{\Phi}_{u_r}^{-1}(t/\lambda)$. From the relationship

$$E\{\widehat{\sigma}_r^{2k}\} = \frac{\Gamma\left(k + \frac{m_r-1}{2}\right)\left(\frac{2\sigma_r^2}{m_r-1}\right)^k}{\Gamma\left(\frac{m_r-1}{2}\right)}, \tag{2.2.16}$$

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 $\mathrm{N}\left(0,\ \sigma_r^2/m_r\right)$ random variable evaluated at t/λ. Substituting this expression into equation (2.2.14) 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 equation (2.2.8) reveals that $Q_r^*(z, \lambda, m_r, \widehat{\sigma}_r)$ is an unbiased estimator of $Q_d(z, \lambda, \sigma^2/m_r)$. Finally, noting that the expectation in equation (2.2.11) is conditional on $\overline{W}_r$, it follows that $Q_r^*((x-\overline{W}_r)/\lambda, \lambda, m_r, \widehat{\sigma}_r)$ and $Q_d((x-\overline{W}_r)/\lambda, \lambda, \sigma^2/m_r)$ have the same expectation. Therefore, from equations (2.2.6), (2.2.10), and (2.2.4),

$$E\{\widehat{g}_x(x)\} = E\{\widehat{f}_x(x)\} = E\left\{\frac{1}{n\lambda}\sum_{r=1}^{n} Q\left(\frac{x - X_r}{\lambda}\right)\right\}.$$

2.2.2 Estimation of $f_x(x)$ when Measurement Errors are Homoscedastic

The results of the previous section also apply to the case where the measurement error variance is constant, $\sigma_1^2 = \cdots = \sigma_n^2 = \sigma^2$, and data are pooled to estimate the

common variance σ^2. Next, we present an estimator for $f_x(x)$ under this assumption and show that it, too, generalizes the Stefanski-Carroll estimator.

Let $\{U_{r,j}\}_{r=1,j=1}^{n,\ m_r}$ be independent and identically distributed $\mathrm{N}(0,\ \sigma^2)$ random variables, independent of $\{X_r\}_{r=1}^n$, and let $\widehat{\sigma}^2$ be the pooled estimator of the measurement error variance based on $d = \sum_{r=1}^n (m_r - 1)$ degrees of freedom,

$$\widehat{\sigma}^2 = \frac{1}{d} \sum_{r=1}^{n} (m_r - 1)\widehat{\sigma}_r^2. \tag{2.2.17}$$

Given $X_1, \ldots, X_n$, the sample means $\overline{W}_1, \ldots, \overline{W}_n$ are distributed as independent $\mathrm{N}(X_r,\ \sigma^2/m_r)$ random variables, and $d\widehat{\sigma}^2/\sigma^2$ is distributed as a Chi-squared(d) random variable, independently of $\overline{W}_1, \ldots, \overline{W}_n$. Let $T_{r,d} = Z_{r,1}/\sqrt{Z_{r,1}^2 + \cdots + Z_{r,d}^2}$ where the $Z_{r,j}$ are independent $\mathrm{N}(0,\ 1)$ random variables. It follows from Theorem 2.1.1.1 that the estimator

$$\widehat{\theta} = E\left\{ \frac{1}{\lambda}\, Q\left(\frac{x - \left\{ \overline{W} + i(\frac{d}{m_r})^{1/2} \widehat{\sigma} T_{r,d} \right\}}{\lambda} \right) \,\middle|\, \overline{W}, \widehat{\sigma}^2 \right\}$$

is conditionally unbiased for $\lambda^{-1} Q\{(x - X_r)/\lambda\}$, i.e.,

$$E(\widehat{\theta}\,|X_r) = \frac{1}{\lambda}\, Q\left(\frac{x - X_r}{\lambda} \right).$$

Thus the estimator defined as

$$\widehat{f}_x(x) = \frac{1}{n\lambda} \sum_{r=1}^{n} E\left\{ Q\left(\frac{x - \left\{ \overline{W}_r + i(\frac{d}{m_r})^{1/2} \widehat{\sigma} T_{r,d} \right\}}{\lambda} \right) \,\middle|\, \overline{W}_r, \widehat{\sigma}^2 \right\} \tag{2.2.18}$$

has the key property in equation (2.2.4), that conditional on the true data, it is an unbiased estimator of the true-data kernel density estimator, just as in the case of heteroscedastic measurement errors.

To see that this estimator is also a generalization of the Stefanski-Carrol deconvoluting kernel density estimator, $\widehat{g}_x(x)$ in equation (2.2.6), first define the function

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

so that equation (2.2.18) can be written as

$$\widehat{f}_x(x)=\frac{1}{n\lambda}\sum_{r=1}^{n}Q_r^\dagger\left(\frac{x-\overline{W}_r}{\lambda},\lambda,d,m_r,\widehat{\sigma}\right). \tag{2.2.19}$$

The functions $Q_r^\dagger\left(z,\lambda,d,m_r,\widehat{\sigma}\right)$ and $Q_r^*\left(z,\lambda,m_r,\widehat{\sigma}_r\right)$ in equation (2.2.9) are related via

$$Q_r^*\left(z,\lambda,m_r,\widehat{\sigma}_r\right)=Q_r^\dagger\left(z,\lambda,m_r-1,m_r,\widehat{\sigma}_r\right).$$

We use the arguments of the previous section to show that $Q_r^\dagger\left(z,\lambda,d,m_r,\widehat{\sigma}\right)$ is an unbiased estimator of $Q_d(z,\lambda,\sigma^2/m_r)$. As in equation (2.2.11) we have the alternative expression for $Q_r^\dagger\left(z,\lambda,d,m_r,\widehat{\sigma}\right)$,

$$Q_r^\dagger\left(z,\lambda,d,m_r,\widehat{\sigma}\right)=\frac{1}{2\pi}\int e^{-itz}E\left\{\exp\left(\frac{-t}{\lambda}\left(\frac{d}{m_r}\right)^{1/2}\widehat{\sigma}T_{r,d}\right)\,\middle|\,\overline{W}_r,\widehat{\sigma}^2\right\}\Phi_Q(t)dt. \tag{2.2.20}$$

The expectation in equation (2.2.20) is the characteristic function of $T_{r,d}$. Here, $T_{r,d}=\boldsymbol{T}^T\boldsymbol{e}_1$, where $\boldsymbol{T}^T$ is a random vector uniformly distributed on the d-dimensional unit sphere, and $\boldsymbol{e}_1$ is the $d\times 1$ dimensional unit vector having a one in the first position. Substituting this characteristic function, defined analogous to equation (2.2.12), into equation (2.2.20) and simplifying leaves

$$Q_r^\dagger\left(z,\lambda,d,m_r,\widehat{\sigma}\right)=\frac{1}{2\pi}\int e^{-itz}\Phi_Q(t)\widehat{\Phi}_{u_r}^{-1}(t/\lambda)dt,$$

where

$$\widehat{\Phi}_{u_r}^{-1}(t/\lambda) = \sum_{k=0}^{\infty} \frac{\left(\frac{t^2}{4\lambda^2}\frac{d}{m_r}\widehat{\sigma}^2\right)^k \Gamma\left(\frac{d}{2}\right)}{k!\,\Gamma\left(k+\frac{d}{2}\right)}.$$

To evaluate the expectation of $Q_r^{\dagger}\left(z, \lambda, d, m_r, \widehat{\sigma}\right)$ note that

$$E\{\widehat{\sigma}^{2k}\} = \frac{\Gamma\left(k+\frac{d}{2}\right)\left(\frac{2\sigma^2}{d}\right)^k}{\Gamma\left(\frac{d}{2}\right)},$$

and so

$$E\left\{\widehat{\Phi}_u^{-1}(t/\lambda)\right\} = e^{t^2\sigma^2/2m_r\lambda^2}.$$

Thus,

$$E\{Q_r^{\dagger}\left(z, \lambda, d, m_r, \widehat{\sigma}\right)\} = \frac{1}{2\pi}\int e^{-itz}e^{t^2\sigma^2/2m_r\lambda^2}\Phi_Q(t)dt. \tag{2.2.21}$$

Because

$$Q_d(z, \lambda, \sigma^2/m_r) = \frac{1}{2\pi}\int e^{-itz}e^{t^2\sigma^2/2m_r\lambda^2}\Phi_Q(t)dt, \tag{2.2.22}$$

the quantity on the right-hand side of equation (2.2.21) is the Stefanski-Carroll deconvoluting kernel defined for the case of $\mathrm{N}(0,\ \sigma^2/m_r)$ measurement errors. Again it follows that

$$E\left\{Q_r^{\dagger}\left(\frac{x-\overline{W}_r}{\lambda}, \lambda, d, m_r, \widehat{\sigma}\right)\right\} = E\left\{Q_d\left(\frac{x-\overline{W}_r}{\lambda}, \lambda, \frac{\sigma^2}{m_r}\right)\right\},$$

so that from equations (2.2.6) and (2.2.19), $\widehat{f}_x(x)$ is seen to have the same expectation as the Stefanski-Carroll deconvoluting estimator.

2.2.3 Monte Carlo Estimation

In most cases, the estimators in equations (2.2.3) and (2.2.18) do not have simple closed-form expressions. However, a Monte Carlo approximation to the condi-

tional expectations in equations (2.2.3) and (2.2.18) follows directly from equation (2.1.2). Consider equation (2.2.3), the expression for $\widehat{f}_x$ under the assumption of heteroscedastic measurement errors. 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 $\widehat{f}_x(x)$,

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

Note that when $m_r = 2$, $T_{r,b}$ is equal to either 1 or -1. Thus in the special case of two replicate measurements, the conditional expectation in $\widehat{f}_x$ may be evaluated directly, and

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

In the case of homoscedastic measurement error, the Monte Carlo version of (2.2.18) is

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

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

2.3 Mean Integrated Squared Error

In this section, we derive expressions for the mean integrated squared error (MISE) of the estimators in equations (2.2.3) and (2.2.18). A main objective is to compare

the asymptotic properties of these estimators 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 Stefanski-Carroll deconvoluting kernel density estimator in equation (2.2.5) achieves this rate of convergence (Stefanski & Carroll 1990), and does so with the optimal bandwidth of $\lambda = \sigma\{\log(n)\}^{-1/2}$ (Stefanski 1990).

We begin by examining the estimator in equation (2.2.3) for the case of heteroscedastic measurement errors. For each X_r, $r = 1, \ldots, n$, we assume that a fixed number of replicate measurements, $m_r \geq 2$, are observed and used to estimate σ_r^2. For this case, the independence of the summands in equation (2.2.3) simplifies the derivation of the MISE and results in closed-form expressions for special cases of m_r. In practice, when replicate measurements are available, it is likely that the number of replicates is small. We examine in detail the case of two replicate measurements, i.e., $m_r = 2$ for all r. We also present a weighted version of this estimator, a natural extension in the case of nonconstant variances, where weights are selected to minimize the asymptotic integrated variance.

Finally, we discuss the asymptotic analysis of the estimator in equation (2.2.18) for the case of homoscedastic measurement errors. Here measurements are pooled to estimate the common measurement error variance σ^2, and so the summands in equation (2.2.18) are not independent. We use an approximation to this estimator in the examination of its asymptotic properties.

2.3.1 Heteroscedastic Measurement Errors: Estimation of the Sequence of Variances $\sigma_1^2, \ldots, \sigma_n^2$

Consider the heteroscedastic-variance estimator, $\widehat{f}_x(x)$ in equation (2.2.3). The mean integrated squared error of $\widehat{f}_x(x)$ has the representation

$$\text{MISE}\{\widehat{f}_x(x)\} = \int \text{Var}\{\widehat{f}_x(x)\}dx + \int \text{Bias}^2\{\widehat{f}_x(x)\}dx. \tag{2.3.23}$$

From the key property in equation (2.2.4), $\widehat{f}_x(x)$ has the same expectation as the true-data kernel density estimator. Therefore it has the same bias and same integrated squared bias as the true-data kernel density estimator, and as $\lambda \to 0$,

$$\int \text{Bias}^2\{\widehat{f}_x(x)\}dx \sim \frac{\lambda^4}{4}\mu_{Q,2}^2 \int \{f_x''(x)\}^2 dx, \tag{2.3.24}$$

where $\mu_{Q,2}^2 = \int z^2 Q(z)dz$. The integrated variance of $\widehat{f}_x(x)$ can be partitioned as

$$\int \text{Var}\{\widehat{f}_x(x)\}dx = \int E\{\widehat{f}_x^2(x)\}dx - \int \{E(\widehat{f}_x(x))\}^2 dx.$$

Determination of the latter term in the partition also follows from results for the true-data kernel density estimator,

$$\int \{E(\widehat{f}_x(x))\}^2 dx = \frac{1}{2\pi}\int |\Phi_f(t)|^2 \Phi_Q^2(\lambda t)dt,$$

where $\Phi_f(t)$ is the characteristic function of $f_x(x)$.

Determination of the former term is more involved as direct evaluation of $E\{\widehat{f}_x^2(x)\}$ is difficult. Instead, we approach the problem of finding an expression for $\int \text{Var}\{\widehat{f}_x(x)\}dx$ through the alternative expression for $\widehat{f}_x(x)$ in equation (2.2.10). Using this repre-

sentation, the integrated variance is

$$\int \mathrm{Var}\{\widehat{f}_x(x)\}dx = \frac{1}{n^2\lambda^2}\sum_{r=1}^{n}\int \mathrm{Var}\left\{Q^*\left(\frac{x-\overline{W}_r}{\lambda},\lambda,m_r,\widehat{\sigma}_r\right)\right\}dx,$$

which can be partitioned as $V_1 - V_2$ where

$$V_1 = \frac{1}{n^2\lambda^2}\sum_{r=1}^{n}\int E\left\{\left[Q^*\left(\frac{x-\overline{W}_r}{\lambda},\lambda,m_r,\widehat{\sigma}_r\right)\right]^2\right\}dx, \tag{2.3.25}$$

and

$$V_2 = \frac{1}{n^2\lambda^2}\sum_{r=1}^{n}\int \left\{E\left[Q^*\left(\frac{x-\overline{W}_r}{\lambda},\lambda,m_r,\widehat{\sigma}_r\right)\right]\right\}^2 dx. \tag{2.3.26}$$

Evaluation of V_2 is straightforward. Because

$$E\left[Q^*\left(\frac{x-\overline{W}_r}{\lambda},\lambda,m_r,\widehat{\sigma}_r\right)\right] = E\left\{Q\left(\frac{x-X_r}{\lambda}\right)\right\} = \lambda\int Q(z)f_x(x-\lambda z)dz,$$

it follows from substitution into equation (2.3.26) and an application of Parseval's Identity that

$$\begin{aligned} V_2 &= \frac{1}{n}\int\left\{\int Q(z)f_x(x-\lambda z)dz\right\}^2 dx \\ &= \frac{1}{2\pi n}\int |\Phi_f(t)|^2\Phi_Q^2(\lambda t)dt. \end{aligned} \tag{2.3.27}$$

Now consider the expression for V_1. After making a change of variables with $z = (x-\overline{W}_r)/\lambda$ and interchanging the operations of expectation and integration, V_1 becomes

$$V_1 = \frac{1}{n^2\lambda}\sum_{r=1}^{n} E\int \left\{Q^*\left(z,\lambda,m_r,\widehat{\sigma}_r\right)\right\}^2 dz.$$

From equation (2.2.14) and Parseval's Identity,

$$\int \left\{Q^*\left(z,\lambda,m_r,\widehat{\sigma}_r\right)\right\}^2 dx = \frac{1}{2\pi}\int \Phi_Q^2(t)\widehat{\Phi}_{u_r}^{-2}(t/\lambda)dt. \tag{2.3.28}$$

$\Phi_Q(t)$ is the known characteristic function of the density $Q(x)$ and does not depend on any random quantities. Thus, evaluating the expectation in equation (2.3.25) only requires finding the expectation of $\widehat{\Phi}_{u_r}^{-2}(t/\lambda)$. From equation (2.2.15),

$$\widehat{\Phi}_{u_r}^{-2}(t/\lambda) = \left\{ \sum_{k=0}^{\infty} \frac{\Gamma\left(\frac{m_r-1}{2}\right)\left(\frac{t^2}{4\lambda^2}\frac{m_r-1}{m_r}\widehat{\sigma}_r^2\right)^k}{\Gamma\left(k+\frac{m_r-1}{2}\right)k!} \right\}^2 .$$

The following notation will be useful. Define

$$(a)_k = \frac{\Gamma(a+k)}{\Gamma(a)}, \tag{2.3.29}$$

and denote the generalized hypergeometric series by (Erdelyi 1953)

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

We also make use of the relationships

$$\{{}_0F_1(b;x)\}^2 = {}_2F_3(b,b-\frac{1}{2};b,b,2b-1;4x) = {}_1F_2(b-\frac{1}{2};b,2b-1;4x). \tag{2.3.31}$$

A proof of the first equality appears in Bailey (1928); the second follows upon simplification after substitution into equation (2.3.30).

Using this notation we write equation (2.2.15) as

$$\widehat{\Phi}_{u_r}^{-1}(t/\lambda) = {}_0F_1\left(\frac{(m_r-1)}{2};\frac{t^2\widehat{\sigma}_r^2(m_r-1)}{4m_r\lambda^2}\right),$$

and so it follows from equation (2.3.31) that

$$\begin{aligned}\widehat{\Phi}_{u_r}^{-2}(t/\lambda) &= {}_1F_2\left(\frac{(m_r-2)}{2};\frac{(m_r-1)}{2},m_r-2;\frac{t^2\widehat{\sigma}_r^2(m_r-1)}{m_r\lambda^2}\right)\\ &= \sum_{k=0}^{\infty}\frac{\left(\frac{m_r-2}{2}\right)_k\left(\frac{t^2\widehat{\sigma}_r^2(m_r-1)}{m_r\lambda^2}\right)^k}{\left(\frac{m_r-1}{2}\right)_k(m_r-2)_k\,k!}.\end{aligned} \tag{2.3.32}$$

Similarly we write equation (2.2.16) as

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

and evaluate the expectation of $\widehat{\Phi}_{u_r}^{-2}(t/\lambda)$ in equation (2.3.32),

$$E\left\{\widehat{\Phi}_{u_r}^{-2}(t/\lambda)\right\} = \sum_{k=0}^{\infty} \frac{\left(\frac{m_r-2}{2}\right)_k \left(\frac{2t^2\sigma_r^2}{m_r\lambda^2}\right)^k}{(m_r-2)_k\, k!}. \tag{2.3.33}$$

Thus, after evaluating the expectation of equation (2.3.28), equation (2.3.25) becomes,

$$V_1 = \frac{1}{2\pi n^2 \lambda} \sum_{r=1}^{n} \int \Phi_Q^2(t) \left\{ \sum_{k=0}^{\infty} \frac{\left(\frac{m_r-2}{2}\right)_k \left(\frac{2t^2\sigma_r^2}{m_r\lambda^2}\right)^k}{(m_r-2)_k\, k!} \right\} dt. \tag{2.3.34}$$

For $\lambda \to 0$ as $n \to \infty$ it is apparent from equations (2.3.27) and (2.3.34) that V_2 is negligible compared to V_1. Thus keeping the leading variance term and the bias term in (2.3.23) we have that for $n \to \infty$ and $\lambda \to 0$,

$$\begin{aligned} \mathrm{MISE}\{\widehat{f}_x(x)\} &\sim \frac{1}{2\pi n^2 \lambda} \sum_{r=1}^{n} \int \Phi_Q^2(t) \left\{ \sum_{k=0}^{\infty} \frac{\left(\frac{m_r-2}{2}\right)_k \left(\frac{2t^2\sigma_r^2}{m_r\lambda^2}\right)^k}{(m_r-2)_k\, k!} \right\} dt \\ &+ \frac{\lambda^4}{4} \mu_{Q,2}^2 \int \{f_x''(x)\}^2 dx. \end{aligned}$$

The infinite sum makes the asymptotic behavior of this expression difficult to examine for general sequences m_r, $r = 1, 2, \ldots$. By the ratio test, this sum converges for $m_r \geq 2$ because

$$\lim_{k\to\infty} \left\{ \frac{\left(\frac{m_r-2}{2}\right)_{k+1} \left(\frac{2t^2\sigma_r^2}{m_r\lambda^2}\right)^{k+1}}{(m_r-2)_{k+1}\,(k+1)!} \right\} \left\{ \frac{\left(\frac{m_r-2}{2}\right)_k \left(\frac{2t^2\sigma_r^2}{m_r\lambda^2}\right)^k}{(m_r-2)_k\, k!} \right\}^{-1} = \lim_{k\to\infty} \frac{\left(\frac{m_r-2}{2}+k\right)\left(\frac{2t^2\sigma_r^2}{m_r\lambda^2}\right)}{(m_r-2+k)(k+1)},$$

which is equal to 0. Closed-form expressions of the sum can be obtained for special cases of m_r. We focus the remainder of our analysis on the important case of two replicate measurements, i.e., $m_r = 2$ for all r.

Heteroscedastic Measurement Errors with Variances Estimated from Two Replicate Measurements

When two replicate measurements are observed for each X_r, $r = 1, \ldots, n$, the expression for V_1 in equation (2.3.34) simplifies greatly, as a result of the following lemma.

