DECONVOLUTION IN RANDOM EFFECTS MODELS VIA NORMAL MIXTURES

A Dissertation

by

NATHANIEL A. LITTON

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

August 2009

Major Subject: Statistics

DECONVOLUTION IN RANDOM EFFECTS MODELS VIA NORMAL MIXTURES

A Dissertation

by

NATHANIEL A. LITTON

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Approved by:

Chair of Committee,	Jeffrey D. Hart
Committee Members,	Richard DeBlassie
	Michael Sherman
	Clifford Spiegelman
Head of Department,	Simon Sheather

August 2009

Major Subject: Statistics

ABSTRACT

Deconvolution in Random Effects Models via

Normal Mixtures. (August 2009)

Nathaniel A. Litton, B.S., Cornell University;

M.S., Texas A&M University

Chair of Advisory Committee: Dr. Jeffrey D. Hart

This dissertation describes a minimum distance method for density estimation when the variable of interest is not directly observed. It is assumed that the underlying target density can be well approximated by a mixture of normals. The method compares a density estimate of observable data with a density of the observable data induced from assuming the target density can be written as a mixture of normals. The goal is to choose the parameters in the normal mixture that minimize the distance between the density estimate of the observable data and the induced density from the model. The method is applied to the deconvolution problem to estimate the density of X_i when the variable $Y_i = X_i + Z_i$, $i = 1, \ldots, n$, is observed, and the density of Z_i is known. Additionally, it is applied to a location random effects model to estimate the density of Z_{ij} when the observable quantities are p data sets of size n given by $X_{ij} = \alpha_i + \gamma Z_{ij},\ i = 1, \ldots, p,\ j = 1, \ldots, n$, where the densities of α_i and Z_{ij} are both unknown.

The performance of the minimum distance approach in the measurement error model is compared with the deconvoluting kernel density estimator of Stefanski and Carroll (1990). In the location random effects model, the minimum distance estimator is compared with the explicit characteristic function inversion method from Hall and Yao (2003). In both models, the methods are compared using simulated and real data sets. In the simulations, performance is evaluated using an integrated squared error

criterion. Results indicate that the minimum distance methodology is comparable to the deconvoluting kernel density estimator and outperforms the explicit characteristic function inversion method.

To Dana and Mom

ACKNOWLEDGMENTS

When reflecting upon my time in graduate school, I find that I could not have made it through without the kind support of many people along my journey. I am fortunate to have had the opportunity to surround myself with so many influential people.

First, I am greatly indebted to my advisor, Dr. Jeffrey D. Hart, for his encouragement, keen insight, and invaluable guidance throughout my studies. I have been truly fortunate to have had such a great advisor and mentor. I would like to express special thanks to Dr. Cliff Spiegelman for retaining me as a research assistant for several semesters and for serving as a committee member. I would also like to thank Dr. Michael Sherman and Dr. Dante DeBlassie for their time and suggestions while serving as members of my committee.

My most appreciative and loving thanks go to my parents and brother. Their dedicated support and steady encouragement throughout the years have been beyond measure. Were it not for my mother's unconditional love and sacrifice, I would not be where I am today. Last, but most significantly, I would like to express the warmest thanks to my fiancée and soon-to-be wife, Dana. Without her encouragement, patience, and love, this dissertation would not have been written.

TABLE OF CONTENTS

Page

ABSTRACT iii

DEDICATION v

ACKNOWLEDGMENTS vi

TABLE OF CONTENTS vii

LIST OF TABLES ix

LIST OF FIGURES x

CHAPTER

I INTRODUCTION 1

1.1 Deconvolution in Measurement Error Model 2
1.1.1 Density Function Estimation 2
1.1.2 Bandwidth Selection 5
1.1.3 Distribution Function Estimation 6
1.1.4 Optimality Results 7
1.2 Deconvolution in Location Random Effects Model 8
1.2.1 Density Function Estimation 10
1.2.2 Distribution Function Estimation 11
1.3 Other Related Results 13

II DECONVOLUTION IN MEASUREMENT ERROR MODEL . 14

2.1 Minimum Distance Method 14
2.2 Approximating Minimum Distance Estimate Using Normal Mixtures 15
2.2.1 Metric 15
2.2.2 Discretizing the Metric 17
2.2.3 Estimating Mixing Proportions $\mathbf{p}$ 18
2.2.4 Estimating σ 18

CHAPTER Page

2.2.5 Estimating Number of Components L 19
2.2.6 Outline for Showing Consistency of Estimator . . . 22
2.3 Simulated Examples 22
2.4 Applications 37

III DECONVOLUTION IN LOCATION RANDOM EFFECTS MODEL 39

3.1 Minimum Distance Method 39
3.2 Identifiability of Error Density 41
3.2.1 Existing Identifiability Results 41
3.2.2 Identifiability from Residuals 42
3.3 Approximating Minimum Distance Estimate Using Normal Mixtures 48
3.3.1 Minimum Distance Using Univariate Distribution of ϵ_0 49
3.3.2 Minimum Distance Using Bivariate Distribution of (ϵ_1, ϵ_2) 50
3.3.3 Estimating Mixing Proportions 51
3.3.4 Estimating σ and Number of Components L 52
3.3.5 Outline for Showing Consistency of Estimator . . . 54
3.4 Simulated Examples 54
3.4.1 Hall and Yao's Explicit Characteristic Function Inversion Method 56
3.4.2 Minimum Distance Method Using a Normal Mixture 62
3.5 Extension to Location-Scale Random Effects Model 70
3.6 Applications 72

IV CONCLUSION 79

4.1 Summary 79
4.2 Further Research 80

REFERENCES 82

APPENDIX A 85

APPENDIX B 95

VITA 98

LIST OF TABLES

TABLE		Page
1	Median ISE of density estimates under normal measurement error using minimum distance estimation. The numbers in parentheses are interquartile ranges of ISE for 100 replications.	25
2	Median ISE of density estimates reported by Delaigle and Gijbels under normal measurement error using the deconvoluting kernel density estimator. The numbers in parentheses are interquartile ranges of ISE for 500 replications.	26
3	Median ISE of density estimates under Laplace measurement error using minimum distance estimation. The numbers in parentheses are interquartile ranges of ISE for 100 replications.	31
4	Median ISE of density estimates reported by Delaigle and Gijbels under Laplace measurement error using the deconvoluting kernel density estimator. The numbers in parentheses are interquartile ranges of ISE for 500 replications.	32
5	Median ISE of density estimates using Hall and Yao's method. The numbers in parentheses are interquartile ranges of ISE for 100 replications.	57
6	Median ISE of minimum distance estimates for error densities using univariate distribution of residuals. The numbers in parentheses are interquartile ranges of ISE for 100 replications.	63
7	Median ISE of minimum distance estimates for error densities using bivariate distribution of residuals. The numbers in parentheses are interquartile ranges of ISE for 100 replications.	64

LIST OF FIGURES

FIGURE Page

1 Example of how the distance metric typically changes with the number of normal mixture components. 21

2 Target densities of increasing difficulty to estimate; density 1 (top left), density 2 (top right), density 3 (bottom left), density 4 (bottom right). 24

3 Minimum distance estimates for $N(0,1)$ density (density 1) under normal measurement error at sample size n=50 (top left), n=100 (top right), n=250 (bottom left), n=500 (bottom right). 27

4 Minimum distance estimates for $\chi^2(3)$ density (density 2) under normal measurement error at sample size n=50 (top left), n=100 (top right), n=250 (bottom left), n=500 (bottom right). 28

5 Minimum distance estimates for normal mixture density (density 3) under normal measurement error at sample size n=50 (top left), n=100 (top right), n=250 (bottom left), n=500 (bottom right). . . . 29

6 Minimum distance estimates for gamma mixture density (density 4) under normal measurement error at sample size n=50 (top left), n=100 (top right), n=250 (bottom left), n=500 (bottom right). . . . 30

7 Minimum distance estimates for $N(0,1)$ density (density 1) under Laplace measurement error at sample size n=50 (top left), n=100 (top right), n=250 (bottom left), n=500 (bottom right). 33

8 Minimum distance estimates for $\chi^2(3)$ density (density 2) under Laplace measurement error at sample size n=50 (top left), n=100 (top right), n=250 (bottom left), n=500 (bottom right). 34

9 Minimum distance estimates for normal mixture density (density 3) under Laplace measurement error at sample size n=50 (top left), n=100 (top right), n=250 (bottom left), n=500 (bottom right). 35

FIGURE Page

10 Minimum distance estimates for gamma mixture density (density 4) under Laplace measurement error at sample size n=50 (top left), n=100 (top right), n=250 (bottom left), n=500 (bottom right). 36

11 Minimum distance estimates for log daily saturated fat assuming normal (left) and Laplace error densities (right) when $\sigma_Z^2 = 0$, $1/5\sigma_X^2$, $1/3\sigma_X^2$, and $1.5\sigma_X^2$. 37

12 Residual densities for different values of n when error density is $N(0,1)$ (solid), Laplace$(0, \sqrt{1/2})$ (dotted-dash) and Unif$(-\sqrt{3}, \sqrt{3})$ (dashed). 44

13 Residual densities for n = 3, 4, 5, and 8 compared with their respective underlying error density. 45

14 Residual densities for n = 10, 15, 20, and 25 compared with their respective underlying error density. 46

15 Residual densities for n = 30, 50, 100, and 250 compared with their respective underlying error density. 47

16 Target error densities of increasing difficulty to estimate; density 1 (top left), density 2 (top right), density 3 (bottom left), density 4 (bottom right). 55

17 Density estimates for $N(0,1)$ error density (density 1) using Hall and Yao's method at sample size p=100 (top left), p=250 (top right), p=500 (bottom left), p=1000 (bottom right). 58