Lemma 2.3.1.1 *Let $m = 2$ and let u be any real number. Then*

$$\sum_{k=0}^{\infty} \frac{\left(\frac{m-2}{2}\right)_k u^k}{(m-2)_k \, k!} = \frac{1}{2}(1 + e^u).$$

Proof. Let $m > 0$ and let k be an integer greater than 0. Then $(m)_0 = 1$ and $(m+1)_k = (m)_{k-1}$ from equation (2.3.29). Then,

$$\frac{\left(\frac{m-2}{2}\right)_k}{(m-2)_k} = \frac{\frac{1}{2}\left(\frac{m}{2}\right)_{k-1}}{(m-1)_{k-1}}.$$

Therefore,

$$\begin{aligned} \sum_{k=0}^{\infty} \frac{\left(\frac{m-2}{2}\right)_k u^k}{(m-2)_k \, k!} &= 1 + \sum_{k=1}^{\infty} \frac{\left(\frac{m-2}{2}\right)_k u^k}{(m-2)_k \, k!} \\ &= 1 + \frac{1}{2} \sum_{k=1}^{\infty} \frac{\left(\frac{m}{2}\right)_{k-1} u^k}{(m-1)_{k-1} \, k!} \end{aligned}$$

which, substituting $m = 2$, becomes

$$1 + \frac{1}{2} \sum_{k=1}^{\infty} \frac{u^k}{k!} = \frac{1}{2}(1 + e^u).$$

□

Thus when $m_r = 2$ for all r, the expectation in equation (2.3.33) simplifies to

$$E\left\{\widehat{\Phi}_{u_r}^{-2}(t/\lambda)\right\} = \frac{1}{2}e^{t^2\sigma_r^2/\lambda^2} + \frac{1}{2},$$

and V_1 in equation (2.3.34) becomes

$$\begin{aligned} V_1 &= \frac{1}{4\pi n^2\lambda}\sum_{r=1}^{n}\int \Phi_Q^2(t)e^{t^2\sigma_r^2/\lambda^2}dt + \frac{1}{4\pi n\lambda}\int \Phi_Q^2(t)dt \\ &= \frac{1}{4\pi n^2\lambda}\sum_{r=1}^{n}\int \Phi_Q^2(t)e^{t^2\sigma_r^2/\lambda^2}dt + o(\{n\lambda^2\}^{-1}), \end{aligned}$$

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

$$\text{MISE}\{\widehat{f}_x(x)\} \sim \frac{1}{4\pi n^2\lambda}\sum_{r=1}^{n}\int \Phi_Q^2(t)e^{t^2\sigma_r^2/\lambda^2}dt + \frac{\lambda^4}{4}\mu_{Q,2}^2\int\{f_x''(x)\}^2dx. \qquad (2.3.35)$$

Asymptotic analysis of the MISE ($n \to \infty, \lambda \to 0$) in equation (2.3.35) is difficult because of the dependence of the MISE on the particular sequence of variances $\sigma_1^2,,\sigma_2^2,\ldots$. 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. This is equivalent to assuming that $\sigma_1^2,\ldots,\sigma_n^2$ are independent and identically distributed with common density $f_{\sigma^2}(s)$, independent of $X_1,\ldots,X_n$. Under this assumption,

$$\begin{aligned} \frac{1}{n}\sum_{r=1}^{n}\int \Phi_Q^2(t)e^{t^2\sigma_r^2/\lambda^2}dt &\to \int\int \Phi_Q^2(t)e^{t^2v/\lambda^2}dt f_{\sigma^2}(v)dv \qquad (2.3.36) \\ &= \int \Phi_Q^2(t)M_{\sigma^2}(t^2/\lambda^2)dt, \qquad (2.3.37) \end{aligned}$$

where $M_{\sigma^2}(s)$ is the moment generating function, $E(e^{s\sigma^2})$, assumed to exist for all real s. Substituting (2.3.37) for (2.3.36) in (2.3.35) results in the approximation

$$\text{MISE}_n(\lambda) = \frac{1}{4\pi n\lambda}\int \Phi_Q^2(t)M_{\sigma^2}(t^2/\lambda^2)dt + b\lambda^4, \qquad (2.3.38)$$

where $b = \mu_{Q,2}^2 \int \{f_x''(x)\}^2 dx/4$. The expression in equation (2.3.38) is more amenable to asymptotic analysis. We now consider an interesting and informative special case.

Suppose that $\sigma_1^2, \ldots, \sigma_n^2$ are independent, identically distributed Uniform $(c,\ d)$ random variables where $0 \leq c < d$. Then $f_{\sigma^2}(s)$ has moment generating function $M_{\sigma^2}(t) = (e^{dt} - e^{ct})/\{(d-c)t\}$. Let $Q(x)$ be a kernel function such that $Q(x) \propto \{\sin(x)/x\}^{2k}$ for some $k \geq 2$. Kernels of this form are entire functions and satisfy the regularity conditions stated in Theorem 2.1.1.1. Moreover, appropriately scaled, $Q(x)$ is an even probability density function whose characteristic function vanishes outside the finite interval, $[-A,\ A]$ where $A = 2k$.

Now consider $\widehat{f}_x(x)$ in equation (2.2.3), calculated with this kernel. Setting $x = t/\lambda$, equation (2.3.38) becomes

$$\text{MISE}_n(\lambda) = \frac{1}{2n\pi(d-c)} \int_0^{A/\lambda} \Phi_Q^2(\lambda x)(e^{dx^2} - e^{cx^2})x^{-2}dx + b\lambda^4. \tag{2.3.39}$$

First, let $\{\lambda_n\}$ be a sequence of bandwidths such that $\lambda_n \to 0$ as $n \to \infty$. We seek the optimal sequence of bandwidths, i.e., the sequence for which $\text{MISE}_n(\lambda_n)$ converges to 0 at the fastest possible rate. A general form for the optimal sequence of bandwidths is suggested by the results for the known-variance deconvoluting kernel density estimator, and we examine sequences $\{\lambda_n\}$ where $\lambda_n \propto \{\log(n)\}^{-1/2}$. We show next that $\text{MISE}_n(\lambda_n)$ converges most quickly when $\lambda_n = A\{d/\log(n)\}^{1/2}$.

Let $\gamma \geq dA^2$ and consider the bandwidth $\lambda_n = \{\gamma/\log(n)\}^{1/2}$. It is clear that for any $0 < \gamma < \infty$, as n increases, the bias term in equation (2.3.39) converges to 0. We now examine the asymptotic behavior of the variance term in equation (2.3.39).

Define

$$I_n = \frac{1}{2n\pi(d-c)} \int_0^{A\{\log(n)/\gamma\}^{1/2}} \Phi_Q^2(x\{\gamma/\log(n)\}^{1/2})(e^{dx^2} - e^{cx^2})x^{-2}dx, \qquad (2.3.40)$$

and note that because $\Phi_Q(t) \leq 1$ for all t,

$$I_n \leq \frac{1}{2n\pi(d-c)} \int_0^{A\{\log(n)/\gamma\}^{1/2}} (e^{dx^2} - e^{cx^2})x^{-2}dx.$$

It follows from l'Hopital's rule that

$$\begin{aligned} \lim_{n\to\infty} I_n &\propto \lim_{n\to\infty} \frac{e^{(dA^2\log(n)/\gamma)} - e^{(cA^2\log(n)/\gamma)}}{n\{\log(n)\}^{3/2}} \\ &\leq \lim_{n\to\infty} \frac{n^{(dA^2/\gamma-1)} - n^{(cA^2/\gamma-1)}}{\{\log(n)\}^{3/2}} \\ &\leq \lim_{n\to\infty} \frac{n^{(dA^2/\gamma-1)}}{\{\log(n)\}^{3/2}}. \end{aligned}$$

When $dA^2/\gamma - 1 \leq 0$, $n^{(dA^2/\gamma-1)} \leq 1$. Hence $\text{MISE}_n(\lambda)$ converges to 0 whenever $\gamma \geq dA^2$, or equivalently, whenever $\lambda_n > A\{d/\log(n)\}^{1/2}$.

Next let $\gamma < dA^2$. We consider the behavior of $\text{MISE}_n(\lambda_n)$ for the sequence of bandwidths $\{\lambda_n\}$ where $\lambda_n = \{\gamma/\log(n)\}^{1/2}$. Let $\epsilon > 0$ and set $\gamma = d(A-\epsilon)^2$. Substitution into equation (2.3.40) yields

$$I_n = \frac{1}{2n\pi(d-c)} \int_0^{A\{\log(n)/(A-\epsilon)^2 d\}^{1/2}} \Phi_Q^2(x\{(A-\epsilon)^2 d/\log(n)\}^{1/2})(e^{dx^2} - e^{cx^2})x^{-2}dx,$$

which, letting $y = x\{\log(n)\}^{-1/2}$ and simplifying, becomes

$$I_n = \frac{1}{2\pi(d-c)} \int_0^{A/(A-\epsilon)d^{1/2}} \Phi_Q^2(y(A-\epsilon)d^{1/2})y^{-2}(n^{(dy^2-1)} - n^{(cy^2-1)})\{\log(n)\}^{-1/2}dy. \qquad (2.3.41)$$

Define

$$D = \frac{n^{(dy^2-1)} - n^{(cy^2-1)}}{y^2\{\log(n)\}^{1/2}},$$

and consider the point $y = \delta A/\{(A-\epsilon)d^{1/2}\}$ where $(A-\epsilon)/A < \delta < 1$. This point lies inside the range of integration in equation (2.3.41). Then,

$$\lim_{n\to\infty} D \propto \lim_{n\to\infty} \frac{n^{\{\delta A/(A-\epsilon)\}^2} - n^{\{c^{1/2}\delta A/d^{1/2}(A-\epsilon)\}^2}}{\{\log(n)\}^{1/2}} = \infty$$

because $\{\delta A/(A-\epsilon)\}^2 > 1$. Therefore for sequences $\{\lambda_n\}$ where $\lambda_n < A\{d/\log(n)\}^{1/2}$, I_n in equation (2.3.41) diverges, and thus $\text{MISE}_n(\lambda_n)$ diverges. The smallest bandwidth sequence for which the integrated variance converges to 0, $\lambda_n = A\{d/\log(n)\}^{1/2}$, minimizes the bias term in equation (2.3.38), and thus is the optimal sequence of bandwidths. It can be seen from equation (2.3.39) that with this bandwidth, the variance term in $\text{MISE}_n(\lambda_n)$ decreases at an exponential rate, much more rapidly than the polynomial rate of the bias term. Therefore, the rate at which $\text{MISE}_n(\lambda_n)$ converges to 0 is dictated by the bias term and is $A^4\{d/\log(n)\}^2$.

For the special case just analyzed the measurement error variances were assumed to be Uniform(c, d), $0 < c < d$. The rate of convergence of our estimator is proportional to $\{\log(n)\}^{-2}$, which is 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 d, 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 rates of convergence, and that the rate of convergence is driven by the larger error variances when the latter are heteroscedastic. The asymptotic results are interesting, but their relevance in

finite samples is limited, because of the very large samples needed for the asymptotic approximations to be valid as indicated by the $\{\log(n)\}^{-2}$ rate.

2.3.2 A Weighted Estimator for the Case of Heteroscedastic Measurement Errors

A commonly-used strategy to reduce variability in an estimator whose components have unequal variances is to weight each component by a factor that is proportional to the inverse of its variance. In this section we investigate this strategy to reduce the variability of $\widehat{f}_x(x)$ in equation (2.2.3). As $\widehat{f}_x(x)$ is the sum of independent components that have a common mean but different variances, it is reasonable to expect that weighting will reduce its overall variability.

We present a weighted estimator, $\widetilde{f}_x(x)$. Optimal weights, $w_1, \ldots, w_n$, are derived under the assumption that the measurement error variances are known, and are selected to minimize the asymptotic integrated variance of $\widehat{f}_x(x)$. We examine the asymptotic properties of the estimator calculated with these optimal weights, which we denote by $\widetilde{f}_{opt}(x)$. The results of this analysis provide intuition for the properties of the weighted estimator $\widetilde{f}_x(x)$, which uses the set of estimated weights $\widehat{w}_1, \ldots, \widehat{w}_n$ calculated with the unknown measurement error variances replaced by their sample estimates. Because we are substituting estimated weights for true weights, and these 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. Nev-

ertheless the simulation studies reported in Section 2.4 indicate that the approximate weighting has a significant effect in finite samples.

The optimal weighted estimator has the general form

$$\widetilde{f}_{opt}(x) = \frac{1}{\lambda}\sum_{r=1}^{n} w_r E\left\{ Q\left(\frac{x - \left\{ \overline{W}_r + i(\frac{m_r-1}{m_r})^{1/2}\widehat{\sigma}_r T_r \right\}}{\lambda} \right) \middle| \, \overline{W}_r, \widehat{\sigma}_r^2 \right\}, \qquad (2.3.42)$$

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 $\widetilde{f}_x(x)$ is $\int \mathrm{Var}\{\widetilde{f}_x(x)\}dx = \widetilde{V}_1 - \widetilde{V}_2$ where

$$\widetilde{V}_1 = \frac{1}{\lambda^2}\sum_{r=1}^{n} w_r^2 \int E\left\{ \left[Q^*\left(\frac{x - \overline{W}_r}{\lambda}, \lambda, m_r, \widehat{\sigma}_r \right) \right]^2 \right\} dx, \qquad (2.3.43)$$

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 equation (2.2.9).

Let $0 < B < \infty$ and suppose that $w_r \leq B/n$ for $r = 1, \ldots, n$. It follows from equation (2.3.27) that $\widetilde{V}_2 = o\{(n\lambda)^{-1}\}$. Thus asymptotically, $\widetilde{V}_1$ is the dominant term in the integrated variance, and we select weights to minimize this quantity. Let

$$h_r(\sigma_r^2, m_r, \lambda) = \int \Phi_Q^2(t) \left\{ \sum_{k=0}^{\infty} \frac{\left(\frac{m_r-2}{2}\right)_k \left(\frac{2t^2\sigma_r^2}{m_r\lambda^2}\right)^k}{(m_r-2)_k \, k!} \right\} dt \qquad (2.3.44)$$

and note from equation (2.3.34) that $h_r(\sigma_r^2, m_r, \lambda)$ represents the contribution to the asymptotic integrated variance from the rth component of $\widehat{f}_x(x)$. Define the set of weights,

$$w_r = \frac{\{h_r(\sigma_r^2, m_r, \lambda)\}^{-1}}{\sum_{r=1}^{n}\{h_r(\sigma_r^2, m_r, \lambda)\}^{-1}} \qquad (2.3.45)$$

for $r = 1, \ldots, n$. We now show that these weights minimize equation (2.3.43). Using equations (2.3.28), (2.3.33) and (2.3.44), we write V_1 as

$$\widetilde{V}_1 = \frac{1}{2\pi\lambda} \sum_{r=1}^{n} w_r^2 h_r(\sigma_r^2, m_r, \lambda), \tag{2.3.46}$$

which after substituting the weights defined in equation (2.3.45) becomes

$$\widetilde{V}_{1,w} = \frac{1}{2\pi\lambda} \left\{ \sum_{r=1}^{n} \left\{ h_r(\sigma_r^2, m_r, \lambda) \right\}^{-1} \right\}^{-1}. \tag{2.3.47}$$

Now let u_r, $r = 1, \ldots, n$ be any other set of weights such that $u_r \geq 0$ for all r and $\sum_{r=1}^{n} u_r = 1$. Substituting these weights into equation (2.3.46) gives

$$\widetilde{V}_{1,u} = \frac{1}{2\pi\lambda} \sum_{r=1}^{n} u_r^2 h_r(\sigma_r^2, m_r, \lambda).$$

Using the facts that $\{\sum_{r=1}^{n} u_r\}^2 = 1$ and

$$\left\{ \sum_{r=1}^{n} u_r \right\}^2 \leq \left\{ \sum_{r=1}^{n} u_r^2 h_r(\sigma_r^2, m_r, \lambda) \right\} \left\{ \sum_{r=1}^{n} \left\{ h_r(\sigma_r^2, m_r, \lambda) \right\}^{-1} \right\},$$

by the Cauchy-Schwartz inequality, it follows that $\widetilde{V}_{1,w} \leq \widetilde{V}_{1,u}$. Thus $\widetilde{V}_1$ in equation (2.3.43) is minimized for the set of weights defined in equation (2.3.45), and combining equations (2.3.44) and (2.3.47) is

$$\widetilde{V}_1 = \left\{ 2\pi\lambda \sum_{r=1}^{n} \left\{ \int \Phi_Q^2(t) \left\{ \sum_{k=0}^{\infty} \frac{\left(\frac{m_r-2}{2}\right)_k}{(m_r-2)_k} \frac{\left(\frac{2t^2\sigma_r^2}{m_r\lambda^2}\right)^k}{k!} \right\} dt \right\}^{-1} \right\}^{-1}. \tag{2.3.48}$$

Substituting the sample variances, $\widehat{\sigma}_1^2, \ldots, \widehat{\sigma}_n^2$ for the true variances in equation (2.3.45) forms the set of estimated weights, $\widehat{w}_1, \ldots, \widehat{w}_n$, and the estimator

$$\widetilde{f}_x(x) = \frac{1}{\lambda} \sum_{r=1}^{n} \widehat{w}_r E \left\{ Q \left(\frac{x - \left\{ \overline{W}_r + i(\frac{m_r-1}{m_r})^{1/2} \widehat{\sigma}_r T_r \right\}}{\lambda} \right) \middle| \, \overline{W}_r, \widehat{\sigma}_r^2 \right\}. \tag{2.3.49}$$

Mean Integrated Squared Error of $\widetilde{f}_{opt}(x)$

We now consider the asymptotic properties of the weighted estimator, $\widetilde{f}_{opt}(x)$, calculated with the optimal weights in equation (2.3.45). We derive an expression for the MISE of $\widetilde{f}_{opt}(x)$ 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 $\widetilde{f}_x(x)$ in equation (2.3.49). The estimated variances in $\widetilde{f}_x(x)$ complicate a direct examination of its asymptotic properties. Instead, we evaluate this estimator via simulation in Section 2.4.

The estimator $\widetilde{f}_{opt}(x)$ has the same key property in equation (2.2.4) as the heteroscedastic-variance estimator, and so has the same integrated squared bias as the kernel density estimator of $f_x(x)$ based on the true data. Combining equations (2.3.24) and (2.3.48), we have that as $n \to \infty$ and $\lambda \to 0$,

$$\begin{aligned} \text{MISE}\{\widetilde{f}_{opt}(x)\} &\sim \left\{ 2\pi\lambda \sum_{r=1}^{n} \left\{ \int \Phi_Q^2(t) \left\{ \sum_{k=0}^{\infty} \frac{\left(\frac{m_r-2}{2}\right)_k}{(m_r-2)_k} \frac{\left(\frac{2t^2\sigma_r^2}{m_r\lambda^2}\right)^k}{k!} \right\} dt \right\}^{-1} \right\}^{-1} \\ &+ \frac{\lambda^4}{4}\mu_{Q,2}^2 \int \{f_x''(x)\}^2 dx. \end{aligned}$$

The asymptotic properties of this expression are difficult to examine in full generality. The expression simplifies in the case of two replicate measurements and we continue our analysis under this assumption.

From Lemma 2.3.1.1 and equation (2.3.48), when $m_r = 2$ for all r, as $n \to \infty$ and $\lambda \to 0$,

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

and the MISE becomes

$$\text{MISE}\{\widetilde{f}_{opt}(x)\} \sim \left\{ 4\pi\lambda \sum_{r=1}^{n} \left\{ \int \Phi_Q^2(t) e^{t^2\sigma_r^2/\lambda^2} dt \right\}^{-1} \right\}^{-1} + \frac{\lambda^4}{4} \mu_{Q,2}^2 \int \{f_x''(x)\}^2 dx. \tag{2.3.50}$$

The asymptotic behavior of the MISE depends on the sequence of variances $\sigma_1^2, \ldots, \sigma_n^2$. As in Section 2.3.1, we make the assumption that these variances are independent and identically distributed with density function $f_{\sigma^2}(s)$, and are independent of $X_1, \ldots, X_n$. From the relationship

$$\sum_{r=1}^{n} \left\{ \int \Phi_Q^2(t) e^{t^2\sigma_r^2/\lambda^2} dt \right\}^{-1} \sim n \int \left\{ \int \Phi_Q^2(t) e^{t^2 v/\lambda^2} dt \right\}^{-1} f_{\sigma^2}(s) ds,$$

it follows that the MISE in equation (2.3.50) can be approximated by

$$\text{MISE}_n(\lambda) = \left\{ 4\pi n\lambda \int \left\{ \int \Phi_Q^2(t) e^{t^2 v/\lambda^2} dt \right\}^{-1} f_{\sigma^2}(s) ds \right\}^{-1} + b\lambda^4, \tag{2.3.51}$$

where $b = \mu_{Q,2}^2 \int \{f_x''(x)\}^2 dx/4$. In Section 2.3.1, we showed that when $f_{\sigma^2}(s)$ is the Uniform$(c,\ d)$ density and $Q(x) \propto \{\sin(x)/x\}^{2k}$ for some $k \geq 2$, the MISE of the estimator for the case of heteroscedastic measurement errors converges to 0 at a rate proportional to $\{\log(n)\}^{-2}$ We now reconsider this case and show that the MISE in equation (2.3.50) converges to 0 at the same rate.

Let $\sigma_1^2, \ldots, \sigma_n^2$ be independent, identically distributed Uniform $(c,\ d)$ random variables where $0 \leq c < d$, and let $Q(x) \propto \{\sin(x)/x\}^{2k}$ for some $k = 2, 3, \ldots$. Note that the characteristic function $\Phi_Q(t)$ vanishes outside the interval $[-A,\ A]$ where $A = 2k$. Under these assumptions, equation (2.3.51) becomes

$$\text{MISE}_n(\lambda) = \left\{ 2\pi n\lambda(d-c) \int_c^d \left\{ \int_0^A \Phi_Q^2(t) e^{t^2 s/\lambda^2} dt \right\}^{-1} ds \right\}^{-1} + b\lambda^4. \tag{2.3.52}$$

We again consider the sequence of bandwidths $\{\lambda_n\}$ where $\lambda_n \propto \{\log(n)\}^{-1/2}$, and show that the optimal sequence is given by $\lambda_n = A\{d/\log(n)\}^{1/2}$. It is clear that for bandwidths of this form, the bias term in equation (2.3.52) converges to 0. We now investigate the asymptotic behavior of the variance term in equation (2.3.52), which we write as

$$\widetilde{V}_n(\lambda) = \left\{ 2\pi n\lambda(d-c) \int_c^d \left\{ \int_0^A \Phi_Q^2(t) e^{t^2 s/\lambda^2} dt \right\}^{-1} ds \right\}^{-1}. \tag{2.3.53}$$

Consider the innermost integral in this expression. Define

$$I_n = \frac{1}{n} \int_0^A \Phi_Q^2(t) e^{t^2 s/\lambda^2} dt, \tag{2.3.54}$$

which letting $y = ts^{1/2}/\lambda$ becomes

$$I_n = \frac{1}{n} \int_0^{As^{1/2}/\lambda} \Phi_Q^2(y\lambda s^{-1/2}) e^{y^2} dy.$$

It is seen from equation (2.3.53) that if $I_n \to 0$, then $\widetilde{V}_n(\lambda) \to 0$. Consider the bandwidth $\lambda_n = \{\gamma/\log(n)\}^{1/2}$ where $\gamma \geq dA^2$. Using the fact that $\Phi_Q^2(t) \leq 1$ for all t and applying l'Hopital's rule and simplifying gives

$$\begin{aligned} \lim_{n\to\infty} I_n &\leq \lim_{n\to\infty} \frac{1}{n} \int_0^{A\{s\log(n)/\gamma\}^{1/2}} e^{y^2} dy \\ &= \lim_{n\to\infty} \frac{As^{1/2} e^{A^2 s\log(n)/\gamma}}{2n\{\log(n)\}^{1/2}} \\ &\propto \lim_{n\to\infty} \frac{s^{1/2} n^{(A^2 s/\gamma - 1)}}{\{\log(n)\}^{1/2}}. \end{aligned}$$

Because $\gamma \geq A^2 d$, it follows that $A^2 s/\gamma - 1 \leq s/d - 1 \leq 0$ for all s in the interval $(c,\ d)$, and so $\lim_{n\to\infty} I_n = 0$. Therefore, $\widetilde{V}_n(\lambda_n)$ converges to 0 for all sequences of bandwidths $\lambda_n \geq A\{d/\log(n)\}^{1/2}$.