18 Density estimates for $\chi^2(3)$ error density (density 2) using Hall and Yao's method at sample size p=100 (top left), p=250 (top right), p=500 (bottom left), p=1000 (bottom right). 59

19 Density estimates for exp(1) error density (density 3) using Hall and Yao's method at sample size p=100 (top left), p=250 (top right), p=500 (bottom left), p=1000 (bottom right). 60

20 Density estimates for normal mixture error density (density 4) using Hall and Yao's method at sample size p=100 (top left), p=250 (top right), p=500 (bottom left), p=1000 (bottom right). . . . 61

FIGURE Page

21 Minimum distance estimates for $N(0,1)$ error density (density 1) at sample size p=100 (top left), p=250 (top right), p=500 (bottom left), p=1000 (bottom right). 65

22 Minimum distance estimates for $\chi^2(3)$ error density (density 2) at sample size p=100 (top left), p=250 (top right), p=500 (bottom left), p=1000 (bottom right). 66

23 Minimum distance estimates for exp(1) error density (density 3) at sample size p=100 (top left), p=250 (top right), p=500 (bottom left), p=1000 (bottom right). 67

24 Minimum distance estimates for normal mixture error density (density 4) at sample size p=100 (top left), p=250 (top right), p=500 (bottom left), p=1000 (bottom right). 68

25 Kernel density estimates for residual density under LRE model (left) and LSRE model (right). 73

26 Minimum distance estimates of error density under LRE model when the normal mixture has L=1, 3, 5, 7, 9, and 11 components. . . 74

27 Minimum distance estimates of error density under LSRE model when the normal mixture has L=1, 3, 5, 7, 9, and 11 components. . . 77

28 Density estimates for error density under LRE model (left) and LSRE model (right). 78

CHAPTER I

INTRODUCTION

There is extensive literature devoted to density and distribution function estimation in the case where the variable of interest is not directly observed. While there are a number of popular problems encountered in the literature, density estimation for two classes of problems is of primary interest in this dissertation: the measurement error model and a location random effects model, both of which are defined later.

In this dissertation, estimation is performed using a minimum distance methodology. An early result on minimum distance dates back to Wolfowitz (1957), who develops a minimum distance method for finding strongly consistent estimators. He considers a stochastic structure where the distribution function of an observed process depends on a distribution function of random variables and a vector of unknown parameters. Then, the minimum distance method provides a way, at least in principle, to obtain strongly consistent estimators of the unknown parameters and distribution function of random variables. This holds provided that the system is identifiable, and a continuity assumption is satisfied. While the methodology proposed by Wolfowitz applies to estimating distribution functions, it motivates a minimum distance procedure that can be used for density estimation to estimate the target density f_X in the measurement error model and the error density f_Z in the location random effects model.

The journal model is *Journal of the American Statistical Association.*

1.1 Deconvolution in Measurement Error Model

The first problem is the classic deconvolution problem, which arises when considering a basic measurement error model where knowledge of the underlying target density is desired. For this model, consider a random variable $X_i,\ i = 1, \ldots, p$, with an unknown probability density f_X and a random variable $Z_i,\ i = 1, \ldots, p$, with a known probability density f_Z. The random variable of interest is X_i, the random noise variable is Z_i, and the observable quantity is the contaminated data

$$Y_i = X_i + Z_i, \quad i = 1, \ldots, n \tag{1.1}$$

having density f_Y. Without loss of generality, the random noise variable Z_i is assumed to have mean 0. The goal is to deduce the probability density of X_i based on the contaminated sample observations and the known error distribution.

The deconvolution problem has been studied in a variety of contexts. Many of the methods rely on a Fourier inversion method, which can be problematic when the characteristic function corresponding to f_Z, denoted by ψ_Z, is near 0. This happens because the estimator of f_X relies on the ratio of the Fourier transforms of f_Y and f_Z, which is related to the well known result that deconvolution is more difficult for smoother error densities. Specific contributions to the deconvolution literature are discussed in the following sections.

1.1.1 Density Function Estimation

Perhaps the most widely studied kernel-type deconvolution method for density estimation is the deconvoluting kernel density estimator proposed by Stefanski and

Carroll (1990). The estimator is defined as

$$\hat{f}_X(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \exp\{-itx\}\frac{\psi_K(\lambda t)\hat{\psi}(t)}{\psi_Z(t)}dt,$$

where K is a kernel, and λ is the smoothing parameter. This estimator is natural in that it is obtained by explicitly solving for the characteristic function of the unknown density and performing Fourier inversion to obtain an estimate of the target density. Here it is assumed that the characteristic function of Z is nonvanishing everywhere. The estimator is shown to be consistent for a continuous and bounded target density under mild conditions. For normal errors, a cross-validation approximation to the integrated squared error (ISE) is derived so that the bandwidth can be computed as the zero of this equation. Lastly, the authors show that the best possible rates of convergence for $\hat{f}_X$ in terms of the integrated mean squared error (IMSE) are $(\log n)^{-2}$, $(\log n)^{-1}$, and $n^{-4/9}$ when the error densities are normal, Cauchy, and Laplace, respectively.

Liu and Taylor (1989) propose a kernel-type estimator that is similar to the deconvoluting kernel density estimator of Stefanski and Carroll (1990). For a kernel K, the estimator is

$$\hat{f}_{M_n}(x) = \frac{1}{2n\pi}\sum_{j=1}^{n} \mathrm{Re}\int_{-M_n}^{M_n} \exp\{it(Y_j - x)\}\frac{\psi_K(th_n)}{\psi_Z(t)}dt$$

for sequences M_n and h_n. Using several assumptions similar to Stefanski and Carroll such as a nonvanishing characteristic function of Z, strong uniform consistency of the estimator is shown. The authors are able to derive exact expressions for the bias and variance of the estimator under the assumption of a symmetric error density and kernel function. However, since the general forms of the exact bias and variance are complicated, an upper bound for the mean squared error (MSE) is derived. As a practical choice of M_n, the value that minimizes the upper bound on the MSE is

recommended.

Devroye (1989) constructs an estimator of the target density in the deconvolution problem satisfying

$$\lim_{n\to\infty} E \int |f_n - f| = 0.$$

The proposed estimator is defined as

$$f_n(x) = \begin{cases} 0 & \text{if } |x| \geq T \\ \frac{1}{2\pi} \operatorname{Re} \int_{R-A_r} \exp\{-itx\} \frac{\psi_K(th)\hat{\psi}(t)}{\psi_Z(t)} dt & \text{if } |x| < T, \end{cases}$$

where T is a tail parameter, h is a smoothing parameter, and r is a noise-control parameter. Other than the truncation at T and integration over $R - A_r$, the estimator is similar to those of Stefanski and Carroll (1990) and Liu and Taylor (1989).

While the previous papers construct estimators for the target density, Carroll and Hall (2004) propose both a kernel and orthogonal series low order approximation to the density of interest. The motivation for this is due to the extremely slow optimal rates of convergence for an estimator of f_X. They note that because of this, the density f_X is not estimable for practical purposes. Thus, the proposed estimator does not estimate f_X, but rather an approximation to f_X that only depends on easily computed averages. The regression setting is also considered since these same methods apply to constructing accurate low order approximations to the underlying function in nonparametric regression with errors in the explanatory variables. Numerical properties are investigated for both approximations in the regression and density estimation settings.

In the classic deconvolution problem, the distribution of X is assumed to be known so as to insure identifiability of the target density. However, Diggle and Hall (1993) assume that in addition to observing the data contaminated by error, another set of data from the error distribution is also observed. That is, independent random

samples are observed from densities f_X and f_Z. The authors introduce a damping factor d to define their nonparametric estimator for f_X as

$$\hat{f}_X(x) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \exp\{-itx\} d(t) \frac{\hat{\psi}_Y(t)}{\hat{\psi}_Z(t)} dt,$$

where $\hat{\psi}_Y$ and $\hat{\psi}_Z$ denote the empirical characteristic functions. The asymptotic MISE of the estimator is derived for the special case when the damping factor is defined as $d(t) = I_{(-p,p)}(t)$. This is then used as the basis for an ad hoc data-driven approach to select a value of p for the truncated estimator.

1.1.2 Bandwidth Selection

Several bandwidth selection methods are commonly used for the deconvoluting kernel density estimator. Hesse (1999) proposes a bandwidth selection method for the deconvoluting kernel density estimator when the error distribution is ordinary smooth. The bandwidth is selected to minimize an estimate of the ISE. As a practical consideration, the continuous optimization problem of solving for the bandwidth is replaced by a version discretized over an appropriate interval. The deconvoluting kernel density estimator using the proposed bandwidth is shown to be asymptotically optimal for both ISE and MISE.

Delaigle and Gijbels (2004) perform a simulation study comparing several bandwidth selection methods when using the deconvoluting kernel density estimator of Stefanski and Carroll (1990). An expression for the asymptotic dominating term of the mean integrated squared error (AMISE) forms the basis for several bandwidth selectors. The two plug-in type estimators for the bandwidth that are investigated select the bandwidth by minimizing the expression for AMISE(h) with respect to h. However, the expression for AMISE(h) depends on the unknown quantity $R(f_X'') = \int (f_X''(x))^2 dx$, and so it must be estimated. The first method for doing this

is to assume that f_X is a normal density. A problem with this is that the estimator for $R(f_X'')$ is not consistent, and thus the selected bandwidth will not be consistent for the bandwidth h_{MISE} that minimizes MISE. The other plug-in type method relies on computing the rth derivative of the deconvoluting kernel density estimator. In the simulation study, these two methods are compared with a cross-validation and bootstrap bandwidth selector. Results indicate that the plug-in bandwidth selector which makes use of the rth derivative of the deconvoluting kernel estimator and the bootstrap bandwidth selector outperform the cross validation method.