We now consider $\widetilde{V}_n(\lambda_n)$ for bandwidths $\lambda_n = \{\gamma/\log(n)\}^{1/2}$ where $\gamma \leq dA^2$. For $\epsilon > 0$, let $\gamma = d(A-\epsilon)^2$. After a change of variables with $x = \{t^2 s/\lambda \log(n)\}^{1/2}$ equation (2.3.54) becomes

$$I_n = \frac{s^{1/2}}{n\lambda\{\log(n)\}^{1/2}} \int_0^{\{A^2 s/\lambda^2 \log(n)\}^{1/2}} \Phi_Q^2\{x\lambda\{\log(n)/s\}^{-1/2}\} e^{x^2 \log(n)} dy,$$

which upon substituting $\lambda_n = \{d(A-\epsilon)^2/\log(n)\}^{1/2}$ and simplifying, is

$$I_n = s^{1/2} \int_0^{\{A^2 s/((A-\epsilon)^2 d)\}^{1/2}} \Phi_Q^2\{x(A-\epsilon)(d/s)^{1/2}\}\{\log(n)\}^{1/2} n^{(x^2-1)} dx. \qquad (2.3.55)$$

Whenever $x > 1$, $\{\log(n)\}^{1/2} n^{(x^2-1)} \to \infty$. But for each s in the interval $(d(A-\epsilon)^2/A^2,\ d)$, the upper limit of integration in equation (2.3.55) exceeds one, and so $I_n \to \infty$.

Thus, $\widetilde{V}_n$ diverges for bandwidths $\lambda_n < A\{d/\log(n)\}^{1/2}$. We conclude, then, that the MISE of the estimator $\widetilde{f}_{opt}(x)$ converges to 0 most quickly with the sequence of bandwidths given by $\lambda_n = A\{d/\log(n)\}^{1/2}$, and converges at the rate of $A^4\{d/\log(n)\}^2$. Comparing these results with the results of Section 2.3.1 reveals 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, it is reasonable to expect that in finite samples, weighting will reduce the variance of the estimator, and we investigate this via simulation in Section 2.4.

2.3.3 Homoscedastic Measurement Error: Pooled Estimation of σ^2

We now consider the case of homoscedastic measurement error, $\sigma_1^2 = \ldots, = \sigma_n^2 = \sigma^2$, and the common variance σ^2 is estimated by the pooled estimator of variance defined in equation (2.2.17), based on $d = \sum_{r=1}^{n}(m_r - 1)$ degrees of freedom. Because we are assuming that $m_r \geq 2$ for all r, it follows that $d \geq n$.

Recall that the estimator in equation (2.2.18) for the case of homoscedastic measurement errors has the key property in equation (2.2.4), that conditional on the true data it is an unbiased estimator of the true-data kernel density estimator. Consequently, $\widehat{f}_x(x)$ has the same bias and same integrated squared bias as the true-data kernel density estimator.

Because the summands in equation (2.2.18) share the common variance estimator $\widehat{\sigma}^2$, they are not independent. This complicates direct calculation of $\mathrm{Var}\{\widehat{f}_x(x)\}$ and the integrated variance. However, we now argue that under certain smoothness and regularity conditions the homoscedastic-variance estimator has the same integrated variance as the estimator with $\widehat{\sigma}^2$ replaced by the true variance σ^2. This result is intuitively plausible in light of the fact that the known-variance estimator of $f_x(x)$ converges at the rate of $\{\log(n)\}^{-2}$ and the pooled estimator of variance converges at the much faster rate of $n^{-1/2}$. This is the essence of the argument that we now sketch.

Because the conditional expectation in $\widehat{f}_x(x)$ is real-valued, we need only consider

the real part of the function $Q()$. For notational convenience we denote $Re[Q()]$ by Q_{Re}. We have,

$$\begin{aligned}
\widehat{f}_x(x) &= \frac{1}{n\lambda}\sum_{r=1}^{n} E\left\{ Q\left(\frac{x - \left\{ \overline{W}_r + i(\frac{d}{m_r})^{1/2}\widehat{\sigma} T_{r,d} \right\}}{\lambda} \right) \,\middle|\, \overline{W}_r, \widehat{\sigma}^2 \right\} \\
&= \frac{1}{n\lambda}\sum_{r=1}^{n} E\left\{ Q_{Re}\left(\frac{x - \left\{ \overline{W}_r + i(\frac{d}{m_r})^{1/2}\widehat{\sigma} T_{r,d} \right\}}{\lambda} \right) \,\middle|\, \overline{W}_r, \widehat{\sigma}^2 \right\} \\
&= E\left\{ \frac{1}{n\lambda}\sum_{r=1}^{n} Q_{Re}\left(\frac{x - \left\{ \overline{W}_r + i(\frac{d}{m_r})^{1/2}\widehat{\sigma} T_{r,d} \right\}}{\lambda} \right) \,\middle|\, \overline{W}_1, \ldots, \overline{W}_n, \widehat{\sigma}^2 \right\}
\end{aligned} \tag{2.3.56}$$

where $T_{r,d} = Z_{r,1}/\sqrt{Z_{r,1}^2 + \cdots + Z_{r,d}^2}$.

Our argument is based on a Taylor Series expansion. Recall that $d \geq n$, and thus as $n \to \infty$,

$$\sqrt{d}\; T_{r,d} = \frac{Z_{r,1}}{\sqrt{\frac{1}{d}(Z_{r,1}^2 + \cdots + Z_{r,d}^2)}} \to Z_{r,1}$$

and

$$\widehat{\sigma}^2 \to \sigma^2$$

almost surely by the Strong Law of Large Numbers. The fact that for large n, $\widehat{\sigma} d^{1/2} T_{r,d}$ is close to $\sigma Z_{1,r}$ suggests the suitability of the one-term Taylor Series approximation

$$\begin{aligned}
\frac{1}{n\lambda}\sum_{r=1}^{n} Q_{Re}\left(\frac{x - \left\{ \overline{W}_r + i(\frac{d}{m_r})^{1/2}\widehat{\sigma} T_r \right\}}{\lambda} \right) &\approx \frac{1}{n\lambda}\sum_{r=1}^{n} Q_{Re}\left(\frac{x - \left\{ \overline{W}_r + i m_r^{-1/2}\sigma Z_{r,1} \right\}}{\lambda} \right) \\
&+ \frac{1}{n\lambda}\sum_{r=1}^{n} (\widehat{\sigma} d^{1/2} T_{r,d} - \sigma Z_{r,1}) Q'_{Re}\left(\frac{x - \left\{ \overline{W}_r + i m_r^{-1/2}\alpha_r \right\}}{\lambda} \right),
\end{aligned} \tag{2.3.57}$$

where α_r is an interior point on the line segment that joins $d^{1/2}\widehat{\sigma} T_r$ and $\sigma Z_{r,1}$.

At best, the MISE of $\widehat{f}_x(x)$ converges to 0 at the rate of $\{\log(n)\}^{-2}$ (Carroll & Hall 1988), and achieves this rate with the bandwidth $\lambda = \sigma\{\log(n)\}^{-1/2}$ (Stefanski 1990). We now show that relative to the first term, the contribution to the MISE from the second term on the right-hand side of equation (2.3.57) is minor. We use an application of the Cauchy-Schwartz inequality to show that the second term is $O_p(\{n\lambda^2\}^{-1})$, and so is asymptotically negligible for bandwidths on the order of $\{\log(n)\}^{-1/2}$.

It follows from the Cauchy-Schwartz inequality that

$$\left| \frac{1}{n\lambda} \sum_{r=1}^{n} (\widehat{\sigma} d^{1/2} T_{r,d} - \sigma Z_{r,1}) Q' \left(\frac{x - \left\{ \overline{W}_r + i m_r^{-1/2} \sigma Z_{r,1} \right\}}{\lambda} \right) \right|^2 \leq A_1 A_2$$

where

$$A_1 = \frac{1}{n\lambda} \sum_{r=1}^{n} (\widehat{\sigma} d^{1/2} T_r - \sigma Z_{r,1})^2 \tag{2.3.58}$$

and

$$A_2 = \frac{1}{n\lambda} \sum_{r=1}^{n} \left\{ Q' \left(\frac{x - \left\{ \overline{W}_r + i m_r^{-1/2} \alpha \right\}}{\lambda} \right) \right\}^2 . \tag{2.3.59}$$

Consider A_1 in equation (2.3.58). Because $T_{r,d} = Z_{r,1} / \sqrt{Z_{r,1}^2 + \cdots + Z_{r,d}^2}$ and $d/n > 1$,

$$\begin{aligned} A_1 &= \frac{1}{n\lambda} \left(\frac{\widehat{\sigma}}{\sqrt{(Z_{r,1}^2 + \ldots + Z_{r,d}^2)/d}} - \sigma \right)^2 \sum_{r=1}^{n} Z_{r,1}^2 \\ &\leq \frac{1}{n\lambda} \left\{ d^{1/2} \left(\frac{\widehat{\sigma}}{\sqrt{(Z_{r,1}^2 + \ldots + Z_{r,d}^2)/d}} - \sigma \right)^2 \right\} \frac{1}{n} \sum_{r=1}^{n} Z_{r,1}^2 . \end{aligned}$$

From the facts that $\sqrt{(Z_{r,1}^2 + \ldots + Z_{r,d}^2)/d} = o_p(1)$, $d^{1/2}(\widehat{\sigma} - \sigma) = O_p(1)$ and $\frac{1}{n} \sum_{r=1}^{n} Z_{r,1}^2 = O_p(1)$, it follows that $A_1 = O_p(\{n\lambda\}^{-1})$.

Now consider A_2 in equation (2.3.59). Provided that $Q'_{Re}(\cdot)$ is bounded, $A_2 = O_p(\lambda^{-1})$. Therefore, $A_1 A_2 = O(\{n\lambda^2\}^{-1})$, so that $A_1 A_2 \xrightarrow{P} 0$ for any λ such that $n\lambda^2 \to \infty$.

Thus, from equation (2.3.56),

$$\widehat{f}_x(x) \approx \frac{1}{n\lambda} \sum_{r=1}^{n} E\left\{ Q_{Re}\left(\frac{x - \left\{ \overline{W}_r + i m_r^{-1/2} \sigma Z_{r,1} \right\}}{\lambda} \right) \middle| \overline{W}_r \right\}, \tag{2.3.60}$$

and we use this expression to approximate the MISE of $\widehat{f}_x(x)$. First define the function

$$Q_r^{\ddagger}(z, \lambda, m_r, \sigma) = E\left\{ Q_{Re}\left(z - \frac{i\sigma Z_{r,1}}{m_r^{1/2}\lambda} \right) \middle| \overline{W}_r, \right\}, \tag{2.3.61}$$

so that $\widehat{f}_x(x)$ in equation (2.3.60) can be written as

$$\widehat{f}_x(x) \approx \frac{1}{n\lambda} \sum_{r=1}^{n} Q_r^{\ddagger}\left(\frac{x - \overline{W}_r}{\lambda}, \lambda, m_r, \sigma \right).$$

Applying the Fourier inversion formula and interchanging the operations of integration and expectation in equation (2.3.61) yields

$$Q_r^{\ddagger}(z, \lambda, m_r, \sigma) = \frac{1}{2\pi} \int e^{-itz} E\left\{ \exp\left(\frac{-t\sigma Z_{r,1}}{\lambda m_r^{1/2}} \right) \middle| \overline{W}_r, \right\} \Phi_Q(t) dt. \tag{2.3.62}$$

Because $Z_{r,1}$ is a standard normal random variable,

$$E\{e^{-t\sigma Z_{r,1}/m_r^{1/2}\lambda}\} = e^{t^2\sigma^2/2m_r\lambda^2},$$

so that

$$Q_r^{\ddagger}(z, \lambda, m_r, \sigma) = \frac{1}{2\pi} \int e^{-itz} e^{t^2\sigma^2/2m_r\lambda^2} \Phi_Q(t) dt. \tag{2.3.63}$$

Comparison with equation (2.2.22) shows that $Q_r^{\ddagger}(z, \lambda, m_r, \sigma)$ is the same function as the Stefanski-Carroll deconvoluting kernel defined for the case of $\mathrm{N}(0,\ \sigma^2/m_r)$

measurement errors. Therefore it has the same expectation and same bias as the true-data kernel density estimator. The approximate integrated variance of $\widehat{f}_x(x)$ is $\int \mathrm{Var}\{\widehat{f}_x(x)\}dx \approx V_1 - V_2$ where

$$V_1 = \frac{1}{n^2\lambda^2}\sum_{r=1}^{n}\int E\left\{\left[Q^{\ddagger}\left(\frac{x-\overline{W}_r}{\lambda},\lambda,m_r,\sigma\right)\right]^2\right\}dx, \tag{2.3.64}$$

and

$$V_2 = \frac{1}{n^2\lambda^2}\sum_{r=1}^{n}\int \left\{E\left[Q^{\ddagger}\left(\frac{x-\overline{W}_r}{\lambda},\lambda,m_r,\sigma_r\right)\right]\right\}^2 dx. \tag{2.3.65}$$

First consider V_2. Because

$$E\left\{Q^{\ddagger}\left(\frac{x-\overline{W}_r}{\lambda},\lambda,m_r,\widehat{\sigma}_r\right)\right\} = E\left\{Q\left(\frac{x-X_r}{\lambda}\right)\right\} = \lambda\int Q(z)f_x(x-\lambda z)dz,$$

it follows from substitution into equation (2.3.65) and Parseval's Identity that

$$\begin{aligned} V_2 &= \frac{1}{n}\int\left\{\int Q(z)f_x(x-\lambda z)dz\right\}^2 dx \\ &= \frac{1}{2\pi n}\int |\Phi_f(t)|^2\Phi_Q^2(\lambda t)dt. \end{aligned} \tag{2.3.66}$$

Now consider V_1. From equation (2.3.63) and Parseval's Identity,

$$\int\left\{Q^{\ddagger}\left(\frac{x-\overline{W}_r}{\lambda},\lambda,m_r,\sigma\right)\right\}^2 = \frac{1}{2\pi}\int e^{t^2\sigma^2/m_r\lambda^2}\Phi_Q(t)dt,$$

which upon substitution into equation (2.3.64) gives

$$V_1 = \frac{1}{2\pi n^2\lambda}\sum_{r=1}^{n}\int \Phi_Q(t)^2 e^{t^2\sigma^2/m_r\lambda^2}dt. \tag{2.3.67}$$

Combining equations (2.3.24), (2.3.66), and (2.3.67) yields an approximation to the IMSE for the case of homoscedastic measurement errors, and as $n \to \infty$ and $\lambda \to 0$,

$$\mathrm{MISE}\{\widehat{f}_x(x)\} \sim \frac{1}{2\pi n^2\lambda}\sum_{r=1}^{n}\int \Phi_Q(t)^2 e^{t^2\sigma^2/m_r\lambda^2}dt + \frac{\lambda^4}{4}\mu_{Q,2}^2\int\{f_x''(x)\}^2dx. \tag{2.3.68}$$

Finally, we note that for equal numbers of replicate measurements, i.e., $m_r = m$ for all r, this approximation is equivalent to the MISE of the Stefanski-Carroll estimator. For each $r = 1, \ldots, n$, $\overline{W}_r$ measures X_r with a N(0, σ^2/m) measurement error, and so measurement errors are identically distributed with variance that is assumed known. The approximation in equation (2.3.68) simplifies to

$$\text{MISE}\{\widehat{f}_x(x)\} \sim \frac{1}{2\pi n\lambda}\int \Phi_Q(t)^2 e^{t^2\sigma^2/m\lambda^2}dt + \frac{\lambda^4}{4}\mu_{Q,2}^2\int\{f_x''(x)\}^2dx,$$

which is the MISE of the Stefanski-Carroll estimator for the case of N(0, σ^2/m) measurement errors.

2.4 Simulation Study

We performed a simulation study to investigate the performance of the proposed measurement error-corrected estimators. We included in the study the so-called naive estimator computed from the observed data and ignoring the presence of measurement error. Specifically, we took the naive estimator to be the kernel density estimator calculated with the sample means of the replicate measurements,

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

where $\phi(x)$ is the standard normal density. The measurement error-corrected estimators in equations (2.2.3), (2.3.42), and (2.2.18) were calculated with the kernel $Q(x) \propto \{\sin(x)/x\}^4$, where $Q(x)$ was scaled to have mean 0 and variance 1.

Estimator performance was evaluated with the integrated squared error (ISE), defined in general for an estimator $\widehat{f}(x)$ as

$$\text{ISE}\{\widehat{f}(x)\} = \int \{\widehat{f}(x) - f_x(x)\}^2 dx. \tag{2.4.69}$$

For each simulated data set, estimators were calculated at their optimal bandwidths, found by minimizing equation (2.4.69) in λ. 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 Section 2.5. For additional comparison, however, we also calculated the naive estimator using the popular plug-in bandwidth selection method,

$$\widehat{\tau} = 1.06\ \widehat{\sigma}^2_{\overline{W}}\ n^{-1/5}, \tag{2.4.70}$$

where $\widehat{\sigma}^2_{\overline{W}}$ is the sample variance of the means $\overline{W}_1, \ldots, \overline{W}_n$ (Silverman 1986).

We considered three factors in this study. First, we investigated the effect of sample size on the relative performance of the estimators. The logarithmic rates of convergence of our estimators suggest that very large sample sizes will be required before realizing a benefit from a correction for normal measurement error. To determine when the extra effort becomes worthwhile, we considered six different sample sizes, $n = 100, 500, 1000, 2000$ and 2500. Second, we examined the effect of the true-data density. The shape of the true-data density will influence the success with which it can be separated from the measurement error density. True data, $X_1, \ldots, X_n$, were generated from two densities, 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 $\sigma_1^2 = \cdots = \sigma_n^2 = 1$ for the case of constant variances, and $\sigma_1^2, \ldots, \sigma_n^2$ chosen uniformly over the interval $(0,\ 2)$ for the case of nonconstant variances. All estimators were computed on each set of observed data, and we used the results to gain an understanding of their robustness to assumptions on the measurement error variances.

In practice, generally only a small number of replicate measurements is observed, with two replicates being the most common. In these simulations, we considered only the case of $m_r = 2$ replicate measurements for each $r = 1, \ldots, n$. For notational convenience, in the following discussion of our results, we refer to the estimator for heteroscedastic errors in equation (2.2.3) as $\widehat{f}_{het}(x)$, the weighted estimator in equation (2.3.42) as $\widehat{f}_{wt}(x)$, and the estimator for homoscedastic errors in equation (2.2.18) as $\widehat{f}_{hom}(x)$. All results are based on fifty simulated data sets.

Average integrated squared errors are plotted by sample size in Figure 2.1 for $X_r \sim \mathrm{N}(0,\ 1)$ and Figure 2.2 for $X_r \sim$ Chi-squared(4). In each case, the average integrated squared error was most variable for the sample size of 100. Although $\widehat{f}_{hom}(x)$ consistently had a smaller integrated squared error than $\widehat{f}_{naive}(x)$ at this sample size, paired t-tests of the differences were not significant at the 0.05 level. For the simulations that considered heteroscedastic measurement errors, $\widehat{f}_{het}(x)$ had a significantly higher average integrated squared error than all other estimators for sample sizes of $n \geq 500$. It was also much more variable, its standard error pooled across these

sample sizes was 0.0007 for $X_r \sim \text{N}(0,\ 1)$ and 0.0022 for $X_r \sim$ Chi-squared(4). By contrast, the standard error pooled across both the remaining estimators and sample sizes of $n \geq 500$ was 0.0003 for $X_r \sim \text{N}(0,\ 1)$ and 0.0008 for $X_r \sim$ Chi-squared(4). In both sets of simulations, the estimators $\widehat{f}_{hom}(x)$ and $\widehat{f}_{wt}(x)$ had significantly lower average integrated squared errors than all other estimators for $n \geq 500$. However differences between the two estimators were generally not significant. The variability of the average integrated squared error was more consistent among estimators when measurement errors were homoscedastic. Pooled across estimators and sample sizes of $n \geq 500$, the standard error was 0.0004 for $X_r \sim \text{N}(0,\ 1)$ and 0.0012 for $X_r \sim$ Chi-squared(4). In both cases, when $n \geq 500$, the average integrated squared error of $\widehat{f}_{hom}(x)$ was significantly less than that of all other estimators.

It is clear from these simulations that weighting is effective in reducing the integrated squared error of $\widehat{f}_{het}(x)$. In almost every case, the average integrated squared error of $\widehat{f}_{wt}(x)$ was significantly less than that of $\widehat{f}_{het}(x)$. However, the weighted estimator performed relatively poorly when measurement error variances were constant. This is not surprising, considering that $\widehat{f}_{wt}(x)$ relies on estimates of this constant variance from only two replicate measurements. The estimator for homoscedastic errors performed well for both types of measurement errors, suggesting it as a good choice when there is doubt about the homogeneity of the error variances. Finally, a particularly encouraging conclusion that can be drawn from these simulations is that even for reasonable sample sizes, the measurement error-corrected estimators can outperform the naive estimator. This suggests that given a reliable rule for selecting a

bandwidth, correcting for measurement errors will be worthwhile in many situations.

2.5 A Bootstrap Method for Bandwidth Selection

For practical implementation of the measurement error-corrected estimators, a bandwidth selection rule is required. In this section we describe a bootstrap procedure for selecting a bandwidth. We illustrate its application with a real-data example in Section 2.6. The bootstrap has been used for bandwidth selection in traditional density estimation (see Marron, 1992; Faraway and Jhun, 1990; Taylor, 1989). We briefly outline this method, then describe its extension to the measurement error problem.

A kernel density estimator based on the random sample $X_1, \ldots, X_n$, has the form

$$\widehat{f}(x;h) = \frac{1}{nh} \sum_{r=1}^{n} K\left(\frac{x - X_r}{h}\right),$$

where $K(x)$ is a standardized probability density function and h is the bandwidth. Let $\widehat{h}_0$ be a bandwidth estimator calculated from the random sample. The initial estimator of the density $f_x(x)$ is $\widehat{f}(x;\widehat{h}_0)$, and plays the role of the true density in the bootstrap world. A bootstrap sample from this density, $X_1^*, \ldots, X_n^*$, is generated as follows. For $r = 1, \ldots, n$, let X_{π_r} be randomly drawn with replacement from $X_1, \ldots, X_n$, and let Y_r be a random variable with density $K(x)$. The random variable X_r^*, where

$$X_r^* = X_{\pi_r} + \widehat{h}_0 Y_r,$$

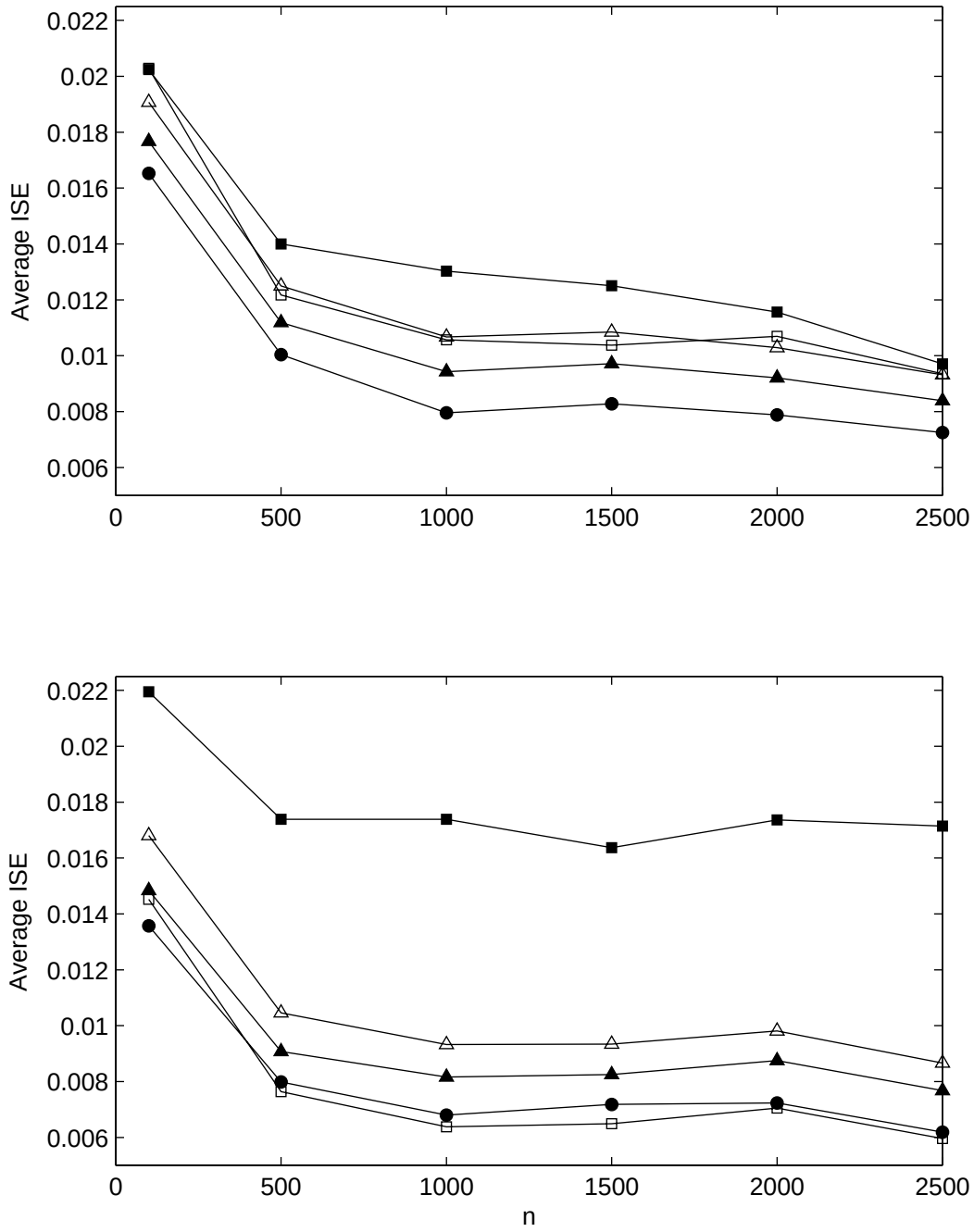

Figure 2.1: Average ISE by sample size for $X \sim \mathrm{N}(0,\ 1)$. Top panel: homoscedastic measurement errors, $\mathrm{N}(0,\ 1)$. Bottom panel: heteroscedastic measurement errors, $\mathrm{N}(0,\ \sigma_r^2)$ with σ_r^2 uniform on $(0,\ 2)$. Open triangle: $\widehat{f}_{nv}$ with plug-in bandwidth; Closed triangle: $\widehat{f}_{nv}$ with optimal bandwidth; Closed square: $\widehat{f}_{het}$; Open square: $\widehat{f}_{wt}$; Closed circle: $\widehat{f}_{hom}$.

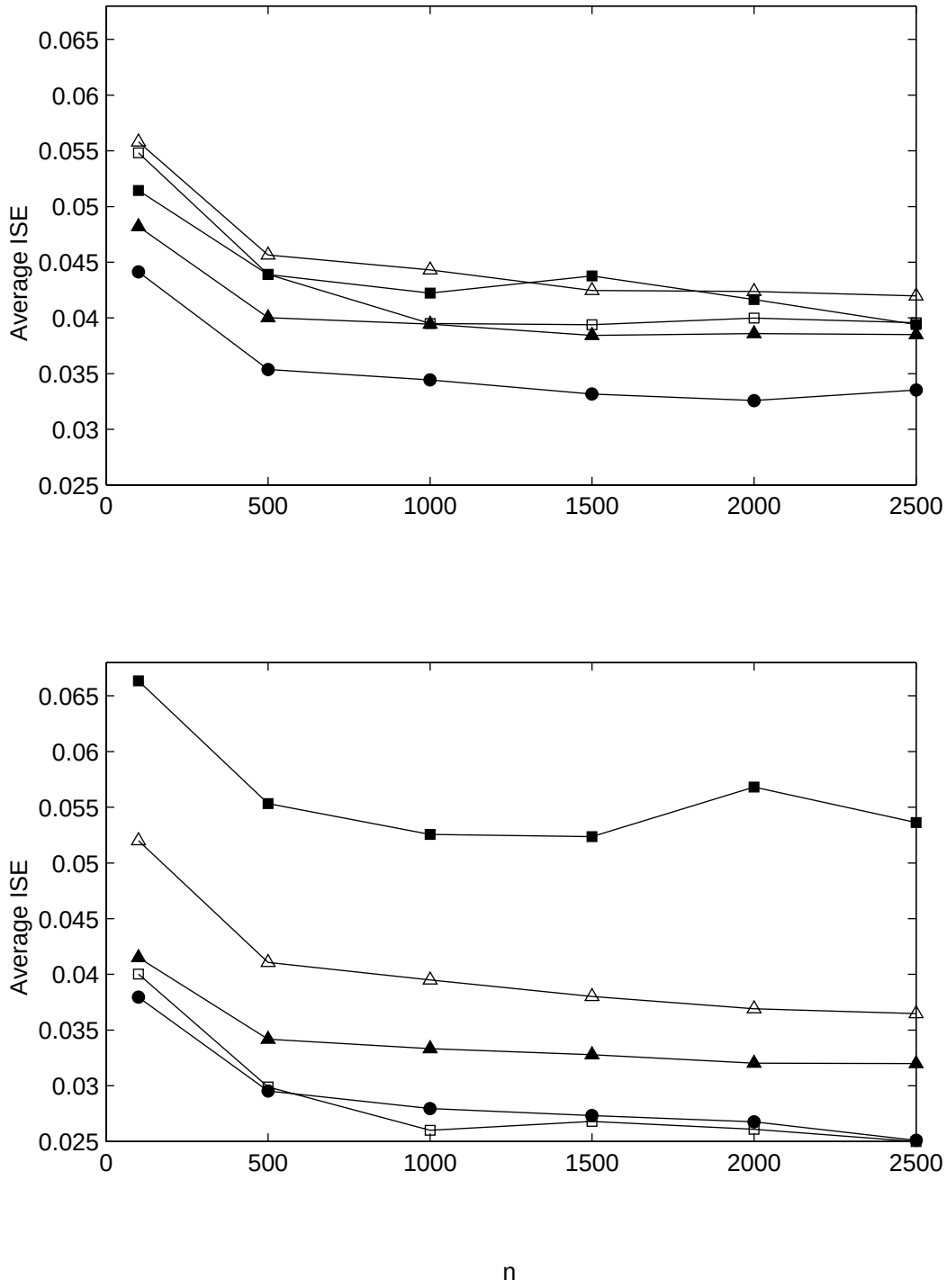

Figure 2.2: Average ISE by sample size for $X \sim \chi^2_4$. Top panel: homoscedastic measurement errors, N(0, 1). Bottom panel: heteroscedastic measurement errors, N(0, σ_r^2) with σ_r^2 uniform on (0, 2). Open triangle: $\widehat{f}_{nv}$ with plug-in bandwidth; Closed triangle: $\widehat{f}_{nv}$ with optimal bandwidth; Closed square: $\widehat{f}_{het}$; Open square: $\widehat{f}_{wt}$; Closed circle: $\widehat{f}_{hom}$.

has density function $\widehat{f}(x;\widehat{h}_0)$. The bootstrap sample, $X_1^*,\ldots,X_n^*$ is used to calculate

$$\widehat{f}^*(x;h)=\frac{1}{nh}\sum_{r=1}^{n}K\left(\frac{x-X_r^*}{h}\right).$$

Bootstrap samples are used to calculate the bootstrap mean integrated squared error

$$\text{MISE}^*(h)=E_{BS}\left[\int\left\{\widehat{f}^*(x;h)-\widehat{f}(x;\widehat{h}_0)\right\}^2dx\right] \tag{2.5.71}$$

where E_BS denotes expectation with respect to the bootstrap distribution.

The final bootstrap estimator of the bandwidth is found by minimizing $\text{MISE}^*(h)$. The minimization problem can be solved empirically by computing a large number of bootstrap density estimates, $\widehat{f}_1^*(x;h),\ldots,\widehat{f}_n^*(x;h)$, for a dense grid of bandwidths h, averaging to approximate the bootstrap expectation, and then choosing that h which achieves the smallest average integrated squared error. An appealing feature of this method, however, is that h can be estimated directly, without any resampling. We present the details of this feature below in the context of the measurement error problem.

We now describe how this method can be extended to estimate bandwidths for our measurement error-corrected estimators. The situation is different in that observed data, because they contain measurement error, are not distributed with the density $f_x(x)$. Thus, given an initial estimate of $f_x(x)$, bootstrap data sets must be generated with measurement errors that are similar to those present in the observed data.

Given the original data $\{W_{r,j}\}_{r=1,j=1}^{n,\ m_r}$, the naive estimator,

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

where $\widehat{\tau}$ is the plug-in bandwidth in equation (2.4.70), provides a reasonable, though overly smoothed, estimate of $f_x(x)$. We use this as our initial estimate of $f_x(x)$, and it plays the role of the true density in the bootstrap world.

A bootstrap sample of true data is generated first. For each $r = 1, \ldots, n$, let $\overline{W}_{\pi_r}$ be randomly drawn with replacement from the set of sample means, and let Z_r be a N(0, 1) random variable. The variable X_r^* where

$$X_r^* = \overline{W}_{\pi_r} + \widehat{\tau} Z_r$$

has density function $\widehat{f}_{naive}(x; \widehat{\tau})$. The random variables $X_1^*, \ldots, X_n^*$ form the bootstrap sample of true values.

The bootstrap true values are used as follows to construct a bootstrap sample of observed values. Consider first the case where measurement errors are heteroscedastic and $\widehat{\sigma}_1^2, \ldots, \widehat{\sigma}_n^2$ are the estimates of the measurement error variances from the original data. For $r = 1, \ldots, n$ and $j = 1, \ldots, m_r$, let $Z_{r,j}$ be a N(0, 1) random variable, and define

$$W_{r,j}^* = X_r^* + \widehat{\sigma}_r Z_{r,j}. \tag{2.5.72}$$

The random variables $\{W_{r,j}^*\}_{r=1,j=1}^{n,\ m_r}$ form the bootstrap observed-data sample, and are used to calculate the bootstrap sample estimates $\overline{W}_1^*, \ldots, \overline{W}_n^*$ and $\widehat{\sigma}_1^{2*}, \ldots, \widehat{\sigma}_n^{2*}$. It follows that for each $r = 1, \ldots, n$, conditional on X_r^* and the original data, $\overline{W}_r^* \sim$ N$(X_r^*, \widehat{\sigma}_r^2/m_r)$ and $(m_r - 1)\widehat{\sigma}_r^{2*}/\widehat{\sigma}_r^2 \sim$ Chi-squared$(m_r - 1)$.

When measurement errors are homoscedastic, a bootstrap observed-data sample is generated by replacing $\widehat{\sigma}_r$ in equation (2.5.72) with $\widehat{\sigma}$, the pooled estimate

of measurement error variance from the original data, based on d degrees of freedom. The bootstrap observed data, $\{W^*_{r,j}\}^{n,\ m_r}_{r=1,j=1}$, are used to calculate the sample estimates $\overline{W}^*_1, \ldots, \overline{W}^*_n$ and $\widehat{\sigma}^{2*}$. Conditional on $X^*_1, \ldots, X^*_n$ and the original data, $\overline{W}^*_r \sim \mathrm{N}(X^*_r, \widehat{\sigma}^2_r/m_r)$ and $d\widehat{\sigma}^{2*}/\widehat{\sigma}^2 \sim$ Chi-squared(d).

From the key property in equation (2.2.4), it follows that the measurement error-corrected estimators computed from the bootstrap samples of observed values have the same expectation and bias as the kernel density estimator computed from the bootstrap sample of true values. For notational convenience, we refer to the measurement error-corrected estimators computed from a bootstrap sample as $\widehat{f}^*_{het}(x;h)$, $\widehat{f}^*_{wt}(x;h)$, and $\widehat{f}^*_{hom}(x;h)$. We also adopt the general notation $\widehat{f}^*_{\cdot}(x;h)$ to represent any of the three estimators.

The bootstrap estimate of the optimal bandwidth for a measurement error-corrected estimator is the value h that minimizes the mean integrated squared error between $\widehat{f}^*_{\cdot}(x;h)$ and $\widehat{f}_{naive}(x,\widehat{\tau})$,

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

The optimal bandwidth can be determined empirically by computing a large number of bootstrap estimates, $\widehat{f}^*_{\cdot,1}(x;h), \ldots, \widehat{f}^*_{\cdot,B}(x;h)$, over a dense grid of bandwidths h, calculating the integrated squared error for each one, and selecting the value of h that achieves the smallest average integrated squared error. However the optimal bandwidth can also be determined analytically, as we show next. We present expressions for $\mathrm{MISE}^*(h)$ for each of the measurement error-corrected estimators. The optimal

bandwidth is found by numerically evaluating $\text{MISE}^*(h)$ over a grid of bandwidths h, and choosing the one for which $\text{MISE}^*(h)$ is smallest.

Analytical determination of the optimal bandwidth is possible due to the fact that in the bootstrap world, the true density of the data is known. The expectation of $\widehat{f}_\cdot^*(x,h)$ is given by

$$E^*\{\widehat{f}_\cdot^*(x;h)\} = \int Q(z)\widehat{f}_{naive}(x-hz;\widehat{\tau})dz, \tag{2.5.73}$$

where $Q(x)$ is the kernel used to calculate $f_\cdot^*(x;h)$. The bootstrap MISE in equation (2.5.71) has the representation

$$\text{MISE}^*(h) = \int \text{Var}^*\{\widehat{f}_\cdot^*(x;h)\}dx + \int \text{Bias}^{*2}\{\widehat{f}_\cdot^*(x;h)\}dx.$$

It follows from equation (2.5.73) that the integrated squared bias of $f_\cdot^*(x;h)$ is

$$\int \text{Bias}^{*2}\{\widehat{f}_\cdot^*(x;h)\}dx = \int \left\{\int Q(z)\widehat{f}_{naive}(x-hz;\widehat{\tau})dz - \widehat{f}_{naive}(x;\widehat{\tau})\right\}^2 dx. \tag{2.5.74}$$

The bootstrap integrated variance can be partitioned as

$$\int \text{Var}^*\{\widehat{f}_\cdot(x;h)\}dx = V_1^* - V_2^*,$$

where

$$V_1^* = \int E[\{\widehat{f}_\cdot(x;h)\}^2]dx$$

and

$$V_2^* = \int [E\{\widehat{f}_\cdot(x;h)\}]^2 dx$$

V_2^* is the same for each of the measurement error-corrected estimators, and from equation (2.5.73),

$$V_2^* = \frac{1}{n}\int \left\{\int Q(z)\widehat{f}_{naive}(x-hz)dz\right\}^2 dx. \tag{2.5.75}$$

V_1^* depends on the estimator used. Consider first the case of heteroscedastic measurement errors and the estimator $\widehat{f}^*_{het}(x;h)$. In Section 2.3.1, we derived an expression for the term V_1 in the integrated variance of the estimator $\widehat{f}_{het}(x;h)$ calculated from the original data. Nothing in that derivation depends on the true density $f_x(x)$. The only difference is the conditional variance of the the sample means, which in the bootstrap world are given by $\widehat{\sigma}^*_r/m_r$ for $r = 1, \ldots, n$. It follows from equation (2.3.34) that for the estimator $\widehat{f}^*_{het}(x;h)$,

$$V_1^* = \frac{1}{2\pi n^2 h} \sum_{r=1}^{n} \int \Phi_Q^2(t) \left\{ \sum_{k=0}^{\infty} \frac{\left(\frac{m_r-2}{2}\right)_k}{(m_r-2)_k} \frac{\left(\frac{2t^2\widehat{\sigma}^2_r}{m_r h^2}\right)^k}{k!} \right\} dt. \tag{2.5.76}$$

Combining this term with equations (2.5.74) and (2.5.75) yields the bootstrap MISE of the estimator $\widehat{f}^*_{het}(x;h)$.

Similarly, for the case of heteroscedastic errors and the weighted estimator, $\widehat{f}^*_{wt}(x;h)$, it follows from equation (2.3.48) that

$$V_1^* = \left\{ 2\pi h \sum_{r=1}^{n} \left\{ \int \Phi_Q^2(t) \left\{ \sum_{k=0}^{\infty} \frac{\left(\frac{m_r-2}{2}\right)_k}{(m_r-2)_k} \frac{\left(\frac{2t^2\widehat{\sigma}^2_r}{m_r h^2}\right)^k}{k!} \right\} dt \right\}^{-1} \right\}^{-1}, \tag{2.5.77}$$

which, with equations (2.5.74) and (2.5.75), yields the bootstrap MISE of the estimator $\widehat{f}^*_{wt}(x;h)$. Note in the bootstrap world, because $\widehat{\sigma}^2_1, \ldots, \widehat{\sigma}^2_n$ are known, the optimal weights for $\widehat{f}^*_{wt}(x)$ are known. Also note that the infinite sums in equations (2.5.76) and (2.5.77) simplify in the case of two replicates according to Lemma 2.3.1.1.

Finally, in Section 2.3.3 we derived an approximation to the integrated variance of $\widehat{f}_{hom}(x;h)$. We use this same approximation to estimate the bootstrap integrated

variance of $\widehat{f}^*_{hom}(x;h)$, which from from the MISE in equation (2.3.68) is given by

$$\int \text{Var}^*\{\widehat{f}^*_{hom}(x;h)\}dx \sim \frac{1}{2\pi n^2 h}\sum_{r=1}^{n}\int \Phi_Q(t)^2 e^{t^2\widehat{\sigma}^2/m_r h^2}dt.$$

Combining this with the bootstrap integrated bias in equation (2.5.74) yields the bootstrap MISE of $\widehat{f}^*_{hom}(x;h)$.

These computations are involved, but are less time-consuming than is resampling for empirical estimation of the bandwidth. Both approaches have potential drawbacks that stem from the deconvoluting kernel's sensitivity to too-small bandwidths. In the empirical averages, a single bootstrap sample that results in a highly variable density estimate can inflate the average integrated squared error of smaller candidate bandwidths, resulting in overestimation of the optimal bandwidth. An alternative is to choose the optimal bandwidth according to the median integrated squared error. Conversely, the analytical approach can lead to underestimation of the bandwidth, likely due to numerical instability in the computation of integrals.