1.1.3 Distribution Function Estimation

In line with the mixture approach of this dissertation, Cordy and Thomas (1997) discuss distribution function estimation in the deconvolution problem when the target distribution can be represented as a mixture of known distributions. Instead of considering the case for a general error distribution, the primary interest is in a normal error with known variance σ^2/n_i. An EM algorithm is suggested for estimating the mixing proportions of the component distributions. While the authors suggest using normal mixture components with common variance for simplicity, they remark that the actual shape of the component distribution is relatively unimportant. Furthermore, there are two methods for constructing large sample confidence intervals for $F_X(x)$ at a fixed value of x. The first is based on a normal approximation to $\hat{F}_X(x)$, while the second approximates the distribution of the likelihood ratio statistic using a chi-square distribution.

Hall and Lahiri (2008) investigate the problem of distribution function estimation in addition to moment and quantile estimation in deconvolution. For example, a natural moment estimator of the target distribution is to base the estimator on empirical moments of the observed data and known theoretical moments of the error

distribution. Many convergence results are derived for the various estimators. It is shown that $\sqrt{n}$ consistency is possible when the error distribution is particularly "rough" in the sense that the tails of the error characteristic function converge to 0 very slowly. An analogous convergence result is shown for moment estimators. An interesting result that is unexplored by previous investigators is that the distribution function converges at an uneven rate. In other words, the rate of convergence for the distribution function of zero-centered data can be an order of magnitude slower at the origin than in the region $\{x : |x| > x_0\}$, for each fixed $x_0 > 0$.

1.1.4 Optimality Results

An early optimality result is proven by Carroll and Hall (1988), who investigate the optimal rates of convergence in the deconvolution problem for any nonparametric estimator of the unknown target density. The intrinsic difficulty in deconvolving a density is apparent by the very slow optimal rates of convergence. For example, it is shown that when the error distribution is normal, the fastest rate of convergence is $(\log n)^{-k/2}$, where the target density f_X has k bounded derivatives. The optimal rates of convergence when the errors are gamma with shape parameter $\alpha > 0$ and Laplace distributions are $n^{-k/(2k+2\alpha+1)}$ and $n^{-k/(2k+5)}$, respectively. The deconvoluting kernel density estimator of Stefanski and Carroll (1990) attains these optimal rates of convergence for the normal, gamma, and Laplace error densities. While the bounds on the rates of convergence are derived for the class of k-times differentiable densities having k bounded derivatives, there is the possibility that much faster rates can be attained by restricting attention to a smaller class of densities.

In addition, Fan (1991) derives optimal global rates of convergence under weighted L_p loss in the deconvolution problem for any nonparametric estimator of the target density. Unfortunately, this does not include the L_p norm since the weight function in

this case is $w(\cdot) = 1$, which is excluded because w must be integrable. For target densities f_X belonging to the class of densities with k bounded derivatives, the optimal global rate of convergence for supersmooth error densities is $O((\log n)^{-k/2})$. Optimal rates of convergence for estimating the rth derivative of the target density are also derived. Due to the nature of the extremely slow logarithmic rates of convergence in the case of supersmooth errors, Fan investigates how large an error variance is acceptable. It is shown that if the error variance is $O(n^{-1/(2k+1)})$, then in terms of rates of convergence, deconvolution is no more difficult than ordinary density estimation. Thus, provided the level of noise is small, deconvolution may not be as hopeless as the rates of convergence suggest.

A more recent result is provided by Butucea (2004), who considers the special case of the deconvoluting kernel density estimator that uses the well known sinc kernel. The estimator is shown to be optimal in the minimax sense when the unknown density belongs to a class of supersmooth functions, and the known error density is ordinary smooth. Convergence rates are established for the MSE and MISE of this estimator.

1.2 Deconvolution in Location Random Effects Model

The second model in which the variable of interest is not directly observed is the location random effects (LRE) model. For this model, let $Z_{ij},\ i = 1, \ldots, p,\ j = 1, \ldots, n,$ be independent and identically distributed (i.i.d.) random variables with common and unknown probability density f_Z having mean 0 and standard deviation 1. Furthermore, let $\alpha_i,\ i = 1, \ldots, p,$ be i.i.d. random variables with unknown density f_α. The observable quantities can be thought of as p data sets of size n given by

$$X_{ij} = \alpha_i + \gamma Z_{ij}, \quad i = 1, \ldots, p, \quad j = 1, \ldots, n. \tag{1.2}$$

The error random variables $Z_{ij},\ i = 1, \ldots, p,\ j = 1, \ldots, n$, are assumed to be independent of $\alpha_i,\ i = 1, \ldots, p$. The goal is to deduce the probability density of the error variable Z_{ij}.

The distribution of the error random variable Z_{ij} is important for a variety of reasons. For example, it can determine the distributions of test statistics in hypothesis testing. When the LRE model in (1.2) holds, the density of Z_{ij} can be consistently estimated in the usual way via the residuals

$$\hat{Z}_{ij} = X_{ij} - \bar{X}_i, \quad i = 1, \ldots, p, \quad j = 1, \ldots, n \tag{1.3}$$

if $n \to \infty$. This holds under mild moment conditions on α_i and Z_{ij}. Here, $\bar{X}_i$ is the mean of the ith data set. However, in some applications, the assumption that n is large may be unreasonable. In fact, in microarray experiments, p can be on the magnitude of thousands or tens of thousands, while n can be very small. In this case, as $p \to \infty$ with n fixed, the density of $\hat{Z}_{ij}$ is not f_Z, and hence the empirical distribution of the $\hat{Z}_{ij}$'s do not consistently estimate the distribution of Z_{ij}. A natural question then becomes "Is it possible to consistently estimate f_Z as $p \to \infty$ for a bounded n?" This is in part answered by Reiersøl (1950) for the LRE model. Reiersøl investigates identifiability in a relation of the form

$$X_i = Y_i + V_i \quad \text{and} \quad Y_2 = \beta_0 + \beta_1 Y_1, \quad i = 1, 2,$$

where X_i, Y_i, and V_i are random variables, and β_0 and β_1 are parameters. Conditions for identifiability of β_0 and β_1 and the distributions of Y_i and V_i are established in several cases. This is of particular importance since setting $\beta_0 = 0$ and $\beta_1 = 1$ reduces the relation to the LRE model. The results imply that the characteristic functions of Y_i and V_i are identifiable for a sample size as small as $n = 2$. Identifiability in this case holds as long as the characteristic functions have discrete zeros. Specific

contributions to estimation in the LRE model are discussed in the following sections.

1.2.1 Density Function Estimation

The earliest known result relating to estimation in the LRE model comes from Horowitz and Markatou (1996). They propose estimators for f_α and f_Z in a situation where the LRE model is a special case of a more general model for panel data. Their density estimators of the error and random effects are obtained by using Fourier inversion on an estimate of the respective characteristic functions. A drawback to this methodology is that it relies on the assumption that the error density f_Z is symmetric. The estimators are shown to be consistent, and rates of convergence are derived for special cases. In addition, they outline an idea for a procedure to possibly extend the methodology to a nonsymmetric error when $n \geq 3$.

Li and Vuong (1998) propose estimators for f_α and f_Z that relax the symmetric error assumption required by Horowitz and Markatou (1996). As with other similar methods, the estimator expresses the characteristic functions of α and Z in terms of the characteristic function of the observed data, and these expressions are then used to obtain estimates of the unknown densities using Fourier inversion. Conditions for identification of the two distributions are addressed. Rates of uniform convergence are derived for the four combinations of the two distributions being either ordinary smooth or supersmooth.

Susko and Nadon (2002) consider the problem of estimating the error density in the LRE model with a small number of repeated measurements. They propose two methods that make the limiting assumption that the error distribution is symmetric. The first estimator computes an empirical characteristic function of $X_{ir} - X_{is},\ i = 1, \ldots, p,\ r \neq s$ and uses Fourier inversion to obtain an estimate of the error density. The authors derive rates of convergence for the MISE in this case. The second

estimator assumes that a normal mixture provides an adequate approximation to the error density. While not fixed or known, the number of normal mixtures is chosen to be as large as possible while maintaining a reasonable computation time. The mixing proportions and the common variance of the normal mixture are estimated via the EM algorithm.

Furthermore, Delaigle et al. (2008) consider the problem of deconvolution with repeated measurements, which is equivalent to the LRE model. They propose a modified kernel estimator for the random effects density

$$\hat{f}_\alpha(v) = \frac{1}{Mh} \sum_{j=1}^{n} w_j \sum_{k=1}^{N_j} \hat{L}\left(\frac{v - X_{jk}}{h}\right),$$

where the weights w_j are nonnegative, $M = \sum_j N_j$, and

$$\hat{L}(u) = \frac{1}{2\pi} \int \exp\{-itu\} \frac{\psi_K(t)}{\hat{\psi}_Z(t/h) + \rho} dt.$$

Here, it is assumed that the kernel function is symmetric and that ρ is a nonnegative ridge parameter. The authors show that there is no first-order loss in performance of the estimator when using replicated data to estimate f_α in comparison to the case when the error distribution is fully known. This result holds for an ordinary smooth error as long as f_α is smoother than half the derivative of f_Z.

1.2.2 Distribution Function Estimation

The problem of distribution estimation in the LRE model is considered in Hall and Yao (2003). They provide general conditions under which both the error and random effects distributions can be estimated when the number of replications is as small as two. In addition, two types of estimators are proposed for estimating the error and random effects distributions. The two methods also apply to estimating

the error and random effects densities. The first estimator is based on explicit characteristic function inversion; that is, an explicit form of each characteristic function is found based on the characteristic function of the observed data. The error and random effects densities and distributions are then found by Fourier inversion using a truncated version of each characteristic function. The second method is a histogram based estimator that uses a type of minimum distance. While the issue of efficiency is not addressed, numerical properties are discussed, and consistency of both estimators is established.

Furthermore, Neumann (2007) proposes a method to estimate the characteristic functions of α and Z using minimum distance. The distribution functions corresponding to these characteristic functions are then taken as estimators for the true error and random effects distributions. Consistency is shown to hold under weaker conditions than in Hall and Yao (2003). While the proposed method uses minimum distance, Neumann (2007) does not discuss a practical way to implement minimum distance to find estimates for the characteristic functions of α and Z.

Most recently, Hart and Canette (2009) investigate a minimum distance method to nonparametrically estimate distribution functions in the LRE and location-scale random effects (LSRE) models. The method relies on having an N-vector estimate of quantiles for the distribution of interest. The algorithm randomly jitters these quantile estimates and then sorts the elements to yield a new quantile estimate. The goodness of the new quantile estimate is assessed using a metric D. If the new quantile estimate yields a smaller value for the metric than the previous estimate, it is accepted as the best estimate of the quantiles for the distribution of interest. This is repeated using local and global iterations until the metric fails to change by a certain amount over a specified number of consecutive iterations. This methodology is used to simultaneously estimate the error and random effects distributions when

$n \geq 2$ in the LRE model. A location free method for estimating the error distribution when $n \geq 3$ is also proposed. This location free method relies on a result that the authors prove regarding identifiability of the error distribution from residuals. Finally, a methodology is proposed for estimating the error distribution and scale random effects distribution in the LSRE model.

1.3 Other Related Results

As with several other authors, Cox and Hall (2002) consider the LRE model in a high dimension, small sample size framework. However, instead of estimating a density or distribution function, they propose an estimator for the moments and cumulants of the error and random effects distributions. The methodology is based on using a general homogeneous polynomial in the data. An exact form for the second, third, and fourth-order moments is explicitly derived.

The regression setting is closely related to the two models under consideration. Beran and Millar (1994) consider the random coefficient regression model

$$Y_i = A_i + X_i B_i, \quad i \geq 1,$$

where A_i and B_i are $p \times 1$ random vectors, and X_i is a $p \times q$ random matrix. A minimum distance method is proposed to nonparametrically estimate the unknown distribution of (A_i, B_i) from the sample $\{(Y_i, X_i) : 1 \leq i \leq n\}$. The main idea behind this methodology is to choose an estimator so that the distribution of (Y_i, X_i) under the model is close to the empirical distribution of the sample. Consistency of the estimator, conditions for identifiability of the distribution of (A_i, B_i), and considerations when choosing the metric D are investigated. A semiparametric version of the model in which the distribution of (A_i, B_i) belongs to a parametric family is also considered.

CHAPTER II

DECONVOLUTION IN MEASUREMENT ERROR MODEL

2.1 Minimum Distance Method

For the measurement error model, the data consist of independent observations $Y_1, \ldots, Y_n$, where

$$Y_i = X_i + Z_i, \quad i = 1, \ldots, n.$$

The random variable X_i is assumed to come from an unknown probability density f_X, and the random noise variable Z_i is assumed to have a known probability density f_Z. Furthermore, the variables X_i and Z_i are assumed to be independent. However, X_i and Z_i are not directly observed, so ordinary density estimation techniques do not apply when estimating f_X.

The main idea behind the proposed method is to compare a density estimate of the observed data to a density induced from the model. That is, a density estimate $\hat{f}_Y$ of the data $y_1, \ldots, y_n$ is first obtained by some well established method. This density estimate is then compared to the known error density convolved with a candidate density $\widetilde{f}_X$ for f_X. The intuition is that the closer the density estimate $\hat{f}_Y$ is to the convolution of f_Z and $\widetilde{f}_X$, then the closer $\widetilde{f}_X$ is to f_X. The issue then becomes one of finding a density $\widetilde{f}_X$ so that the convolution of f_Z and $\widetilde{f}_X$ is very close to $\hat{f}_Y$.

More formally, let $\hat{f}_Y$ denote a kernel density estimate of the observed data $y_1, \ldots, y_n$ given by

$$\hat{f}_Y(y) = \frac{1}{nh} \sum_{j=1}^{n} K\left(\frac{y - y_j}{h}\right), \tag{2.1}$$

where K is a standard normal kernel, and h is the bandwidth. For a discussion of kernel density estimation and bandwidth selection, see Jones et al. (1996) and the

references therein. Let $\widetilde{f}_Y$ denote the density induced from assuming X_i has the density $\widetilde{f}_X$, which is a candidate density for f_X. The density of Y_i is then given by the convolution equation

$$\widetilde{f}_Y(y) = \int_{-\infty}^{\infty} \widetilde{f}_X(y-u) f_Z(u)\, du. \tag{2.2}$$

The estimate of f_X is obtained by choosing a suitable density $\widetilde{f}_X$ that minimizes the density-based metric

$$D^2(\hat{f}_Y, \widetilde{f}_Y) = \int_{-\infty}^{\infty} |\hat{f}_Y(y) - \widetilde{f}_Y(y)|^2 dy.$$

While any number of metrics can be used, Hart and Cañette (2009) point out that density-based metrics work better than those based on cumulative distribution functions. By Parseval's formula, $D^2(\hat{f}_Y, \widetilde{f}_Y)$ can be expressed as

$$D^2(\hat{f}_Y, \widetilde{f}_Y) = \frac{1}{2\pi} \int_{-\infty}^{\infty} |\hat{\psi}_Y(t) - \widetilde{\psi}_Y(t)|^2 dt, \tag{2.3}$$

where $\hat{\psi}_Y(t)$ and $\widetilde{\psi}_Y(t)$ are the characteristic functions corresponding to $\hat{f}_Y(y)$ and $\widetilde{f}_Y(y)$, respectively. A good introduction to characteristic functions can be found in Feller (1971), Lukacs (1970), and Ushakov (1999). Writing the metric in terms of characteristic functions leads to a simplification, as will be seen later.

2.2 Approximating Minimum Distance Estimate Using Normal Mixtures

2.2.1 Metric

The density function f_X is approximated by a normal mixture with L components centered at $\mu_1, \ldots, \mu_L$ having common scale parameter σ. That is, f_X is assumed to

be well approximated by a density of the form

$$\widetilde{f}_X(x) = \sum_{j=1}^{L} p_j \phi_j \left(\frac{x - \mu_j}{\sigma} \right), \tag{2.4}$$

where ϕ_j denotes a standard normal density, and p_j denotes the mixing proportion of ϕ_j. The means $\mu_1, \ldots, \mu_L$ are taken to be equally spaced. Representing f_X as a mixture of normals has the advantage that the density estimate has a parametric and hence tractable form. The characteristic function corresponding to $\hat{f}_Y(y)$ is

$$\hat{\psi}_Y(t) = \exp\{-h^2 t^2/2\} \hat{\psi}(t), \tag{2.5}$$

where

$$\hat{\psi}(t) = \frac{1}{n} \sum_{j=1}^{n} \exp\{y_j i t\}$$

is the empirical characteristic function of $y_1, \ldots, y_n$. Furthermore, the characteristic function corresponding to $\widetilde{f}_Y(y)$ is

$$\widetilde{\psi}_Y(t) = \sum_{j=1}^{L} p_j \psi_j(t) \psi_Z(t), \tag{2.6}$$

where

$$\psi_j(t) = \exp\left\{-\sigma^2 t^2/2\right\} \exp\{\mu_j i t\}$$

is the characteristic function of a normal random variable with mean μ_j and standard deviation σ. Then, by using Equations (2.5) and (2.6), the metric $D(\hat{f}_Y, \widetilde{f}_Y)$ can be written as

$$D^2(\hat{f}_Y, \widetilde{f}_Y) = \frac{1}{2\pi} \int_{-\infty}^{\infty} \left| \exp\left\{-h^2 t^2/2\right\} \hat{\psi}(t) - \sum\nolimits_{j=1}^{L} p_j \psi_j(t) \psi_Z(t) \right|^2 dt. \tag{2.7}$$

2.2.2 Discretizing the Metric

The metric in Equation (2.7) can be approximated by replacing the integral with a Riemann sum. Letting $\mathbf{y}^T = [\ \hat{\psi}_Y(t_1)\ \ \hat{\psi}_Y(t_2)\ \ \cdots\ \ \hat{\psi}_Y(t_m)\ ]$ and defining the matrix $\mathbf{H}$ to be $(h_{ij})_{m\times L} = \psi_j(t_i)\,\psi_Z(t_i)$ reduces the metric to a complex least squares criterion given by

$$\begin{aligned} 2\pi D^2(\hat{f}_Y, \widetilde{f}_Y) &\approx \sum_{i=1}^{m} \left| \exp\left\{-h^2 t_i^2/2\right\} \hat{\psi}(t_i) - \sum\nolimits_{j=1}^{L} p_j \psi_j(t_i)\,\psi_Z(t_i) \right|^2 \Delta t \\ &= \Delta t \cdot |\mathbf{y} - \mathbf{H}\mathbf{p}|^2, \end{aligned} \tag{2.8}$$

where $t_1, \ldots, t_m$ are equally spaced points on $[-t_{\max}, t_{\max}]$ and $\Delta t = t_2 - t_1$. Since the integral defined in Equation (2.7) has infinite upper and lower limits, it becomes necessary to choose a point at which to truncate the integral. The quantity $t_{\max}$ is chosen so that the value of the integral outside the interval $[-t_{\max}, t_{\max}]$ in Equation (2.7) is less than 0.001. Denoting the integrand of the metric in Equation (2.7) by $I(t)$, it is true by the Cauchy-Schwarz inequality that

$$\begin{aligned} I(t) \le (&|\exp\left\{-h^2t^2/2\right\}|^2 + |\exp\left\{-\sigma^2t^2/2\right\}|^2) \\ &\times \left(|\hat{\psi}(t)|^2 + \left| \sum\nolimits_{j=1}^{L} p_j \exp\{\mu_j i t\}\,\psi_Z(t) \right|^2 \right). \end{aligned}$$