2.6 An Application

We illustrate the bootstrap bandwidth procedure described in Section 2.5 with data from a study of automobile emissions. Measurements of carbon monoxide (CO) in automobile emissions were collected with a U.S. EPA remote sensing device, a device that is stationed along a roadway and measures compounds in the exhaust of automobiles as they pass by. For each passing automobile, the CO level in its exhaust was measured and its license plate was photographed. This allowed measurements

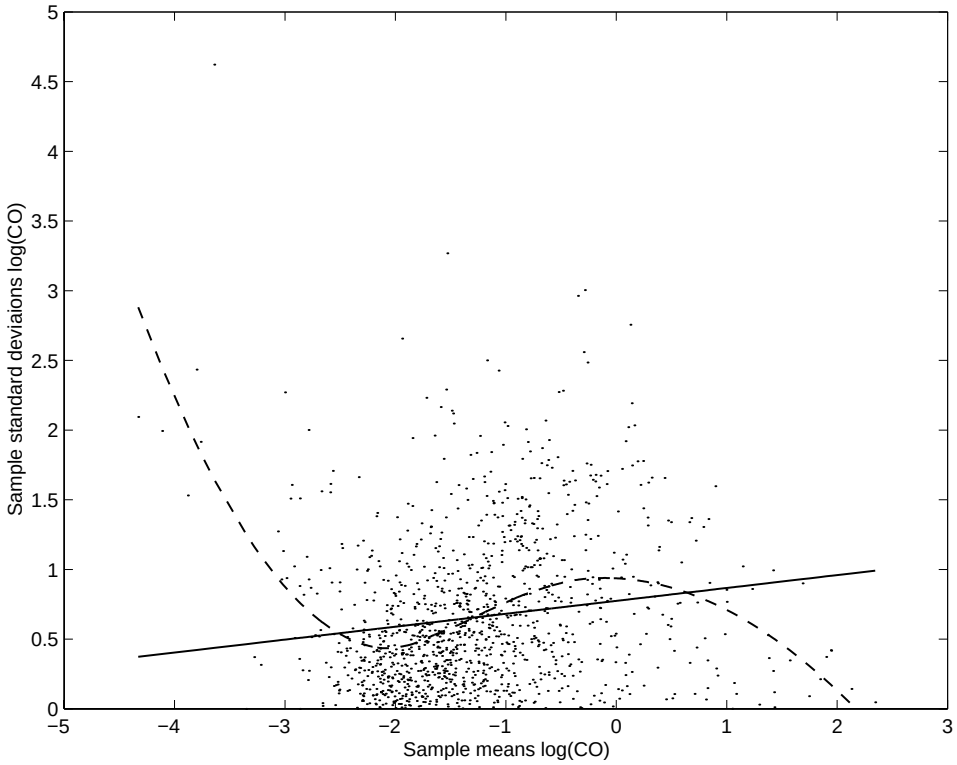

Figure 2.3: Relationship between sample means and sample standard deviations of log(CO) measurements. Solid line: simple linear regression; Dashed line: loess fit.

to be identified to individual automobiles. All measurements were taken at a single location, a highway entrance ramp in North Carolina, over several different dates, so that replicate measurements were available from cars that passed this location multiple times. Of a total of 3002 automobiles observed, 1233 were measured replicate times, 946 twice and 287 three times.

One objective of this study was to characterize CO emissions among the population of automobiles in use in North Carolina. The shape of the density function of CO emissions contains information about the subset of automobiles that are heavy polluters. This subset is of special interest to managers who set regulatory guidelines. The target variable in this study was the average CO emission of a car in use.

Measurements of CO taken from a stationary point on the roadway are subject to error from several sources. First, measurements are subject to variability from environmental conditions such as wind, temperature, and humidity. Second, any car's CO emission is not constant, but depends on factors such as acceleration and engine temperature. Thus, each single observation contains variability as a measurement of the car's long-term average CO emission. Finally, measurements are subject to error from the instrument itself. The observed data in this study, therefore, have multiple sources of error. These errors are compounded in every measurement, and are expected to exhibit more multiplicative than additive behavior. We performed our analysis on the logarithm of the measurements. This transforms the assumed multiplicative error structure into the additive structure for which our method is appropriate.

We estimated the density function of the log-transformed CO measurements from the 1233 automobiles that were measured replicate times. Data showed convincing evidence of heteroscedastic measurement errors. A simple linear regression of the sample standard deviations on the sample means of the transformed data resulted in a significant slope of 0.0936 ($p < .0001$), and a loess fit to the same data indicated a more complex, nonlinear relationship. Data and both fitted models are shown in Figure 3.

For comparison, we fit both the weighted estimator for heteroscedastic errors and the estimator for homoscedastic errors to the log-transformed CO data. Bandwidths for the estimators were selected using both analytical and empirical calculation of the

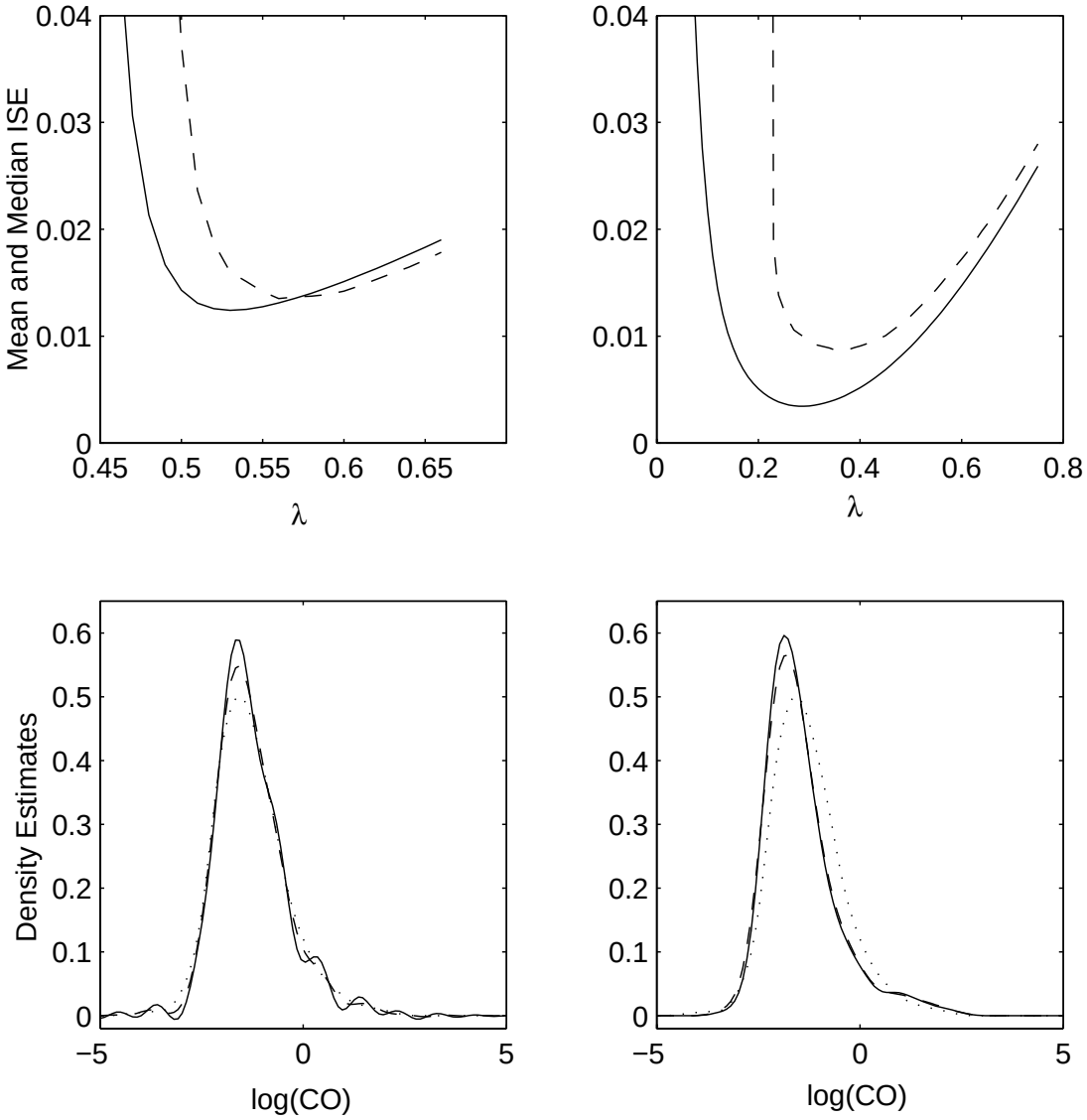

Figure 2.4: Top row: Analytical mean (Solid line) and empirical median (Dashed line) ISE of bootstrap density estimates as a function of bandwidth. Left: Homoscedastic estimator; Right: Weighted estimator; Bottom row: Density estimates with bandwidth minimizing bootstrap MISE (solid line) and median ISE (dashed line). Left: Homoscedastic estimator; Right: Weighted estimator. Dotted line in both panels is the naive estimator with plug-in bandwidth.

Table 2.1: Percentiles of the distribution of log(CO) from five density estimates: Naive kernel density estimate, Homoscedastic error estimator with bootstrap bandwidths minimizing analytical MISE and empirical median ISE, Weighted estimator with bootstrap bandwidths minimizing analytical MISE and empirical median ISE.

$P(x < X)$	X				
	$\widehat{f}_{naive}(x)$	$\widehat{f}_{hom}(x)$		$\widehat{f}_{wt}(x)$	
		Analytical	Empirical	Analytical	Empirical
0.5	-1.4646	-1.4646	-1.4646	-1.7677	-1.6667
0.75	-0.8586	-0.9596	-0.9596	-1.1616	-1.1616
0.9	-0.2525	-0.3535	-0.2525	-0.3535	-0.3535
0.95	0.2525	0.2525	0.2525	0.3535	0.3535
0.99	1.3636	1.2626	1.2626	1.6667	1.6667

bootstrap bandwidth estimate as described in Section 2.5. $\text{MISE}^*(\lambda)$ was evaluated along the grid $\lambda = 0.25, 0.26, \ldots, 0.66$ for the homoscedastic-error estimator and $\lambda = 0.01, 0.02, \ldots, 0.75$ for the weighted estimator. Each empirical estimate was based on 100 bootstrap data sets and the bandwidth was selected as the value that minimized the median integrated squared error.

Results are shown in Figure 2.4. Also shown is the naive estimate, calculated using the sample means of replicate observations and the plug-in bandwidth estimate from equation (2.4.70). Bandwidths estimated to minimize the analytical MISE were $\lambda = 0.53$ for the homoscedastic estimator and $\lambda = 0.29$ for the weighted estimator. Bandwidths estimated to minimize the empirical median ISE were $\lambda = 0.56$ for the homoscedastic estimator and $\lambda = 0.35$ for the weighted estimator. 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 the observed-data estimate. Particular interest in this application is in the upper tail of the density function. For instance, researchers might want to estimate the proportion automobiles whose average emission of CO falls above the legal limit for licensed vehicles. Table 2.6 gives percentiles from the density estimates computed from the data. Upper-tail critical values are largest for the weighted estimator.

Chapter 3

Regression-Assisted Deconvolution

3.1 Introduction

In this chapter, we present a semi-parametric deconvolution estimator for the density function of a random variable X that is measured as W, where $W = X + \sigma_u U$. The method assumes the availability of a covariate vector $\boldsymbol{Z}$ statistically related to $\boldsymbol{X}$, but independent of the error in measuring X, and such that the regression error $X - E(X|\boldsymbol{Z})$ is normally distributed.

Traditionally deconvolution has been studied in a univariate context in the sense that modeling assumptions are made only about the distribution of the true X and the measurement W given X. For many important cases even the best convergence rates for nonparametric estimation of $f_x(x)$ are slow. For example, when U is known to be normally distributed non-parametric deconvolution estimators achieve at best a logarithmic rate of convergence (Carroll & Hall 1988). The large sample size re-

quirements make these estimators impractical in many applications.

Although interest is often focused on the density of a particular variate X, in practice, most data sets are multivariate and thus contain covariates that are correlated with the error-prone variable. The method we present uses information in the covariate $\boldsymbol{Z}$. The idea is to exploit information in the covariates $\boldsymbol{Z}$ to obtain improved deconvolution estimators. The potential improvement is great. Our research is the first to investigate the use of covariate information in deconvolution problems. There are many variations to the basic strategy. We describe the approach in general, but we study only one particular version in this paper, concluding with recommendations for future research along these lines.

3.2 The General Method

We first describe the approach in general terms and then specialize to the particular version studied in this paper.

Let X be the random variable of interest with unknown density function $f_x(x)$. Suppose that X is observed only as W where $W = X + \sigma_u U$, U is a N(0, 1) random variable that is independent of X, and where σ_u is known. Let $\boldsymbol{Z}$ be a $p \times 1$ vector of covariates and suppose that the conditional mean and variance of X given $\boldsymbol{Z}$ are given by

$$E(X|\ \boldsymbol{Z}) = \mu_x(\boldsymbol{Z}, \beta) \quad \text{and} \quad \text{Var}(X|\ \boldsymbol{Z}) = \sigma_x^2(\mu_x(\boldsymbol{Z}, \beta), \theta) \tag{3.2.1}$$

respectively, where $\mu_x()$ and $\sigma_x^2()$ are known functions and β and θ are unknown

parameters. Note that apart from existence of the specified moments, the only restrictive modeling assumption is that the conditional variance of X given $\boldsymbol{Z}$ is a function of the conditional mean, i.e., depends on $\boldsymbol{Z}$ only through $\mu_x(\boldsymbol{Z}, \beta)$. The practical utility of such models in applications is well documented (see, for example, Carroll & Ruppert 1998).

We now add the assumption that the conditional distribution of X given $\boldsymbol{Z}$ is normal, leading to the regression model,

$$X_j = \mu_x(\boldsymbol{Z}_j,\ \beta) + \epsilon_j\ \sigma_x(\mu_x(\boldsymbol{Z}_j,\ \beta), \theta), \qquad j = 1, \ldots, n,$$

where the model errors, $\epsilon_1, \ldots, \epsilon_n$, are N(0, 1) random variables independent of $\boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_n$. Under the stated assumptions the density function of X is,

$$f_x(x) = \int_{-\infty}^{\infty} \frac{1}{\sigma_x(t,\ \theta)}\ \phi\left(\frac{x-t}{\sigma_x(t,\ \theta)}\right) f_\mu(t)dt, \tag{3.2.2}$$

where $f_\mu(t)$ is the density function of $\mu_x(\boldsymbol{Z}, \beta)$ and $\phi(t)$ is the standard normal density. We are implicitly assuming that the covariates are random, not fixed, and that $(X_j,\ \boldsymbol{Z}_j^T)^T$ are independent and identically distributed for $j = 1, \ldots, n$.

Assuming that the measurement error U is independent of $\boldsymbol{Z}$, it follows that the conditional mean and variance of W given $\boldsymbol{Z}$ are

$$E(W|\ \boldsymbol{Z}) = E(X|\ \boldsymbol{Z}) \quad \text{and} \quad \text{Var}(W|\ \boldsymbol{Z}) = \text{Var}(X|\ \boldsymbol{Z}) + \sigma_u^2,$$

and so from equation (3.2.1) we have that

$$E(W|\ \boldsymbol{Z}) = \mu_x(\boldsymbol{Z}, \beta)$$

and

$$\mathrm{Var}(W|\ \boldsymbol{Z}) = \sigma_w^2(\boldsymbol{Z}, \beta, \theta) = \sigma_x^2(\mu_x(\boldsymbol{Z}, \beta), \theta) + \sigma_u^2. \tag{3.2.3}$$

Similarly we have the regression model for W given $\boldsymbol{Z}$,

$$W_j = \mu_x(\boldsymbol{Z}, \beta) + \epsilon_j\ \sqrt{\sigma_u^2 + \sigma_x^2(\mu_x(\boldsymbol{Z}, \beta),\ \theta)}, \qquad j = 1, \ldots, n,$$

and the density function of W is,

$$f_w(w) = \int_{-\infty}^{\infty} \frac{1}{\sqrt{\sigma_x^2(t,\ \theta) + \sigma_u^2}}\ \phi\left(\frac{w - t}{\sqrt{\sigma_x^2(t,\ \theta) + \sigma_u^2}}\right) f_\mu(t)dt, \tag{3.2.4}$$

We exploit the relationships in equations (3.2.1)-(3.2.4) to construct an estimator of $f_x(x)$ that is parametric in the error models (the regression error, $X - \mu_x(\boldsymbol{Z}, \beta)$, and the measurement error, U) and nonparametric in the distribution of $T = \mu_x(\boldsymbol{Z}, \beta)$. The estimator requires regression modeling of W given $\boldsymbol{Z}$, and nonparametric density estimation using the fitted values from the regression model.

Consider a random sample of the observed data, $(W_j,\ \boldsymbol{Z}_j^T)^T$ for $j = 1, \ldots, n$. In the first stage of the analysis, mean and variance function models are fit to the data resulting in estimated mean and variance functions $\widehat{\mu}(\boldsymbol{Z},\ \widehat{\beta})$ and $\widehat{\sigma}_w^2(\widehat{\mu}(\boldsymbol{Z},\ \widehat{\beta}),\ \widehat{\theta})$. In principle, any reasonable method of estimation could be employed at this stage. The key is that the variance is modeled as a function of the mean.

Next the regression model predicted values, $\widehat{\mu}(\boldsymbol{Z}_j,\ \widehat{\beta}),\ j = 1, \ldots, n$, are used in the estimation of f_μ. This nonparametric step could be accomplished using any reasonable method of density estimation. For example, a kernel density estimator of

f_μ is given by

$$\widehat{f}_\mu(t) = \frac{1}{n\lambda}\sum_{j=1}^{n}\phi\left(\frac{t-\widehat{\mu}_x(\boldsymbol{Z}_j,\widehat{\beta})}{\lambda}\right), \tag{3.2.5}$$

where λ is the kernel bandwidth. Provided the model assumptions are valid and that the regression mean function is consistently estimated, the empirical distribution of the predicted values will converge to the distribution of T. It follows that for appropriate bandwidth sequences the kernel density estimator will converge to the density of T. The kernel density estimator estimates one component in the convolution formula for $f_x(x)$ in equation (3.2.2).

The other component in equation (3.2.2) is Gaussian and requires estimation of only the variance function of X given $\boldsymbol{Z}$. This stage is essentially components of variance estimation. In light of (3.2.3) an estimate of $\sigma_x^2(t,\ \theta)$ is given by

$$\widehat{\sigma}_x^2(t,\ \widehat{\theta}) = \widehat{\sigma}_w^2(t,\ \widehat{\theta}) - \sigma_u^2. \tag{3.2.6}$$

Finally, substituting the variance estimator (3.2.6) and the kernel estimator (3.2.5) into equation (3.2.2) results in the estimator

$$\widehat{f}_x(x) = \frac{1}{n\lambda}\sum_{j=1}^{n}\int_{-\infty}^{\infty}\frac{1}{\widehat{\sigma}_x(t,\ \widehat{\theta})}\phi\left(\frac{x-t}{\widehat{\sigma}_x(t,\ \widehat{\theta})}\right)\phi\left(\frac{t-\widehat{\mu}_x(\boldsymbol{Z}_j,\widehat{\beta})}{\lambda}\right)dt, \tag{3.2.7}$$

which we refer to as a regression-assisted deconvolution estimator of $f_x(x)$.

The regression-assisted deconvolution estimator is appealing for its reliance on regression methods that are familiar to statisticians. Modeling the conditional mean and variance of X given $\boldsymbol{Z}$ is critical. However the estimator does not rely on any particular technique for fitting these models. Indeed, models may be fit using any

appropriate regression method, including linear, nonlinear, semi-parametric and non-parametric methods. Equally important is the assumption that regression errors are normally distributed. Fortunately, violations of this assumption can be detected with a variety of techniques, and model transformations can often be employed to yield normally distributed regression errors.

In general, computing the regression-assisted deconvolution estimator requires numerical integration of the expression in equation (3.2.7). However, when the conditional variance of X given $\boldsymbol{Z}$ is constant, i.e., $\mathrm{Var}(X|\ \boldsymbol{Z}) = \sigma_x^2$, equation (3.2.7) simplifies to

$$\widehat{f}_x(x) = \frac{1}{n\sqrt{\widehat{\sigma}_x^2 + \lambda^2}} \sum_{j=1}^{n} \phi\left(\frac{x - \widehat{\mu}_x(\boldsymbol{Z}_j, \widehat{\beta})}{\sqrt{\widehat{\sigma}_x^2 + \lambda^2}}\right). \tag{3.2.8}$$

This form of the estimator also shows that the bandwidth parameter in equation (3.2.8) is superfluous whenever $\widehat{\sigma}_x^2$ is estimated well enough to avoid instability caused by $\widehat{\sigma}_x^2$ too close to, or less than, 0. For small to moderate sample sizes, when σ_w^2 is close to σ_u^2, it is possible $\widehat{\sigma}_x^2$ will be negative. However, when $\widehat{\sigma}_x^2$ is well-estimated, setting $\lambda = 0$ in equation (3.2.8) yields

$$\widehat{f}_x(x) = \frac{1}{n} \sum_{j=1}^{n} \frac{1}{\widehat{\sigma}_x}\ \phi\left(\frac{x - \widehat{\mu}_x(\boldsymbol{Z}_j, \widehat{\beta})}{\widehat{\sigma}_x}\right). \tag{3.2.9}$$

This form of the estimator is appealing because it avoids the need to estimate a bandwidth.

The scope of this paper is limited to exploring the feasibility of regression-assisted deconvolution via simulation and examples. However, a few general comments about the properties of the estimators are in order.