This can be further bounded and simplified by writing

$$\begin{aligned} I(t) &\le 2\left(\exp\left\{-h^2t^2\right\} + \exp\left\{-\sigma^2t^2\right\}\right) \\ &\le 4\exp\left\{-\min(h, \sigma_{\min})^2 t^2\right\}. \end{aligned}$$

Here, $\sigma_{\min}$ is the smallest value of σ considered in the algorithm. The cutoff point $t_{\max}$ is then chosen so that

$$\int_{t_{\max}}^{\infty} 8\exp\left\{-\min(h, \sigma_{\min})^2 t^2\right\} dt \le 0.001,$$

which occurs at

$$t_{\max} = \frac{1}{\sqrt{2}\min(h, \sigma_{\min})}\Phi^{-1}\left(1 - \frac{0.001\min(h, \sigma_{\min})}{8\sqrt{\pi}}\right). \tag{2.9}$$

The next several sections describe using the discretized metric to estimate the mixing proportions $\mathbf{p}$, the scale parameter σ, and the number of mixing components L.

2.2.3 Estimating Mixing Proportions $\mathbf{p}$

For a given number of components L and scale σ, there are several possible ways to estimate the vector of mixing proportions $\mathbf{p}$. The most straightforward way of estimating $\mathbf{p}$ is by ordinary least squares. However, it turns out that the structure of the problem does not inherently satisfy the constraints $\sum_{j=1}^{L} p_j = 1$ and $p_j \geq 0$, $j = 1, \ldots, L$. Similarly, when $\mathbf{p}$ is estimated using the method of Lagrange multipliers by enforcing $\sum_{j}^{L} p_j = 1$, the constraint that $p_j \geq 0$, $j = 1, \ldots, L$, is not naturally satisfied from the structure of the problem. Thus, a quadratic programming algorithm is used to find $\mathbf{p}$ such that $D(\hat{f}_Y, \widetilde{f}_Y)$ is maximized subject to the constraints that $\sum_{j=1}^{L} p_j = 1$ and $p_j \geq 0$, $j = 1, \ldots, L$.

2.2.4 Estimating σ

The next step is to estimate the scale parameter σ. The algorithm for choosing σ is motivated from the observation that the best choice of σ does not change beyond some point at which the number of normal mixture components is "large enough" to represent the underlying target density. To estimate σ, a value of L is first selected that is almost certainly too large. For example, in all of the simulated examples in the next section, between five and eight normal mixtures are sufficient to approximate the target density f_X. Thus, a value of $L = 20$ or 25 could be used in this case. After fixing a large value of L, the best choice of σ is chosen to maximize the cross-

validated log-likelihood of the induced density $\widetilde{f}_Y(y)$. That is, for each candidate value of σ under consideration, leave-one-out or k-fold cross-validation can be used on the observed data, although leave-one-out may be too computationally intensive when n is greater than several hundred. To do this, the optimal mixing proportions $\mathbf{p}^*$ that minimize Equation (2.8) using the training data are found. The testing data are then used to assess the fit of the induced density $\widetilde{f}_Y(y)$ using the $\mathbf{p}^*$ that was found by minimizing the metric from the training data. The value of σ that minimizes the log-likelihood averaged across the testing data is selected as the best choice of σ.

Theoretically, the target density f_X is identifiable when the error density f_Z and the density of the observed data f_Y are known. However, the problem is ill-posed, and there are *near identifiability* issues. That is, while there is only one density f_X whose convolution with f_Z yields f_Y, there are many densities vastly different from f_X that yield densities very close to f_Y when convolved with f_Z. However, these poor estimates of f_X typically arise at small values of σ. The k-fold cross-validation method for selecting σ provides a reasonable way of discouraging the algorithm from selecting these poor estimates even though the convolution of the poor estimate with f_Z is close to $\hat{f}_Y$. This is because the density estimates that result from small values of σ computed from different training samples vary widely and therefore do not provide the best fit to the data withheld for testing.

2.2.5 Estimating Number of Components L

To select the means $\mu_1, \mu_2, \ldots, \mu_L$ of the normal mixture components, a rough estimate for the support of X_i must be obtained. That is, using a mixture approach requires finding an interval $(M_{\min}, M_{\max})$ that covers most of the mass of the proba-

bility density function of X_i. One way of selecting $M_{\min}$ and $M_{\max}$ is

$$M_{\min} = \min_i (y_i) - Q_Z\left(1 - \frac{1}{2n}\right) \quad \text{and} \quad M_{\max} = \max_i (y_i) - Q_Z\left(\frac{1}{2n}\right), \tag{2.10}$$

where Q_Z denotes the quantile function of Z_i. This is chosen by assuming that the smallest observed value of the data, $y_{(1)}$, is the sum of the smallest value of X and the largest value of Z that can be reasonably expected given the sample size. While this results in a conservative estimate for the range of X_i that contains most of the mass of the distribution, there is no penalty in the algorithm for doing this. That is, the algorithm will assign an identically zero or near zero mixing proportion to normal mixture components whose locations are too extreme.

For a fixed L, one way to select the location of the components is

$$\mu_j = M_{\min} + j\left(\frac{M_{\max} - M_{\min}}{L+1}\right), \quad j = 1, \ldots, L, \tag{2.11}$$

although many other ways are possible. Furthermore, suppose that L^* normals are sufficient to represent f_X. As mentioned earlier, the estimate of f_X that results from using L^* normal components does not appreciably change as the number of normals increases past this point. This is in opposition to the intuition that the density estimate should become less smooth as the number of normals increases. Thus, using more than L^* normals is a matter of redundancy and does not affect the resulting estimate of f_X. Once a good value of σ is selected as discussed in the previous section, the number of mixture components can then be chosen by looking at the value of the metric in Equation (2.8) computed for increasing values of L. The value of the metric should substantially decrease and then level off as the number of components increases. Once the value of the metric stops changing by some specified amount (i.e., two or three percent), the number of normals at which this occurs is selected as the number of components to use for the mixture. Figure 1 shows a typical example

of how the metric changes as the number of normal mixture components increases. In this case, the density estimate does not substantially improve as the number of components increases beyond six normals.

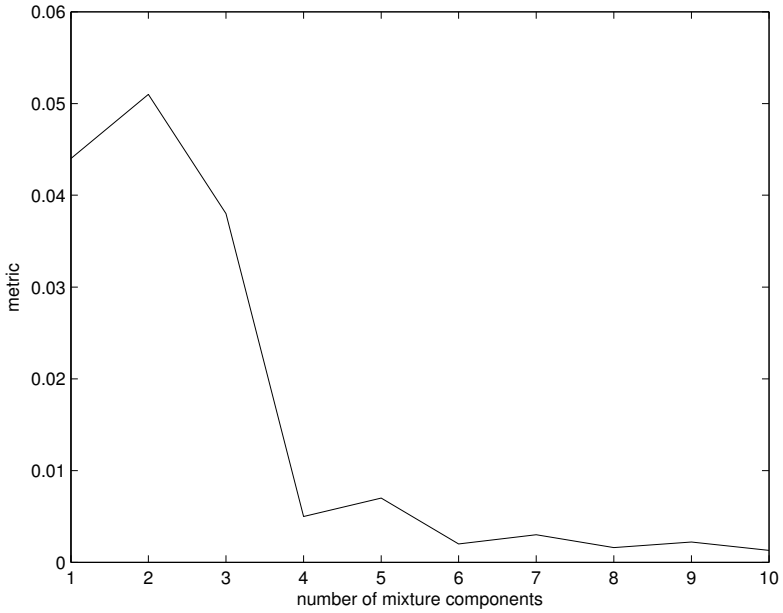

Figure 1. Example of how the distance metric typically changes with the number of normal mixture components.

2.2.6 Outline for Showing Consistency of Estimator

Although a more rigorous mathematical treatment for consistency of the estimator can be found in Appendix A, this section provides a brief outline. Recall that since an explicit form for $\mathbf{p}$ cannot be found, the mixing proportions can be estimated using an optimization routine. This makes it very difficult to establish rates of convergence, and so the treatment of consistency is more of an existence argument.

Since the underlying target density is represented as a finite mixture of normals, it in general is not possible for a fixed L and σ to perfectly represent the target density. There are exceptions to this such as when the target density is normal. Therefore, it holds that as $n \to \infty$ for a fixed L and σ, the estimator approaches some approximation to the target density. It must then hold that the approximation approaches the target density as $L \to \infty$ and $\sigma \to 0$. A "ratchet" argument can then be used to say that as n and L go to infinity and σ goes to 0, the estimator approaches the true target density. Care must be taken to ensure that the support on which the normals are placed goes to infinity and the distance between the normal components approaches 0. Finally, since the convergence argument applies to characteristic functions, conditions for which convergence of characteristic functions implies convergence of density functions must be established.