First suppose that finite-parameter, parametric models and estimation methods (e.g. least squares) are used to estimate the mean and variance functions $\mu_x(\boldsymbol{Z}, \beta)$ and $\sigma_x^2(\mu_x(\boldsymbol{Z}, \beta),\ \theta)$. In this case the mean and variance functions will converge pointwise at the $n^{1/2}$ rate. The simple estimator in equation (3.2.9) is a mean with summands that are continuous functions of $\sqrt{n}$-consistent parameter estimates, and thus would be expected to converge pointwise at the $\sqrt{n}$ rate. So in this case the regression-assisted estimator is expected to converge at the usual parametric rate of $n^{1/2}$, a substantial improvement over the logarithmic rates typically associated with deconvolution of normal measurement error. The estimator in equation (3.2.8) differs from that in equation (3.2.9) only because of the nonzero bandwidth λ. Note that the variance of the estimator in equation (3.2.8) does not diverge as $\lambda \to 0$, and thus there is no variance penalty asymptotically for letting the bandwidth shrink to 0 as $n \to \infty$. This suggests that for certain bandwidth sequences converging to 0 with increasing n, the estimator in equation (3.2.8) will also converge at a parametric rate. The same is expected to be true of the estimator in equation (3.2.7). The indicated conjecture is that for parametric regression modeling the deconvolution estimator converges at the $\sqrt{n}$ parametric rate. This assumes of course that the assumed parametric mean and variance function models, and the assumption of normal regression errors, are correct.

If the mean and variance functions are estimated nonparametrically (thereby eliminating the problem of model misspecification), then we speculate that the regression-assisted estimator will inherit the minimum of the convergence rates of the mean and

variance function estimates. These rates are typically much faster than logarithmic and thus there would still be a substantial improvement over standard deconvolution estimators. Nonparametric mean and variance function modeling eliminates one source of model misspecification, but the issue of normal regression errors remains.

In summary, the key components of this approach that determine the properties of the regression-assisted estimator are the mean and variance function estimates and normality of the regression errors. Fortunately, these are two areas that have been extensively studied in both the theoretical and applied statistics literature. The amount of "off-the-shelf" statistical technology that can be brought to bear on mean and variance function estimation and residual analysis is overwhelming. Thus in any particular application it should be possible to get very good mean and variance function estimates (assuming sufficient data), and also to get a fair assessment of normality of the residuals. Thus the potential for application is great, as is the likelihood of determining when the method is not applicable (nonnormal regression errors).

The remainder of this paper presents simulation results designed to explore the sensitivity of the method to certain key assumptions. The simulation results are followed by an illustrative application of the method to real data.

3.3 Simulations

The regression-assisted deconvolution estimator relies heavily on the regression model assumptions and normality of the regression errors. We performed a simulation study to determine how the success of the method is influenced by three key factors: first, the correct specification of the mean and variance functions, $\mu_x(\boldsymbol{Z}, \beta)$ and $\sigma_x^2(\mu(\boldsymbol{Z}, \beta), \theta)$, second, the precision of the estimated functions, $\widehat{\mu}_x(\boldsymbol{Z}, \widehat{\beta})$ and $\widehat{\sigma}_w^2(\widehat{\mu}(\boldsymbol{Z}, \widehat{\beta}), \widehat{\theta})$, and third, the normality of the regression errors. As this study represents a first look at the feasibility of using covariate information in deconvolution, we restricted our attention to relatively simple cases. All simulations considered multiple linear regression models with constant residual variance.

Simulations investigated how well the regression-assisted deconvolution estimator uncovers features such as bimodality and skewness in the true-data density. Three sets of simulations were performed, each considering a different density for X: 1) the standard normal, 2) a mixture of normals, and 3) a Chi-squared density.

Our interest is in small to moderate sample sizes (at least relative to the sample sizes needed for nonparametric deconvolution). We compared the naive kernel density estimator,

$$\widehat{f}_{naive}(x) = \frac{1}{n\lambda} \sum_{j=1}^{n} \phi\left(\frac{x - W_j}{\lambda}\right), \tag{3.3.10}$$

the true-data kernel density estimator,

$$\widehat{f}_{true}(x) = \frac{1}{n\lambda} \sum_{j=1}^{n} \phi\left(\frac{x - X_j}{\lambda}\right),$$

and the regression-assisted deconvolution estimator in equation (3.2.8). In all cases, the bandwidth parameter λ was calculated using the plug-in method,

$$\lambda = 1.06\widehat{\sigma}_*^2 n^{-1/5}, \tag{3.3.11}$$

where $\widehat{\sigma}_*^2$ is the sample variance of the data used to compute the estimator (Silverman 1986). All simulated data sets contained $n = 100$ observations. The regression-assisted deconvolution estimator in equation (3.2.8) requires an estimate of the model error variance, σ_x^2. This was estimated as $\widehat{\sigma}_x^2 = \text{MSE} - \sigma_u^2$, where MSE is the estimate of mean squared error from the regression of W on $\boldsymbol{Z}$. With the sample size $n = 100$, the mean squared error is not estimated well enough to avoid occasional negative values of $\widehat{\sigma}_x^2$. In such cases, $\widehat{\sigma}_x^2$ was set to 0.

Estimators were compared on the basis of their integrated squared error (ISE), averaged over 200 simulated data sets. The integrated squared error is defined in general for an estimator $\widehat{f}(x)$ as

$$\text{ISE}\{\widehat{f}(x)\} = \int_{-\infty}^{\infty} \{\widehat{f}(x) - f_x(x)\}^2 dx.$$

3.3.1 Simulation 1: Estimation When $f_x(x)$ is the Standard Normal Density

This simulation examined the performance of the regression-assisted deconvolution estimator when $f_x(x)$ is the N(0, 1) density and the conditional mean and variance of X given $\boldsymbol{Z}$ are

$$E(X|\ \boldsymbol{Z}) = \beta^T \boldsymbol{Z} \quad \text{and} \quad \text{Var}(X|\ \boldsymbol{Z}) = \sigma_x^2. \tag{3.3.12}$$

We considered the linear model where for $j = 1, \ldots, n$,

$$X_j = \beta^T \boldsymbol{Z}_j + \sigma_x \epsilon_j, \tag{3.3.13}$$

$\boldsymbol{Z}_j$ is a $p \times 1$, N($\mathbf{0}$, I_p) random vector, and ϵ_j is a N(0, 1) random variable that is independent of $\boldsymbol{Z}_j$. It follows that $X_1, \ldots, X_n$ is a random sample from the normal distribution with mean 0 and variance $\beta^T\beta + \sigma_x^2$. Adding the constraint

$$\beta^T\beta + \sigma_x^2 = 1 \tag{3.3.14}$$

results in the sample $X_1, \ldots, X_n$ of independent N(0, 1) random variables.

The addition of normally distributed measurement error to equation (3.3.13) yields the observed-data model,

$$W_i = \beta^T \boldsymbol{Z}_j + \sigma_x \epsilon_j + \sigma_u U_j, \tag{3.3.15}$$

where for $j = 1, \ldots, n$, U_j is a N(0, 1) measurement error, independent of both $\boldsymbol{Z}_j$ and ϵ_j. $W_1, \ldots, W_n$ is a random sample from the normal distribution with mean 0 and variance $\beta^T\beta + \sigma_x^2 + \sigma_u^2 = 1 + \sigma_u^2$.

Several factors were controlled in the study to allow examination of the regression-assisted deconvolution estimator's dependence on key components of the regression model. We discuss these factors next in detail. Factors and their levels are summarized in Table 3.3.1. To a large extent, these factors determined the way data were generated for the simulation. Our discussion of the study factors is followed by a description of the methods we used to generate data. The section concludes with a presentation and discussion of the simulation results.

Two factors in the study examined how the estimator's performance is affected by the precision with which the true data can be predicted from the regression of W on $\boldsymbol{Z}$. The strength of the relationship between X and $\boldsymbol{Z}$ affects the precision of the predicted values and was controlled through the theoretical coefficient of determination of the regression of X on $\boldsymbol{Z}$,

$$R^2 = \frac{\text{Var}(\beta^T \boldsymbol{Z})}{\text{Var}(X)} = \frac{\beta^T \beta}{\beta^T \beta + \sigma_x^2}. \tag{3.3.16}$$

R^2 was varied at five levels, 0.1, 0.3, 0.5, 0.7 and 0.9. Note that under the constraint in equation (3.3.14),

$$R^2 = \beta^T \beta = 1 - \sigma_x^2. \tag{3.3.17}$$

The measurement error variance also influences the precision of the predicted values. The reliability ratio, ξ, describes the ratio of the variance in the true data to the variance in the observed data,

$$\xi = \frac{\text{Var}(X)}{\text{Var}(W)} = \frac{1}{1 + \sigma_u^2}. \tag{3.3.18}$$

The reliability ratio was also included as a factor in the study and was varied at two levels, 0.7 and 0.9.

A third factor was comprised of the type of estimator used to estimate $f_x(x)$. This factor was designed to investigate the effects of model misspecification on the performance of the regression-assisted deconvolution estimator. Five different estimators made up the levels of this factor. Among these were the naive kernel estimator, $\widehat{f}_{naive}(x)$, the true-data kernel estimator, $\widehat{f}_{true}(x)$, and the regression-assisted deconvolution estimator, $\widehat{f}_x(x)$. As defined, $\widehat{f}_x(x)$ is computed with estimates derived

from the observed-data regression where the mean and variance functions in equation (3.3.12) are known. For the case considered here, this implies that $\widehat{f}_x(x)$ is computed with estimates from the linear regression of W on $\boldsymbol{Z}$, where $\boldsymbol{Z}$ is the vector of p covariates that are correlated with X. Two additional estimators were computed to investigate the effects of misspecifying the mean function by incorrectly identifying p. The regression-assisted deconvolution estimator was computed with estimates from two incorrectly specified models, one an under-fit model with $p_u < p$ covariates and the other an over-fit model with $p_o > p$ covariates. Covariates for the misspecified models were determined as follows. Each simulated data set contained a total of twelve covariates, $p = 4$ of which were correlated with X according to the linear model in equation (3.3.13). The remaining 8 covariates were uncorrelated with X. An under-fit model was constructed by regressing observed data on the entire set of covariates and choosing those $p_u = 2$ covariates whose estimated coefficients had the largest absolute t-statistics. An over-fit model was constructed similarly, choosing the $p_o = 8$ covariates whose estimated coefficients had the largest absolute t-statistics. Predicted values and estimated variances from the under-fit and over-fit models were used to compute the regression-assisted deconvolution estimator, and we denote these estimators by $\widehat{f}_{und}(x)$ and $\widehat{f}_{ovr}(x)$ respectively.

A fourth factor in the study influenced the likelihood that models would be misspecified. This likelihood is determined in part by the strength of the correlations between X and the p true covariates. The strength of these correlations was controlled through assignment of regression coefficients in the parameter vector β. Recall that $\boldsymbol{Z}$

Table 3.1: Factors and levels included in Simulation 1.

Factor	**Levels**
R^2	0.1 0.3 0.5 0.7 0.9
ξ	0.7 0.9
Estimator	$\widehat{f}_x(x)$ $\widehat{f}_{und}(x)$ $\widehat{f}_{ovr}(x)$ $\widehat{f}_{naive}(x)$ $\widehat{f}_{true}(x)$
Coefficient Pattern	$[1,1,1,1]^T$ $[1,1,1/2,1/2]^T$ $[1,3/4,1/2,1/4]^T$
$f_\epsilon(\epsilon)$	N(0, 1) Chi-squared(16) Chi-squared(4)

is a N($\mathbf{0}$, I_p) random vector. Thus, the covariates in $\boldsymbol{Z}$ are mutually independent and have equal variances, so that the relative magnitudes of their individual coefficients determines the relative strengths of their correlations with X. A coefficient pattern vector was used to appropriately scale β and control these relative correlations. From equation (3.3.17) it follows that each value of the factor R^2 determines $\beta^T\beta$. The parameter vector β was calculated as

$$\beta = \beta_* \sqrt{R^2/(\beta_*^T \beta_*)}, \tag{3.3.19}$$

where the components of β_* determine the relative magnitudes of the correlations between X and the covariates in $\boldsymbol{Z}$. With $p = 4$, β_* is a 4×1 vector. Three different values of β_* were used, $[1,1,1,1]^T$, $[1,1,1/2,1/2]^T$, and $[1,3/4,1/2,1/4]^T$. Thus, our simulations considered three correlation patterns between the true data and the covariates, constant correlations, two levels of correlations, and linearly decreasing correlations.

Finally, perhaps most critical to the performance of the regression-assisted de-

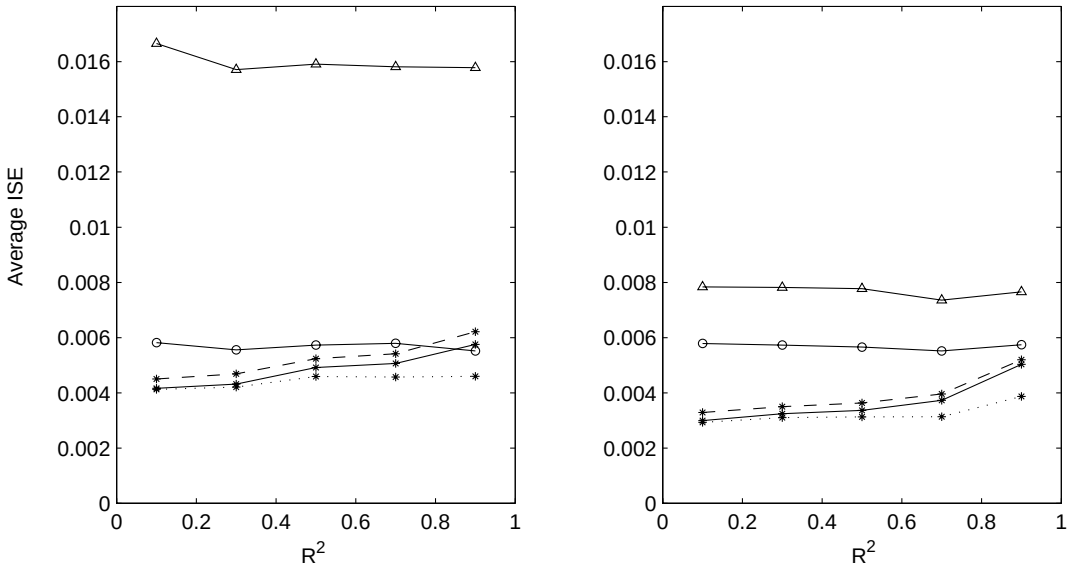

Figure 3.1: Average ISE by R^2 for Simulation 1 with normal model errors. Left: ξ=0.7. Right: ξ=0.9. Circle: $\widehat{f}_{true}(x)$; Triangle: $\widehat{f}_{naive}(x)$; Star, solid line: $\widehat{f}_{x}(x)$; Star, dotted line: $\widehat{f}_{und}(x)$; Star, dashed line: $\widehat{f}_{ovr}(x)$. Pooled standard error of mean ISE is 0.00017.

convolution estimator is the assumption that the density of the model errors, $f_\epsilon(\epsilon)$, is normal. Errors that are skewed are commonly seen in practice, and we investigated the estimator's robustness to this departure from normality. In addition to the normal density, model errors were generated from the standardized Chi-squared(16) density and the standardized Chi-squared(4) density. Note from equation (3.3.13) that in general, when model errors are standardized Chi-square(d) random variables, the density function of X is the convolution of a normal density and the Chi-square(d) density, centered at 0 and scaled to have variance σ_x^2. From equation (3.3.17) it is seen that in the simulations where model errors are Chi-square distributed, $f_x(x)$ changes with R^2, its skewness increasing as R^2 decreases.

The assumption of normal regression errors plays an important role in the regression-

assisted estimator and departures from this assumption will negatively affect the estimator. As part of our simulation we included a study of the power of a common test for detecting nonnormality of residuals. Our intent was to determine whether it would be possible to detect nonnormal regression residuals in those cases where the extent of nonnormality adversely affects the performance of the regression-assisted estimator. In practice, regression residuals can be tested for normality with a variety of available methods. The regression errors from the fit of the observed-data model in equation (3.3.15) have density function defined by the convolution of $f_\epsilon(\epsilon)$ and the N(0, σ_u^2) density. Residuals from the correctly-specified regression model were tested for normality with the D'Agostino-Pearson K^2 statistic (D'Agostino, Belanger & D'Agostino 1990), which has higher power than many others for detecting non-normal skewness and kurtosis in data.

For every combination of factor levels, 200 data sets with $n = 100$ observations were generated as follows. Covariate vectors were generated first. For all data sets, the number of covariates that had nonzero correlations with X was $p = 4$. $\boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_n$ were generated as N(0, I_4) random vectors. Second, the parameter β was determined. For given levels of the factors R^2 and the coefficient pattern, β was calculated from equation (3.3.19). Third, model errors, $\epsilon_1, \ldots, \epsilon_n$ were generated from their standardized density, $f_\epsilon(\epsilon)$, and were scaled to have variance $\sigma_x^2 = 1 - R^2$. These three components were combined according to the true-data linear model in equation (3.3.13), resulting in the true-data sample, $X_1, \ldots, X_n$. Next, an observed-data sample was generated according to equation (3.3.15) by adding normal measurement

Table 3.2: Power of the D'Agostino-Pearson K^2 test in Simulation 1 to detect non-normality in observed-data regression residuals from the correct model fit, when true model errors are standardized Chi-square(4) and Chi-square(16) random variables.

	Chi-square(4)		Chi-square(16)	
R^2	$\xi = 0.7$	$\xi = 0.9$	$\xi = 0.7$	$\xi = 0.9$
0.1	0.54	0.87	0.24	0.40
0.3	0.47	0.82	0.18	0.36
0.5	0.36	0.75	0.15	0.33
0.7	0.26	0.63	0.10	0.27
0.9	0.10	0.31	0.06	0.14

errors to the true data. Given ξ, the measurement error variance was calculated from equation (3.3.18) as $\sigma_u^2 = \xi^{-1} - 1$. Standard normal errors, $U_1, \ldots, U_n$ were scaled by σ_u^2 and added to the true data to form the sample of observations, $W_1, \ldots, W_n$. In a final step, extra covariates were generated as 8×1, $\mathrm{N}(\mathbf{0}, I_8)$ vectors, independent of $X_1, \ldots, X_n$ and $\boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_n$, for computing under-fit and over-fit regression models. An under-fit model was computed with $p_u = 2$ of the twelve total covariates, and an over-fit model was computed with $p_o = 8$ of the twelve covariates.

We discuss the simulation results first for the case of normally distributed model errors. An analysis of variance (ANOVA) calculated with the factors, R^2, ξ, Estimator, and Coefficient Pattern showed that 99% of the variation in integrated squared error that was explained by these factors could be attributed to the main-effects, two-way interactions, and three-way interaction of R^2, ξ, and Estimator. The effect of the coefficient pattern vector was not significant, and reported results are pooled

over this factor.

Average integrated squared errors are plotted in Figure 3.1 and reveal that in all cases considered, the regression-assisted deconvolution estimator yields a significantly smaller integrated squared error than the naive estimator. It is even superior to the true-data estimator for small values of R^2. This seemingly anomalous finding is explained by the fact that the true-data kernel density estimator does not make use of any assumptions about the distribution of X. Underlying the construction of the regression-assisted estimator is the implicit assumption that the density of X has a normal component. The regression-assisted estimator exploits this assumption and thus it is not surprising that it can beat the true-data estimator when the assumption is true, as it is in the case under consideration.

The regression-assisted deconvolution estimator calculated from the under-fit model performs significantly better when R^2 is large than does the estimator calculated from either the correctly-specified or over-fit models. This is explained by the fact that both the data and errors in the regression of W on $\boldsymbol{Z}$ are normally distributed. In this case, the additional error from fitting a model with too few covariates is normally distributed, and is incorporated into the regression errors with no violation of assumptions. Thus, apart from the effect of selecting the best two predictors, the under-fit model is not truly misspecified. The fact that it is sometimes better is likely due to the fewer number of parameters that are estimated in the under-fit model.