2.3 Simulated Examples

Simulations using the previous methodology are implemented to assess the performance of the minimum distance approach with other known deconvolution methods. Specifically, the examples are chosen in order to compare the performance with the deconvoluting kernel estimator introduced in Carroll and Hall (1988) and Stefanski and Carroll (1990) using the bandwidth selection methods compared in Delaigle and

Gijbels (2004). In the latter paper, four bandwidth selection methods are compared using simulated data from several target and error densities. Hence, following Delaigle and Gijbels, the two error densities, $N(0, \sigma_Z^2)$ and Laplace$(0, \sigma_Z)$, for f_Z are considered using sample sizes 50, 100, 250, and 500. The four target densities considered for f_X are defined by

(1) Density 1: $X \sim N(0, 1)$;

(2) Density 2: $X \sim \chi^2(3)$;

(3) Density 3: $X \sim \frac{1}{2}N(-3, 1) + \frac{1}{2}N(2, 1)$;

(4) Density 4: $X \sim \frac{2}{5}$ Gamma$(5, 1) + \frac{3}{5}$ Gamma$(13, 1)$.

Figure 2 shows these densities in order of increasing difficulty to estimate in terms of recovering the general features of the density. That is, detecting bimodality of the normal mixture is more difficult that capturing the skewness of the chi-square. Similarly, the bimodal gamma mixture has two modes of different size, which should be more difficult to recover than the mixture of two normals whose modes are identical. This difficulty in estimation does not refer to asymptotic rates of convergence.

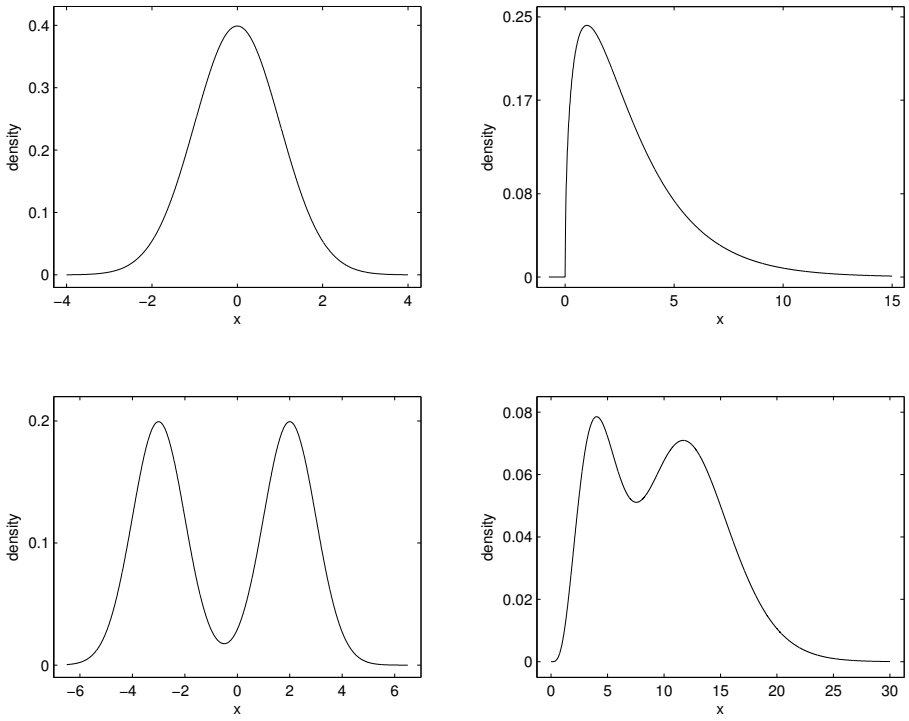

Figure 2. Target densities of increasing difficulty to estimate; density 1 (top left), density 2 (top right), density 3 (bottom left), density 4 (bottom right).

For densities 1, 2, and 3, the error variance is selected so that the noise-to-signal ratio $Var\left(Z\right)/Var\left(X\right)$ is 0.25. Since density 4 has more subtle features to recover, the noise-to-signal ratio for this density is set to 0.10.

To evaluate the performance of the estimator, 100 samples are generated for each of the eight target and error density combinations. That is, a density estimate is computed for each sample using data with density f_Y which are generated by computing $Y_i = X_i + Z_i$. The density estimate $\hat{f}_X$ is computed for each sample and compared with the true density using the ISE criterion given by

$$ISE(\hat{f}_X, f_X) = \int (\hat{f}_X\left(x\right) - f_X\left(x\right))^2 dx.$$

Table 1 summarizes the ISE of the estimates under a normal error density using minimum distance estimation, while Table 2 summarizes the results reported by Delaigle and Gijbels using their best bandwidth selection method. In Table 2, the reported ISE for estimating density 1 makes use of a normal reference bandwidth while the best bandwidth selection method for the other densities is the plug-in method discussed in Chapter I. The missing values in Table 2 denote simulation results that were not reported in their paper. Thus, following Delaigle and Gijbels, the medians and interquartile ranges of the 100 values of ISE are reported for each target density and sample size combination.

Table 1. Median ISE of density estimates under normal measurement error using minimum distance estimation. The numbers in parentheses are interquartile ranges of ISE for 100 replications.

	Target Density			
n	*1* - $N(0,1)$	*2* - $\chi^2(3)$	*3 - Normal Mix*	*4 - Gamma Mix*
50	0.0098 (0.0146)	0.0213 (0.0154)	0.0346 (0.0235)	0.0052 (0.0035)
100	0.0062 (0.0083)	0.0176 (0.0106)	0.0228 (0.0248)	0.0033 (0.0022)
250	0.0029 (0.0044)	0.0143 (0.0094)	0.0139 (0.0229)	0.0022 (0.0013)
500	0.0018 (0.0024)	0.0114 (0.0035)	0.0086 (0.0142)	0.0017 (0.0010)

Table 2. Median ISE of density estimates reported by Delaigle and Gijbels under normal measurement error using the deconvoluting kernel density estimator. The numbers in parentheses are interquartile ranges of ISE for 500 replications.

	Target Density			
n	*1* - $N(0,1)$	*2* - $\chi^2(3)$	*3* - *Normal Mix*	*4* - *Gamma Mix*
50	0.011 (0.012)	–	–	–
100	0.0080 (0.0075)	0.018 (0.0084)	0.027 (0.013)	–
250	0.0051 (0.0042)	–	0.020 (0.011)	0.0023 (0.0011)

The medians and interquartile ranges of the ISE for density 1 are slightly lower than those reported in Delaigle and Gijbels using their best bandwidth selection method. The performance, as measured by the ISE, is comparable to Delaigle and Gijbels using the best bandwidth for densities 2 and 4. However, the medians of the ISE for density 3 are substantially lower than those reported in Delaigle and Gijbels. A lower median appears to come with the price that the interquartile range of the ISE is nearly twice that of the one reported in their paper. Figures 3, 4, 5, and 6 show the resulting minimum distance estimates for each of the four target densities under normal measurement error.

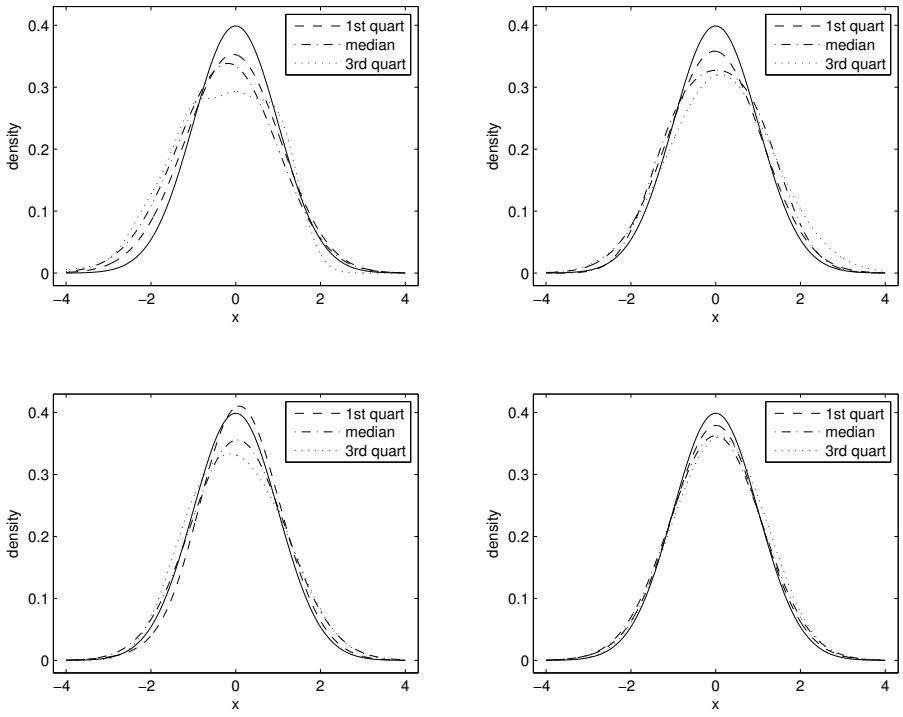

Figure 3. Minimum distance estimates for N(0,1) density (density 1) under normal measurement error at sample size n=50 (top left), n=100 (top right), n=250 (bottom left), n=500 (bottom right).