That the regression-assisted deconvolution estimator performs well when the assumption of normal model errors is met is not surprising. We now discuss the results

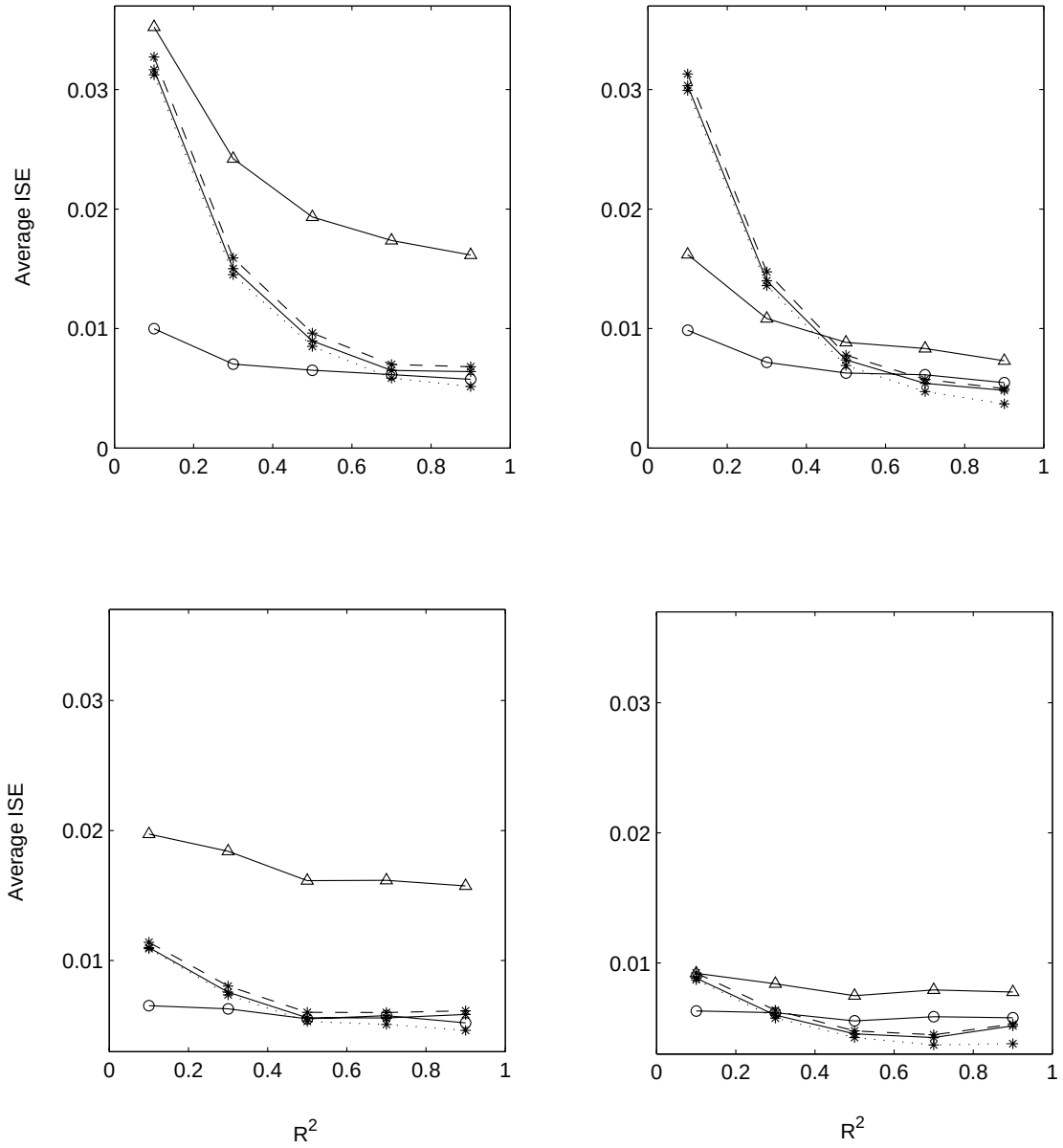

Figure 3.2: Average ISE by R^2 for Simulation 1 with standardized Chi-square model errors. Top row: Standardized Chi-square(4) errors; Left: $\xi = 0.7$; Right: $\xi = 0.9$. Bottom row: Standardized Chi-square(16) errors; Left: $\xi = 0.7$; Right: $\xi = 0.9$. Circle: $\widehat{f}_{true}(x)$; Triangle: $\widehat{f}_{naive}(x)$; Star, solid line: $\widehat{f}_{x}(x)$; Star, dotted line: $\widehat{f}_{und}(x)$; Star, dashed line: $\widehat{f}_{ovr}(x)$. Standard error of mean ISE is 0.00027 for Chisquare(4) errors and 0.00021 for Chisquare(16) errors.

of simulations that considered standardized Chi-squared(16) and standardized Chi-squared(4) model errors. In both sets of simulations, the main-effects and interactions among the factors R^2, ξ, and Estimator accounted for over 99% of the explained variability in the integrated squared error, and results were again pooled over the coefficient-pattern factor. Figure (3.2) shows average integrated squared errors for these simulations and suggests that the regression-assisted deconvolution estimator should be used cautiously when regression errors are believed to be non-normal. The integrated squared error is most strongly influenced by non-normality when R^2 is small and the distribution of the regression errors is highly skewed. However, non-normality was detected consistently in these cases with the D'Agostino-Pearson K^2 test. The power of this test for detecting non-normality in the residuals from the correct model fit is summarized in Table 3.3.1. There was little or no effect on the performance of the regression-assisted deconvolution estimator from under-fitting or over-fitting the regression model.

3.3.2 Simulation 2: Estimation When $f_x(x)$ is a Normal Mixture Density

The aim of this simulation was to develop an understanding of how well the regression-assisted deconvolution estimator uncovers bimodal features in the true-data density. We investigated the case where $f_x(x)$ is a normal mixture density and the conditional mean and variance of X given $\boldsymbol{Z}$ are given by equation (3.3.12), the

same as in Simulation 1. We considered the true-data model

$$X_j = -b\alpha + bY_j + \beta^T \boldsymbol{Z}_j + \sigma_x \epsilon_j, \qquad j = 1, \ldots, n, \tag{3.3.20}$$

where Y_j is a Bernoulli(α) random variable, $\boldsymbol{Z}_j$ is a $p \times 1$, N($\mathbf{0}$, I_p) random vector, ϵ_j is a N(0, 1) random variable, and Y_j, $\boldsymbol{Z}_j$, and ϵ_j are mutually independent. It follows that $X_1, \ldots, X_n$ is a random sample from an $\{\alpha : (1-\alpha)\}$ mixture of normals having means $b(1-\alpha)$ and $-b\alpha$ respectively, and common variances $\beta^T\beta + \sigma_x^2$. From equation (3.3.20), it is seen that $E(X) = 0$ and $\text{Var}(X) = b^2\alpha(1-\alpha) + \beta^T\beta + \sigma_x^2$. We standardized X by constraining this variance to equal one, i.e.

$$b^2\alpha(1-\alpha) + \beta^T\beta + \sigma_x^2 = 1. \tag{3.3.21}$$

Adding normal measurement error to equation (3.3.20) results in the observed-data model,

$$W_j = -b\alpha + bY_j + \beta^T \boldsymbol{Z}_j + \sigma_x \epsilon_j + \sigma_u U_j,$$

where U_j is a N(0, 1) random variable that is independent of Y_j, $\boldsymbol{Z}_j$, and ϵ_j for $j = 1, \ldots, n$. The observed data form a random sample from an $\{\alpha : (1-\alpha)\}$ mixture of normals with means $b(1-\alpha)$ and $-b\alpha$, and common variances $\beta^T\beta + \sigma_x^2 + \sigma_u^2 = 1 + \sigma_u^2$.

The same factors were considered in this simulation as were considered in Simulation 1. Factors and their levels are summarized in Table 3.3.2. Note that for the model in equation (3.3.20), the theoretical coefficient of determination is

$$R^2 = \frac{\text{Var}(bY_j + \beta^T \boldsymbol{Z})}{\text{Var}(X)} = \frac{b^2\alpha(1-\alpha) + \beta^T\beta}{b^2\alpha(1-\alpha) + \beta^T\beta + \sigma_x^2}.$$

Table 3.3: Factors and levels included in the Simulation 2.

Factor	**Levels**
R^2	0.1 0.3 0.5 0.7 0.9
ξ	0.7 0.9
Estimator	$\widehat{f}_x(x)$ $\widehat{f}_{und}(x)$ $\widehat{f}_{ovr}(x)$ $\widehat{f}_{naive}(x)$ $\widehat{f}_{true}(x)$
Coefficient Pattern	$[1,1,1]^T$ $[1,1/2,1/2]^T$ $[1,2/3,1/3]^T$
$f_\epsilon(\epsilon)$	N(0, 1) Chi-squared(16) Chi-squared(4)

R^2 was varied at the five levels, 0.1, 0.3, 0.5, 0.7 and 0.9. From equation (3.3.21),

$$R^2 = b^2\alpha(1-\alpha) + \beta^T\beta. \tag{3.3.22}$$

This induces the constraint that $R^2 \geq b^2\alpha(1-\alpha)$. The parameters b and α determine the appearance of distinct modes in $f_x(x)$, and for small values of R^2, the modes of densities that satisfy this constraint are obscured. In this simulation, we allowed the true-data density to vary with R^2. We fixed $\alpha = 0.7$ and chose b close to its maximum value for each R^2, resulting in the variety of shapes for the true-data density displayed in Figure 3.3. Although this complicates comparisons of estimators among levels of R^2, comparisons within levels of R^2 are straightforward.

The reliability ratio, ξ in equation (3.3.18), was also included as a factor in this simulation and was varied at the two levels 0.7 and 0.9.

To investigate the effects of model misspecification, we again generated extra covariates for computing under-fit and over-fit regression models. The covariates Y and a 3×1 vector $\boldsymbol{Z}$ were related to X by the model in equation (3.3.20). These were

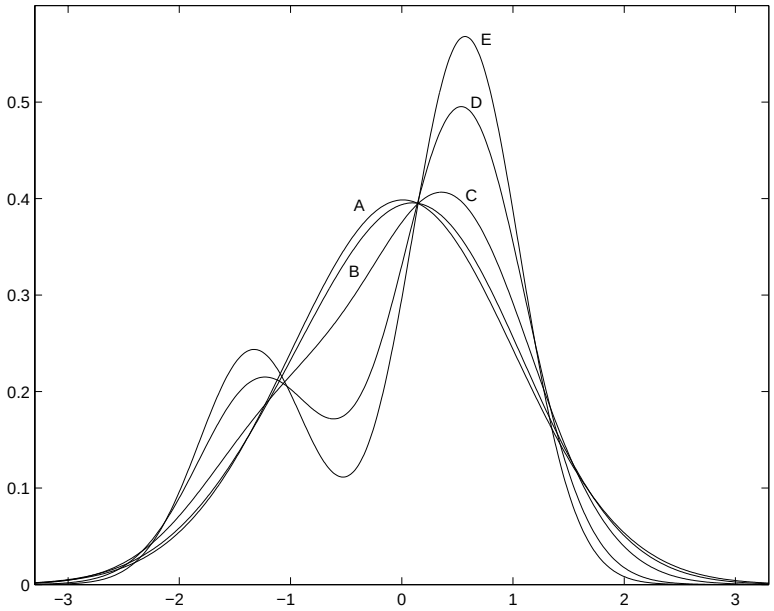

Figure 3.3: 70-30 normal mixture densities with means $\pm b$ for different values of R^2. A: $R^2 = 0.1, b = 0.6$; B: $R^2 = 0.3, b = 1.1$; C: $R^2 = 0.5, b = 1.5$; D: $R^2 = 0.7, b = 1.8$; E: $R^2 = 0.9, b = 1.9$.

used to fit the correctly-specified model and compute the estimator $\widehat{f}_x(x)$. Additional 8×1 vectors of covariates were generated and were with X, and were used to construct under-fit and over-fit regression models for computing the estimators $\widehat{f}_{und}(x)$ and $\widehat{f}_{ovr}(x)$. Observed data were regressed on all twelve covariates. The under-fit model was computed with the covariates corresponding to the largest $p_u = 2$ absolute t-statistics, and the over-fit model with the covariates corresponding to the largest $p_o = 8$ absolute t-statistics.

Also considered was the pattern of correlations between X and $\boldsymbol{Z}$. In this simulation, the 3×1 vector β is determined by R^2, b, and α, and from equation (3.3.22),

$$\beta^T \beta = R^2 - b^2 \alpha (1 - \alpha).$$

The pattern of the regression coefficients was controlled by scaling β with the coeffi-

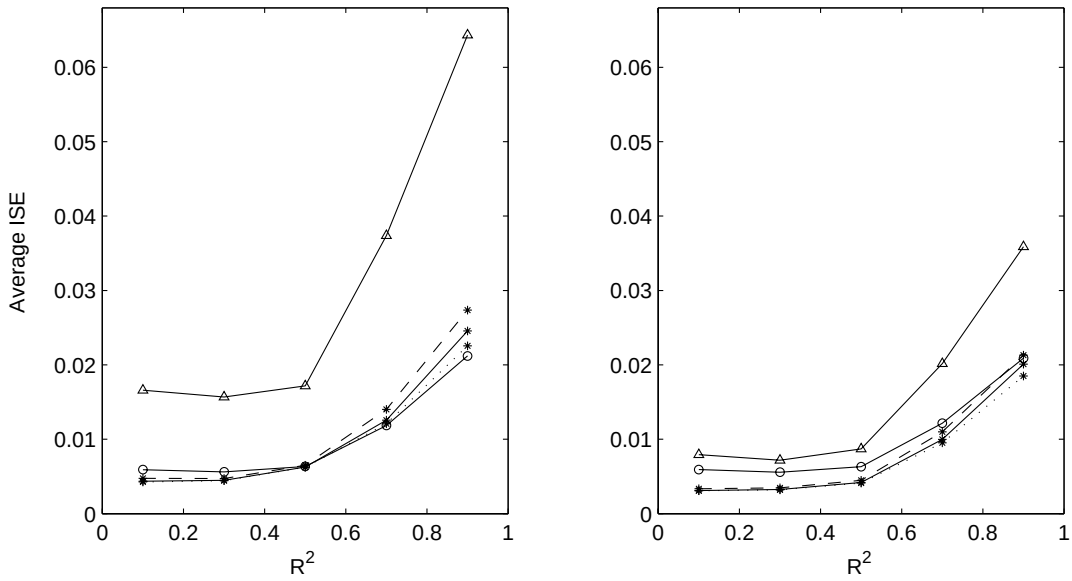

Figure 3.4: Average ISE by R^2 for Simulation 2 with normal model errors. Left: $\xi = 0.7$. Right: $\xi = 0.9$. Circle: $\widehat{f}_{true}(x)$; Triangle: $\widehat{f}_{naive}(x)$; Star, solid line: $\widehat{f}_{x}(x)$; Star, dotted line: $\widehat{f}_{und}(x)$; Star, dashed line: $\widehat{f}_{ovr}(x)$. Pooled standard error of mean ISE is 0.00025.

cient pattern vector β_*,

$$\beta = \beta_* \sqrt{\{R^2 - b^2 p(1-p)\}/(\beta_*^T \beta_*)}. \tag{3.3.23}$$

Three different patterns were considered by assigning to β_* the values $[1,1,1]^T$, $[1,1/2,1/2]^T$ and $[1,2/3,1/3]^T$.

The robustness of the regression-assisted deconvolution estimator to skewed model errors was investigated by generating model errors from standardized Chi-squared(16) and Chi-squared(4) densities. Note that in general, when model errors follow a standardized Chi-squared(d) density, $f_x(x)$ is the convolution of a normal mixture density and the Chi-squared(d) density centered at 0 and scaled to have variance σ_x^2. The likelihood that these departures from normality would be detected in the residuals from the regression of W on $(Y,\ \boldsymbol{Z})$ was investigated for the correctly-specified model.

As in Simulation 1, regression residuals from each simulated data set were tested with the D'Agostino-Pearson K^2 statistic.

A total of 200 data sets with $n = 100$ observations were generated for each combination of the factor levels. Each data set was generated as follows. First, $Y_1, \ldots, Y_n$ were generated as independent Bernoulli(0.7) random variables, and $\boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_n$ were generated as independent N($\mathbf{0}$, I_3) random vectors. Second, given levels of R^2 and the coefficient pattern factor, β was calculated from equation (3.3.23). Third, model errors, $\epsilon_1, \ldots, \epsilon_n$, were generated from $f_\epsilon(\epsilon)$ and scaled by $\sigma_x^2 = 1 - R^2$. With b determined by the level of R^2, these components were combined according to equation (3.3.20) to form the true-data sample, $X_1, \ldots, X_n$. Next the observed-data sample was generated. Given ξ, N(0, 1) measurement errors, $U_1, \ldots, U_n$ were generated and scaled to have variance $\sigma_u^2 = \xi^{-1} - 1$. Adding these errors to the true data produced the observed-data sample $W_1, \ldots, W_n$. Finally, for computing the over-fit and under-fit regression models, additional N($\mathbf{0}$, I_8) random variables were generated independently of $X_1, \ldots, X_n$ and $(Y_1, \boldsymbol{Z}_1), \ldots, (Y_n, \boldsymbol{Z}_n)$.

Average integrated squared errors are displayed in Figure 3.4 for normal model errors and Figure 3.5 for Chi-square model errors. In all cases, an analysis of variance model computed with the factors R^2, ξ, and Estimator and Coefficient Pattern showed that at least 99% of the explained variation in the integrated squared error could be attributed to the main-effects, two-way interactions, and three-way interaction of R^2, ξ, Estimator, and their interactions. Results are pooled over the factor Coefficient Pattern. When model errors are normally distributed, the regression-

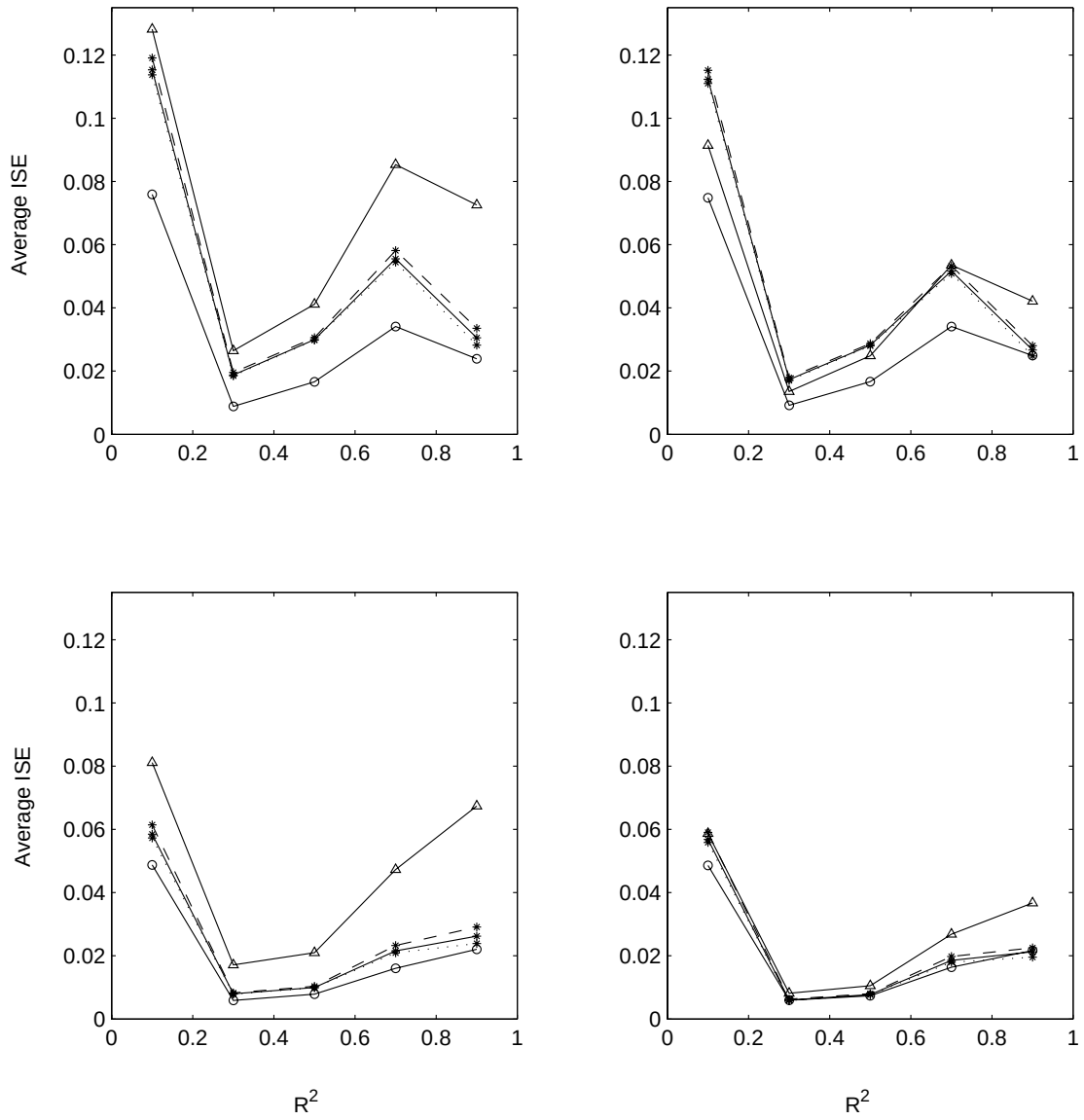

Figure 3.5: Average ISE by R^2 for Simulation 2 with standardized Chi-square model errors. Top row: Standardized Chi-square(4) errors; Left: $\xi = 0.7$; Right: $\xi = 0.9$. Bottom row: Standardized Chi-square(16) errors; Left: $\xi = 0.7$; Right: $\xi = 0.9$. Circle: $\widehat{f}_{true}(x)$; Triangle: $\widehat{f}_{naive}(x)$; Star, solid line: $\widehat{f}_{x}(x)$; Star, dotted line: $\widehat{f}_{und}(x)$; Star, dashed line: $\widehat{f}_{ovr}(x)$. Standard error of mean ISE is 0.00048 for Chisquare(4) errors and 0.00036 for Chisquare(16) errors.

Table 3.4: Power of the D'Agostino-Pearson K^2 test in Simulation 2 to detect non-normality in observed-data regression residuals from the correct model fit, when true model errors are standardized Chi-square(4) and Chi-square(16) random variables.

	Chi-square(4)		Chi-square(16)	
R^2	$\xi = 0.7$	$\xi = 0.9$	$\xi = 0.7$	$\xi = 0.9$
0.1	0.64	0.90	0.23	0.43
0.3	0.56	0.87	0.25	0.39
0.5	0.41	0.83	0.16	0.33
0.7	0.26	0.70	0.09	0.25
0.9	0.09	0.33	0.07	0.11

assisted deconvolution estimator is superior to the naive estimator for all values of R^2 and ξ, and performs at least as well as the true-data estimator. The effects of model misspecification were generally not significant.

The regression-assisted deconvolution estimator is less effective when model errors are Chi-squared distributed, particularly when errors are highly skewed and R^2 is small. Our estimator performed significantly worse than the naive estimator when errors were Chi-squared(4) distributed, R^2 was equal to 0.1, and ξ was equal to 0.9. In general, the D'Agostino-Pearson K^2 statistic was effective in detecting non-normality in the residuals from the regression of W on $\boldsymbol{Z}$ in the situations where the regression-assisted deconvolution estimator performed weakly. The power of this test is summarized in Table 3.3.2 for residuals from the correctly specified model.

3.3.3 Simulation 3: Estimation When $f_x(x)$ is the Standardized Chi-squared(9) Density

In this simulation, we investigated the regression-assisted deconvolution estimator's ability to uncover skewness in the true-data density. We examined the case where $f_x(x)$ is the Chi-square(9) density, standardized to have mean 0 and variance 1. Our objective was to consider the model where for $j = 1, \ldots, n$,

$$X_j = \beta^T \boldsymbol{Z}_j + \sigma_x \epsilon_j,$$

$\boldsymbol{Z}_j$ is a $p \times 1$ vector of covariates and ϵ_j is a N(0, 1) random variable that is independent of $\boldsymbol{Z}_j$.

In Simulations 1 and 2, we specified appropriate distributions for the covariates and the model errors to yield a random sample of true data from the desired density, $f_x(x)$. However with ϵ normally distributed, it is not possible to specify $\boldsymbol{Z}$ so that the density of X in equation (3.3.28) is Chi-squared. Instead, we specified distributions for the true data and model errors first, and used these to determine the covariate vector. Let $\boldsymbol{c} = [c_1, \ c_2, \ldots, c_p]^T$ be a vector of constants, and set

$$\boldsymbol{Z} = \boldsymbol{c}X + \sigma_x \epsilon, \tag{3.3.24}$$

where the components of $\epsilon = [\epsilon_1, \ \epsilon_2, \ldots, \epsilon_p]^T$ are independent N(0, 1) random variables, and are independent of X. We define the parameter vector γ as the theoretical least squares estimate of $\boldsymbol{c}$, i.e.,

$$\gamma = \{E(\boldsymbol{Z}\boldsymbol{Z}^T)\}^{-1} E(X\boldsymbol{Z}). \tag{3.3.25}$$

Evaluating equation (3.3.25) yields

$$\gamma = \boldsymbol{c}(||\boldsymbol{c}||^2 + \sigma_x^2)^{-1}, \tag{3.3.26}$$

and so $\gamma^T X$ is the best linear approximation to the regression of X on Z.

Structuring our simulation in this fashion introduces both nonlinearity and non-normality into the regression of X on $\boldsymbol{Z}$. Denote the conditional mean and variance of X given $\boldsymbol{Z}$ by

$$E(X|\ \boldsymbol{Z}) = \mu_x(\boldsymbol{Z},\ \gamma) \quad \text{and} \quad \text{Var}(X|\ \boldsymbol{Z}) = \sigma_x^2(\boldsymbol{Z},\ \gamma,\ \theta), \tag{3.3.27}$$

and note that here, the conditional variance of X given $\boldsymbol{Z}$ is not a function of $\mu_x(\boldsymbol{Z},\ \gamma)$. The true-data regression model is

$$X_j = \mu_x(\boldsymbol{Z}_j,\ \gamma) + \epsilon_j \sigma_x(\boldsymbol{Z}_j,\ \gamma,\ \theta), \qquad j = 1, \ldots, n, \tag{3.3.28}$$

and $X_1, \ldots, X_n$ is a random sample from the standardized Chi-square(9) density. The mean and variance functions in equation (3.3.28) are nonlinear in γ. Moreover, the regression residuals, $\epsilon_1, \ldots, \epsilon_n$ are both non-normal and are nonlinear in γ. The exact form and extent of the nonlinearity and non-normality in the true-data model are difficult to understand.

An observed-data regression model follows by adding N(0, 1) measurement errors to equation (3.3.28),

$$W_j = \mu_x(\boldsymbol{Z}_j,\ \gamma) + \epsilon_j \sigma_x(\boldsymbol{Z}_j,\ \gamma,\ \theta) + \sigma_u U_j, \qquad j = 1, \ldots, n. \tag{3.3.29}$$

As in Simulations 1 and 2, a variety of factors were controlled to understand

Table 3.5: Factors and levels included in Simulation 3.

Factor	**Levels**
R^2	0.1 0.3 0.5 0.7 0.9
ξ	0.7 0.9
Estimator	$\widehat{f}_x(x)$ $\widehat{f}_{und}(x)$ $\widehat{f}_{ovr}(x)$ $\widehat{f}_{naive}(x)$ $\widehat{f}_{true}(x)$
Coefficient Pattern	$[1,1,1,1]^T$ $[1,1,1/2,1/2]^T$ $[1,3/4,1/2,1/4]^T$

the performance of the regression-assisted deconvolution estimator to several key modeling components. These are described next and summarized in Table 3.3.3.

The coefficient of determination R^2 for the regression of X on $\boldsymbol{Z}$ affects the precision of the predicted values from the regression of W on $\boldsymbol{Z}$. A definition for R^2 follows by noting that from equation (3.3.24), the residual variance for the regression of X on $\boldsymbol{Z}$ can be expressed as

$$\tau^2 = E\{(X - \gamma^T \boldsymbol{Z})^2\}. \tag{3.3.30}$$

Because $\text{Var}(X) = 1$, $R^2 = 1 - \tau^2$ describes the proportion of variation in X explained by regression on $\boldsymbol{Z}$. Substituting the expression for γ in equation (3.3.26) into equation (3.3.30) and simplifying gives

$$R^2 = \frac{||\boldsymbol{c}||^2}{||\boldsymbol{c}||^2 + \sigma_x^2}. \tag{3.3.31}$$

R^2 was varied at the five levels 0.1, 0.3, 0.5, 0.7, and 0.9. Note that as σ_x^2 increases, R^2 decreases, implying that the non-normality of the model errors will be most apparent for small values of R^2.

The reliability ratio was also included as a factor in this simulation and was varied at the two levels, 0.7 and 0.9.

The effects of model misspecification were investigated in several ways. The functions $\mu_x()$ and $\sigma_x^2()$ in equation (3.3.29) are unknown and cannot be modeled directly. Using a multiple linear regression model with constant variance to approximate the true-data model in equation (3.3.28) leads to the approximate observed-data model,

$$W_j \approx \beta^T \boldsymbol{Z}_j + \sigma_x \epsilon_j + \sigma_u U_j \tag{3.3.32}$$

for $j = 1, \ldots, n$. Estimates from this model were used to compute the regression-assisted deconvolution estimator. Thus, these simulations indicate the estimator's sensitivity to overlooking nonlinearity in the mean and variance functions. In addition, because the true variance of X given $\boldsymbol{Z}$ in equation (3.3.27) is not a function of the mean, this approximation further misspecifies the variance function by modeling it as though it were.

The effects of over-fitting and under-fitting the linear regression model in equation (3.3.32) were investigated as well. We set $p = 4$, so that $\boldsymbol{Z}$ in equation (3.3.24) was a 4×1 vector of covariates. Eight additional covariates were generated as independent N(0, 1) random variables, and were uncorrelated with X. These were used to compute the under-fit model with $p_u = 2$ covariates and the over-fit model with $p_o = 8$ covariates. Covariates with the largest absolute t-statistics from the linear regression of W on all twelve covariates were selected for the under-fit and over-fit models. Predicted values and estimated variances from the correctly-specified, under-fit, and

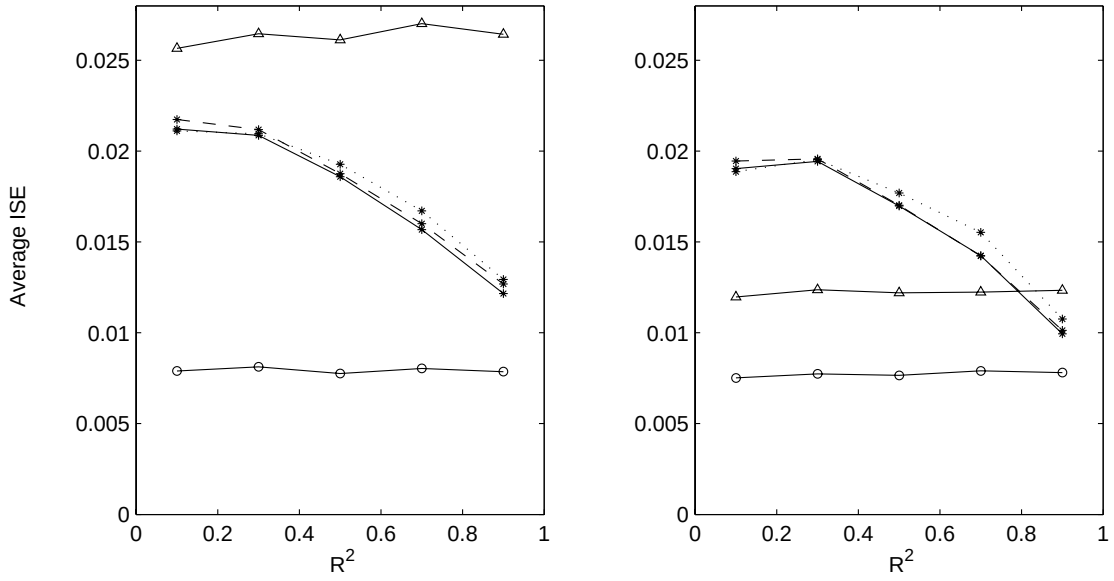

Figure 3.6: Average ISE by R^2 for Simulation 3. Left: reliability ratio=0.7. Right: reliability ratio=0.9. Circle: $\widehat{f}_{true}$; Triangle: $\widehat{f}_{naive}$; Star, solid line: $\widehat{f}_x$ for correct model fit; Star, dashed line: $\widehat{f}_x$ for over-fit model; Star, dotted line: $\widehat{f}_x$ for under-fit model. Pooled standard error of mean ISE is 0.00030.

over-fit models were used to compute the estimators $\widehat{f}_x(x)$, $\widehat{f}_{und}(x)$, and $\widehat{f}_{ovr}(x)$.

A coefficient pattern factor was also included in this simulation. From equation (3.3.26), it is seen that the components of the vector $\boldsymbol{c}$ control the relative strengths of the correlations between X and the $p = 4$ true covariates. Patterns for the magnitudes of the correlations between X and the covariates in $\boldsymbol{Z}$ were determined by specifying $\boldsymbol{c}$. Three patterns were considered for these correlations, $[1, 1, 1, 1]^T$, $[1, 1, 1/2, 1/2]^T$, and $[1, 3/4, 1/2, 1/4]^T$.

Because non-normality is induced in the model errors, $\epsilon_1, \ldots, \epsilon_n$, through the structure of this simulation, we did not consider the density $f_\epsilon(\epsilon)$ as a factor.

A total of 200 data sets with $n = 100$ observations were generated for each combination of the factor levels. Data sets for the simulation were generated as follows.

First, true data $X_1, \ldots, X_n$ were generated as standardized Chi-square(9) random variables. Next, given levels of R^2 and the coefficient pattern factor, the residual variance σ_x^2 was determined from equation (3.3.31). Model errors, $\epsilon_1, \ldots, \epsilon_n$ were generated as independent N(0, 1) random variables and were scaled to have variance σ_x^2. The covariate vectors, $\boldsymbol{Z}_1, \ldots, \boldsymbol{Z}_n$, were then constructed with equation (3.3.24). The sample of observed data was generated next. Given a level of ξ, normal measurement errors $U_1, \ldots, U_n$ were generated and scaled to have variance $\sigma_u^2 = 1 - \xi^{-1}$. These were added to the true data to produce the sample of observed data, $W_1, \ldots, W_n$.

An analysis of variance was computed with the simulation factors and showed that the main-effect and all interactions of the factor Coefficient Pattern explained less than 1% of the variation in the integrated squared error. Average integrated squared errors are plotted in Figure (3.6), with results pooled over the factor Coefficient Pattern. It is evident that the regression-assisted deconvolution estimator performs poorly for small to moderate values of R^2 when the reliability ratio is large. In these cases, the residuals from the regression of W on $\boldsymbol{Z}$ are most affected by the non-normal model errors. Discouragingly, the D'Agostino-Pearson K^2 test was only marginally effective at detecting non-normality in these cases. The power of this test to detect non-normality in the regression residuals from the correct model fit is summarized in Table 3.3.3. The results of this simulation emphasize that if our estimator is to be used, careful thought must be given to the correct model for the conditional mean and variance of X given $\boldsymbol{Z}$.

Table 3.6: Power of the D'Agostino-Pearson K^2 test in Simulation 3 to detect non-normality in observed-data regression residuals from the correct model fit.

R^2	$\xi = 0.7$	$\xi = 0.9$
0.1	0.30	0.54
0.3	0.14	0.27
0.5	0.08	0.14
0.7	0.06	0.07
0.9	0.07	0.05

3.4 An Application

In this section, we illustrate the regression-assisted deconvolution estimator with a real-data example. Data are measurements of carbon monoxide (CO) in automobile emissions, collected with a U.S. EPA remote sensing device. This device is stationed along a roadway and measures the levels of various compounds in the exhaust of passing automobiles. Researchers at the U.S. EPA use density estimates of the levels of various pollutants in automobile exhaust to gain an understanding of the characteristics of the subpopulation of automobiles that are heavy polluters and may be out of compliance with regulations. However, remotely measured concentrations of compounds such as CO are subject to error from several sources including environmental conditions, the automobile's acceleration and engine temperature, and instrument error.

Measurements of CO were recorded over several days at a single location in North

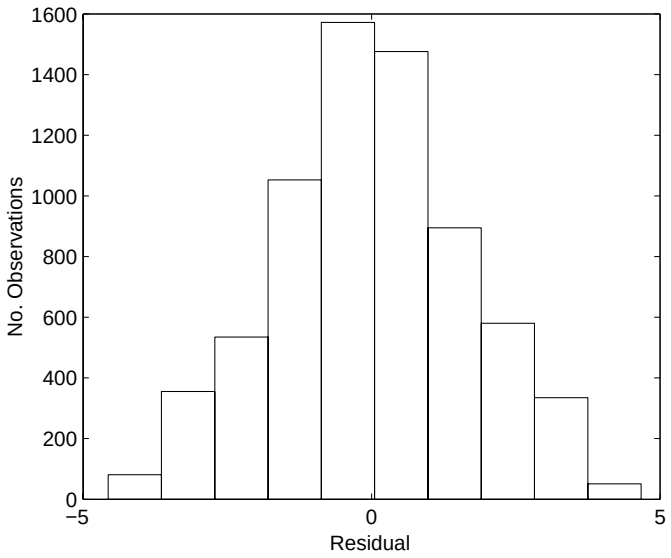

Figure 3.7: Histogram of regression residuals from the linear regression model fit to the observed log-transformed CO measurements.

Carolina. In addition to the CO measurements, the license plate of each passing car was photographed, allowing measurements to be identified to individual automobiles. Covariates describing various characteristics of the automobiles were obtained through the license plate information. These included the date of manufacture, as well as engine features that affect emissions. Because of the multiple sources of error in the measurements, we assumed that the measurement error structure of the original data was multiplicative. We log-transformed data to achieve an additive error structure, and estimated the density function of the log-transformed CO levels. For comparison, we present results from both the regression-assisted deconvolution estimator and the naive estimator that ignores measurement error.

Data represent measurements from a total of 8576 automobiles. Multiple measurements were taken of automobiles that passed the study location more than once,

and these were used to estimate the measurement error variance. Of the 8576 automobiles in the data set, 6933 were measured once, 1478 were measured twice, 165 were measured three or more times. Using the replicate measurements from the $m = 1478$ twice-observed automobiles, we estimated the measurement error variance by

$$\widehat{\sigma}_u^2 = \frac{1}{2m} \sum_{j=1}^{m} (W_1 - W_2)^2,$$

which resulted in the estimate $\widehat{\sigma}_u^2 = 2.1420$.

Both the regression-assisted deconvolution estimator in equation (3.2.8) and the naive estimator in equation (3.3.10) were computed using data from the 6933 automobiles that were observed only once. With the estimate of measurement error variance, the reliability ratio for these data was estimated to be 0.65.

We used multiple linear regression to model the log-transformed CO measurements with the available covariates, and our model resulted in a coefficient of determination of only $R^2 = 0.16$. The hypothesis that the regression residuals from this model were normally distributed was rejected by the D'Agostino-Pearson K^2 test ($p < 0.005$). This is not too surprising, however, considering the large number of observations in the data set. A histogram of the regression residuals is shown in Figure 3.7, and shows that the residuals appear close to normally distributed.

The estimated variance from the observed-data regression was MSE=2.819, so that the model error variance from the true-data regression was estimated as $\widehat{\sigma}_x^2 = 2.819 - 2.142 = 0.677$. This estimated variance and the predicted values from the observed-data regression were used to compute the regression-assisted deconvolution

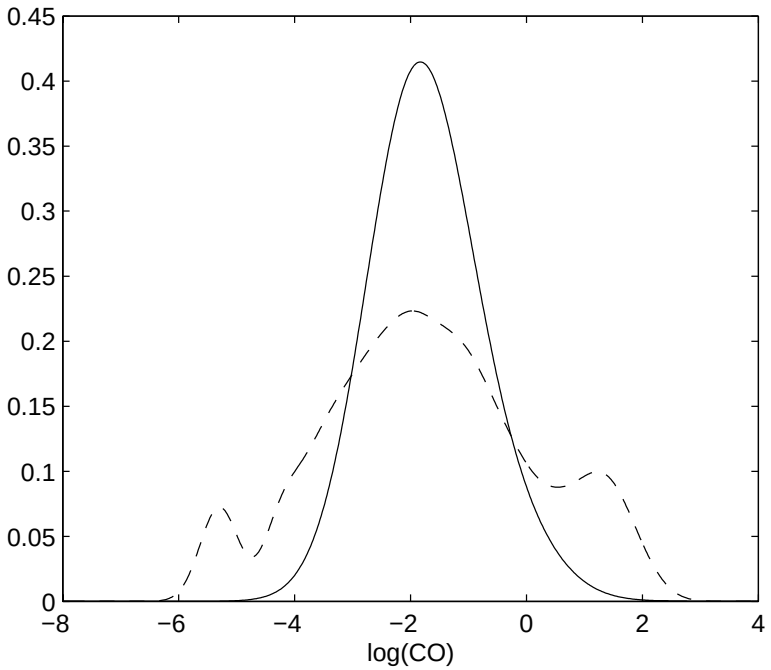

Figure 3.8: Estimates of the density function of log-transformed CO levels in automobile exhaust. Solid line: Regression-assisted deconvolution estimate; Dashed line: Naive estimate.

estimator in equation (3.2.8), with the bandwidth λ set to 0. The naive estimator in equation (3.3.10) was computed directly from the observed measurements, with the plug-in bandwidth $\lambda = 0.329$ estimated with equation (3.3.11).

Density estimates are presented in Figure 3.8. Measurement error, when uncorrected, generally results in a density estimate that is overly flat and smooth. The regression-assisted deconvolution estimate in Figure 3.8 is distinctly higher-peaked and thinner-tailed than the naive estimate, as is typical when the effects of measurement error are corrected. It is clear from the figure that tail probabilities estimated from the regression-assisted deconvolution estimate will be much smaller than those estimated from the naive kernel density estimate. However, the presence of distinct

modes in the naive estimate suggests that our estimator is over-compensating for the effects of measurement error. The three modes might indicate distinct populations in the data. More complicated regression models could be explored and might better address these characteristics, and perhaps increase the model R^2.

3.5 Summary

This chapter introduced a regression-assisted deconvolution estimator for the density function of an error-prone variable X. The estimator assumes that the conditional mean and variance of X given $\boldsymbol{Z}$ can be estimated by the regression of the observed data on $\boldsymbol{Z}$. The estimator also assumes that both regression errors and measurement errors are normally distributed. It was argued in Section 2 that when these assumptions are met, the rate of convergence of the estimator is determined by the minimum rate of convergence of the estimated mean and variance functions.

Results of a simulation study suggested that when the assumptions of the regression model and the normality of regression errors are met, the density of X is well estimated by the regression-assisted deconvolution estimator. In many cases, it is competitive with the true-data kernel density estimator in terms of integrated squared error. The performance of the estimator was seen to depend on the strength of the relationship between X and $\boldsymbol{Z}$ (R^2) and the relative variances of W and X (ξ). Correctly specifying the form of the mean and variance functions (linear, nonlinear) was seen to be critical to the estimator's performance. However model misspecification in

terms of under-fitting and over-fitting the regression model generally was less important. The assumption of normal regression errors was also seen to be very important. Even when not met, however, in many of the cases considered the regression-assisted deconvolution estimator continued to outperform the naive kernel density estimator in terms of integrated squared error.

Our results, both theoretical and empirical, indicate that the regression-assisted deconvolution estimator is a potentially powerful tool for deconvolution. Further research is needed, however, to fully appreciate the estimator's potential.

First, a more rigorous examination of the estimator's asymptotic properties is needed to confirm the heuristic arguments presented in this chapter. Besides lending further credibility to the method, this would provide a framework in which further modifications to the estimator could be easily examined.

Second, our simulation results suggested that the estimator is very sensitive to certain types of model misspecifications. However, our simulations considered only fully parametric regression models. Modeling approaches that make fewer parametric assumptions should be less sensitive to incorrect model specification. The performance of the estimator when models are fit using semi-parametric and nonparametric methods should be investigated. A comparison among methods that rely on a varying degree of parametric assumptions would provide important guidelines for the estimator's use in practice. Understanding the possible drawbacks and gains from imposing increasingly restrictive assumptions would help users decide on the best modeling approach for a given situation.

Finally, to a large extent, the assumption of normal model errors drives the success of the regression-assisted deconvolution estimator, and increases its rate of convergence relative to nonparametric deconvolution estimators. However, it was seen that the estimator can perform poorly when the normality assumption is not met. An alternative approach is to estimate the density of the model errors using nonparametric deconvolution on the observed-data regression residuals. This estimated density, in place of the standard normal density, would then be recombined with the kernel estimate of the predicted values. Additional research is needed to determine whether this approach would offer any advantages over traditional nonparametric deconvolution estimators.

Bibliography

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.

Carroll, R. J. & Ruppert, D. (1988). *Transformation and weighting in regression*. Chapman & Hall Ltd.

Carroll, R. J., Ruppert, D. & Stefanski, L. A. (1995). *Measurement error in nonlinear models*. Chapman & Hall Ltd.

D'Agostino, R. B., Belanger, A. & D'Agostino, Ralph B., J. (1990). A suggestion for using powerful and informative tests of normality. *The American Statistician* **44**, 316–321.

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, Methodological* **55**, 523–531.

Erdelyi, A. e. (1953). *Higher Transcendental Functions Volume 2*. 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. *The Annals of Statistics* **19**, 1257–1272.

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

Faraway, J. J. & Jhun, M. (1990). Bootstrap choice of bandwidth for density estimation. *Journal of the American Statistical Association* **85**, 1119–1122.

Fuller, W. A. (1987). *Measurement error models*. John Wiley & Sons.

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

Parzen, E. (1962). On estimation of a probability density function and mode. *The Annals of Mathematical Statistics* **33**, 1065–1076.

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*. Chapman & Hall Ltd.

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

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

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., Novick, S. J. & Devanarayan, V. (2003). Monte carlo estimation of $\{\sin(\mu)/\mu\}^{2k}$ and other similar estimands with applications ... really!. *North Carolina State University Technical Report*.

Taylor, C. C. (1989). Bootstrap choice of the smoothing parameter in kernel density estimation. *Biometrika* **76**, 705–712.

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

Watson, G. S. (1983). *Statistics on spheres*. John Wiley & Sons.