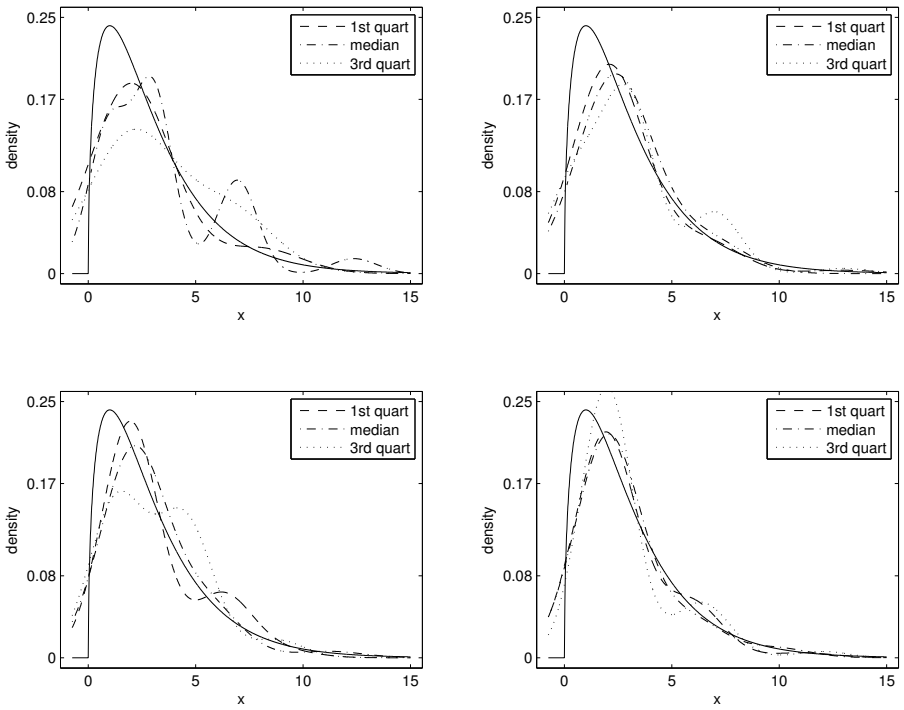

Figure 4. Minimum distance estimates for $\chi^2(3)$ density (density 2) under normal measurement error at sample size n=50 (top left), n=100 (top right), n=250 (bottom left), n=500 (bottom right).

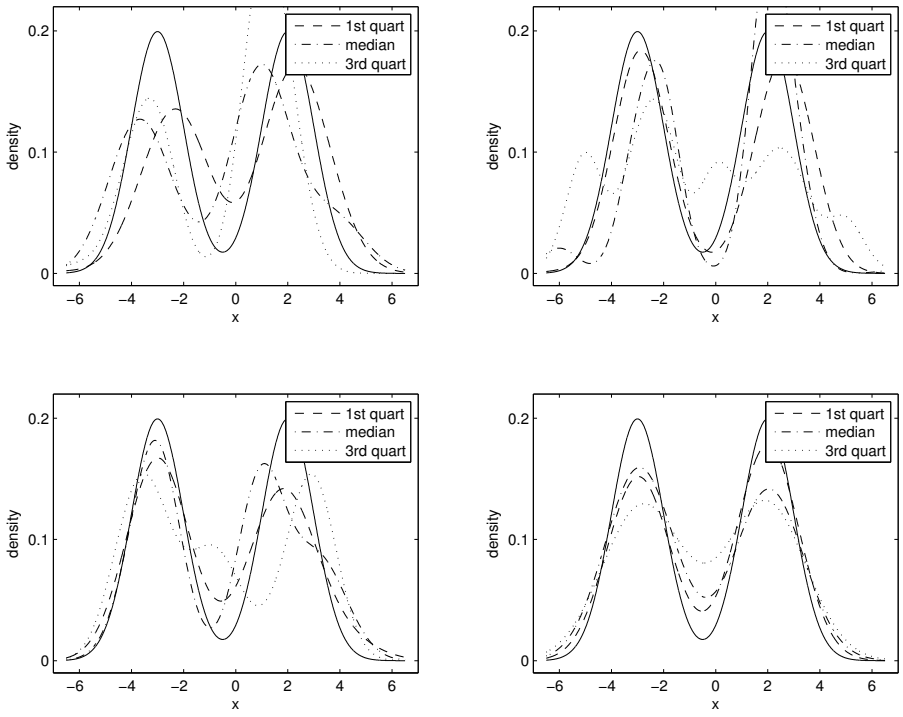

Figure 5. Minimum distance estimates for normal mixture density (density 3) under normal measurement error at sample size n=50 (top left), n=100 (top right), n=250 (bottom left), n=500 (bottom right).

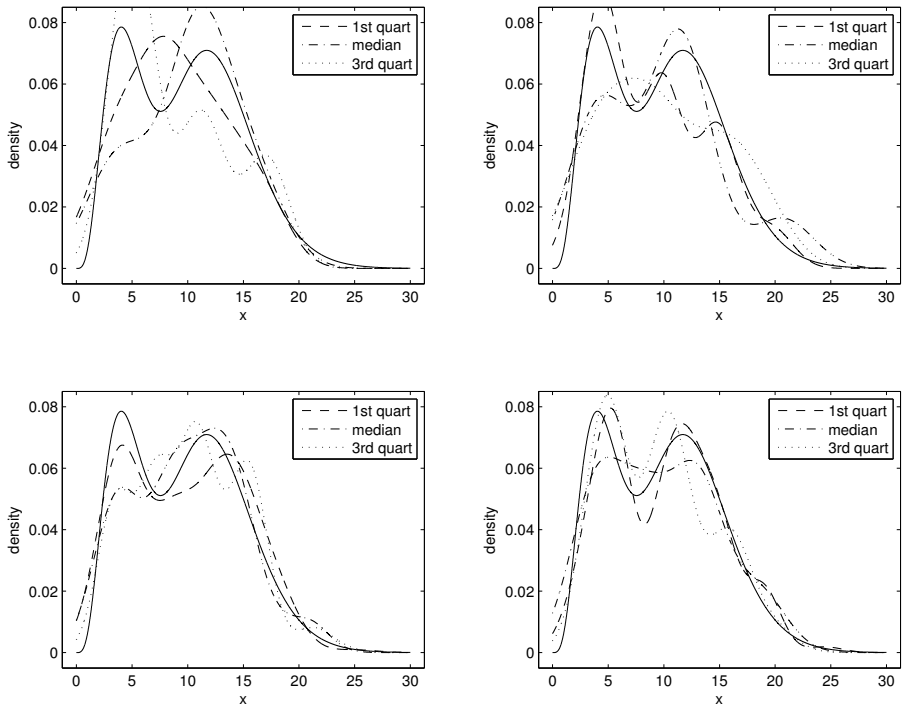

Figure 6. Minimum distance estimates for gamma mixture density (density 4) under normal measurement error at sample size n=50 (top left), n=100 (top right), n=250 (bottom left), n=500 (bottom right).

As seen from the previous plots, the estimates get closer to the true target density as the sample size increases. The most challenging part of the algorithm is estimating σ. This weakness of the algorithm is evident in the worst 10-20% of the 100 density estimates. That is, when a poor estimate occurs, it is usually the result of too small a choice of σ.

The results in Table 3 summarize the ISE of the estimates under a Laplace error density using minimum distance estimation, while Table 4 summarizes the results reported by Delaigle and Gijbels using their best bandwidth selection method. As in the case for normal measurement error, the reported ISE in Table 4 corresponding to density 1 makes use of a normal reference bandwidth while the best bandwidth selection method for the other densities is the plug-in method discussed in Chapter I. The missing values in Table 4 denote simulation results that were not reported in their paper.

Table 3. Median ISE of density estimates under Laplace measurement error using minimum distance estimation. The numbers in parentheses are interquartile ranges of ISE for 100 replications.

	Target Density			
n	*1 -* $N(0,1)$	*2 -* $\chi^2(3)$	*3 - Normal Mix*	*4 - Gamma Mix*
50	0.0098 (0.0150)	0.0181 (0.0086)	0.0224 (0.0200)	0.0043 (0.0029)
100	0.0058 (0.0089)	0.0168 (0.0089)	0.0153 (0.0108)	0.0030 (0.0017)
250	0.0027 (0.0041)	0.0109 (0.0047)	0.0088 (0.0098)	0.0020 (0.0013)
500	0.0018 (0.0019)	0.0083 (0.0044)	0.0056 (0.0064)	0.0015 (0.0012)

Table 4. Median ISE of density estimates reported by Delaigle and Gijbels under Laplace measurement error using the deconvoluting kernel density estimator. The numbers in parentheses are interquartile ranges of ISE for 500 replications.

	Target Density			
n	*1 -* $N(0,1)$	*2 -* $\chi^2(3)$	*3 - Normal Mix*	*4 - Gamma Mix*
50	0.011 (0.011)	–	–	–
100	0.0071 (0.0080)	0.015 (0.0063)	0.018 (0.012)	–
250	0.0041 (0.0035)	–	0.011 (0.0074)	0.0021 (0.0011)

The medians and interquartile ranges of the ISE for densities 1 and 3 are lower than those reported in Delaigle and Gijbels using their best choice of bandwidth while the performance for densities 2 and 4 is comparable. Furthermore, the performance using the Laplace error density is better than the performance when a normal error density is used as can be seen by comparing Tables 1 and 3. This is a consequence of the known result that the smoother the error density, the harder it is to estimate the target density. See Fan (1992) for a detailed discussion. Since the normal density is considered a supersmooth density, and the Laplace is a ordinary smooth density, using a Laplace error density should yield better estimates of the target density for a given sample size and error variance. Figures 7, 8, 9, and 10 show the resulting minimum distance estimates for each of the four target densities under Laplace measurement error.

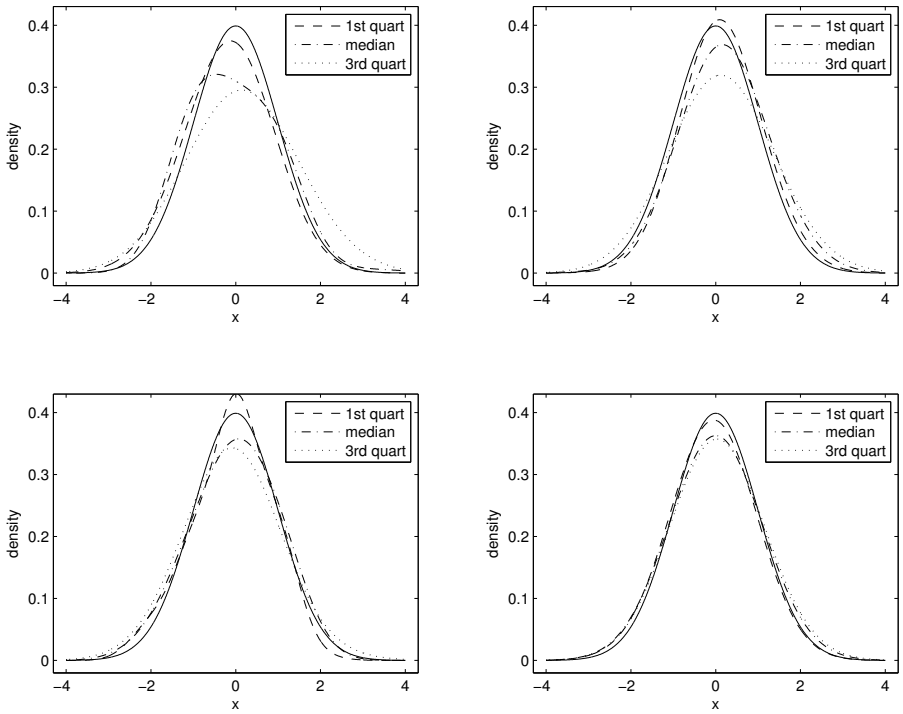

Figure 7. Minimum distance estimates for N(0,1) density (density 1) under Laplace measurement error at sample size n=50 (top left), n=100 (top right), n=250 (bottom left), n=500 (bottom right).

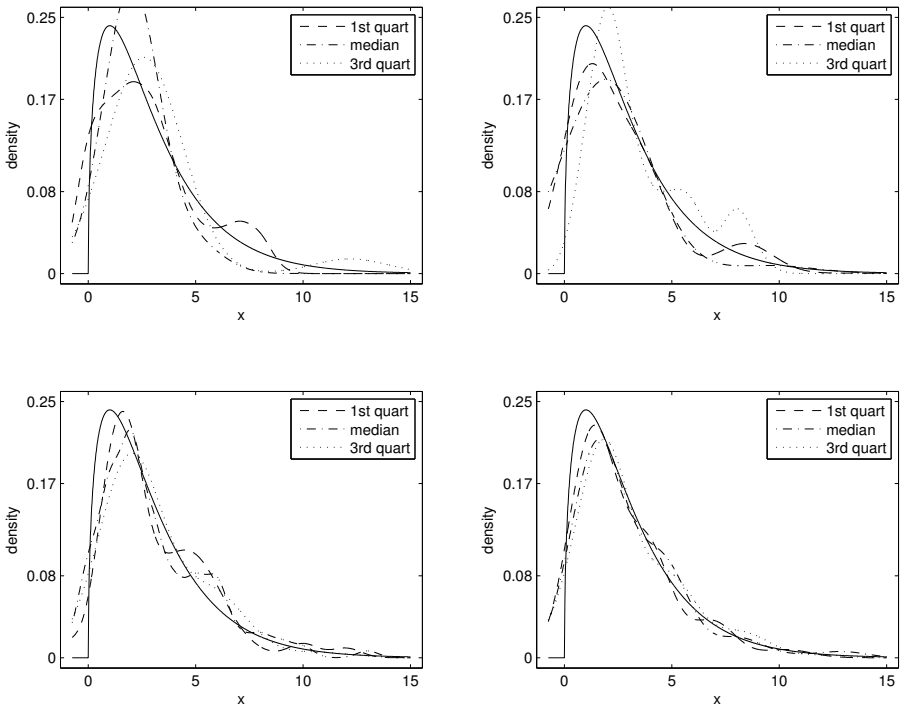

Figure 8. Minimum distance estimates for $\chi^2(3)$ density (density 2) under Laplace measurement error at sample size n=50 (top left), n=100 (top right), n=250 (bottom left), n=500 (bottom right).

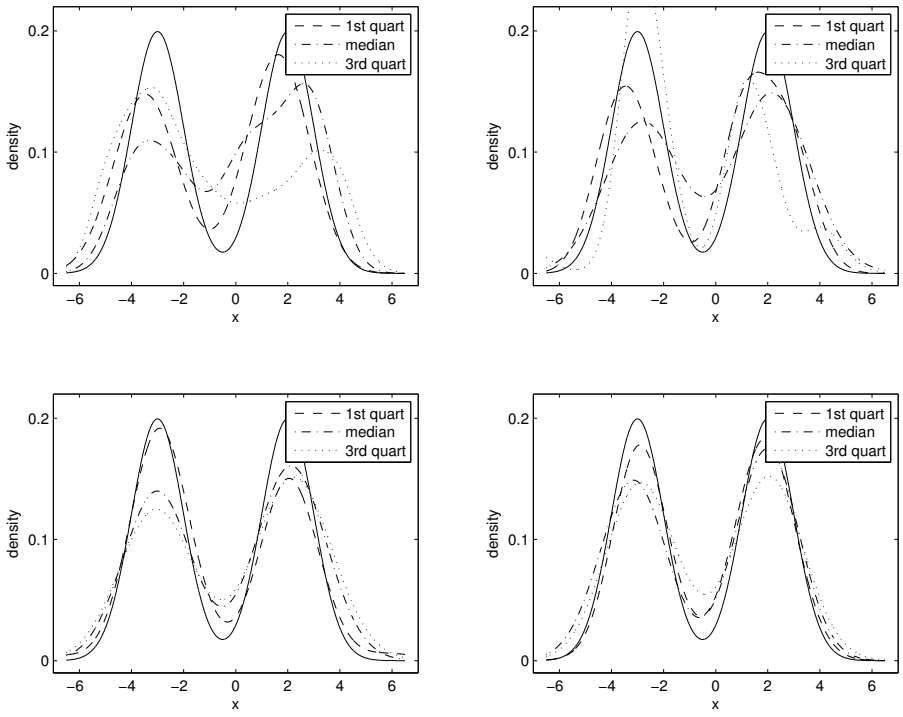

Figure 9. Minimum distance estimates for normal mixture density (density 3) under Laplace measurement error at sample size n=50 (top left), n=100 (top right), n=250 (bottom left), n=500 (bottom right).

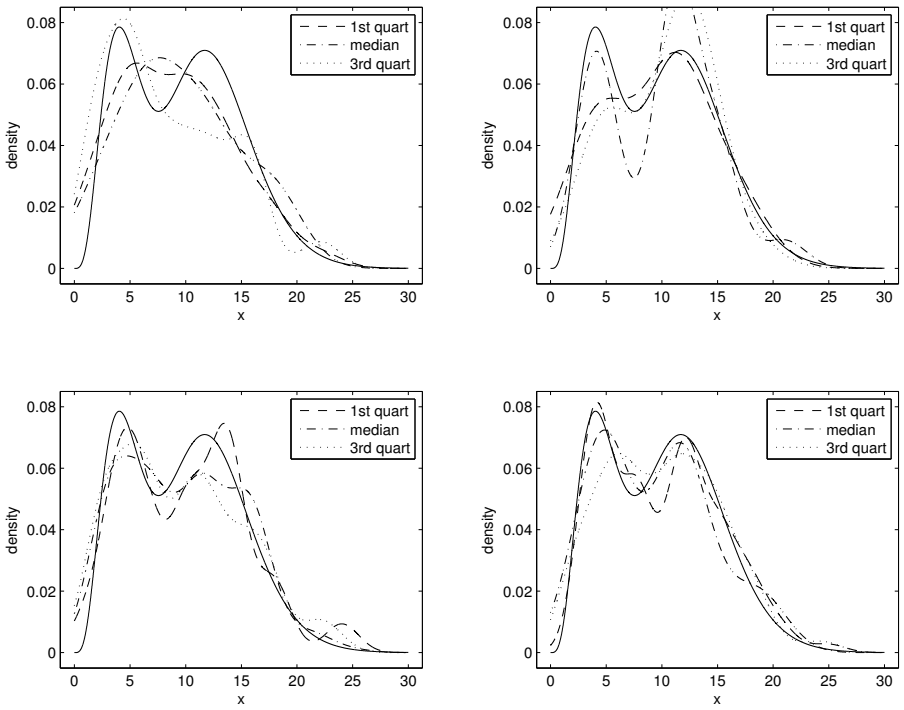

Figure 10. Minimum distance estimates for gamma mixture density (density 4) under Laplace measurement error at sample size n=50 (top left), n=100 (top right), n=250 (bottom left), n=500 (bottom right).

2.4 Applications

The minimum distance methodology is applied to the same 2888 observations that Stefanski and Carroll (1990) use to assess the performance of their estimator. This data set originally came from NHANES I (1970-75) and a cohort study (Jones et al. 1987) that investigated the relationship between long-term daily saturated fat intake and incidence of breast cancer in women between the ages of 25 and 74 years.

The goal is to estimate the density of the long-term logarithm of daily saturated fat intake using the 2888 observations. As pointed out in Stefanski and Carroll (1990), rough estimates of the error variance of the data suggest that 50% or more of the variability in the data may be a result of measurement error. Performing deconvolution with this much noise is challenging, especially for a sample size of 2888. However, some knowledge of the underlying target distribution can be attained by deconvolving the density while assuming less noise. Thus, following Delaigle and Gijbels (2004) and Stefanski and Carroll (1990), normal and Laplace error densities are considered while assuming error variances $\sigma_Z^2 = (1/5)\,\sigma_X^2$, $\sigma_Z^2 = (1/3)\,\sigma_X^2$, and $\sigma_Z^2 = 1.5\sigma_X^2$. In addition, as a basis for comparison, estimates in the error free case when $\sigma_Z^2 = 0$ are computed.

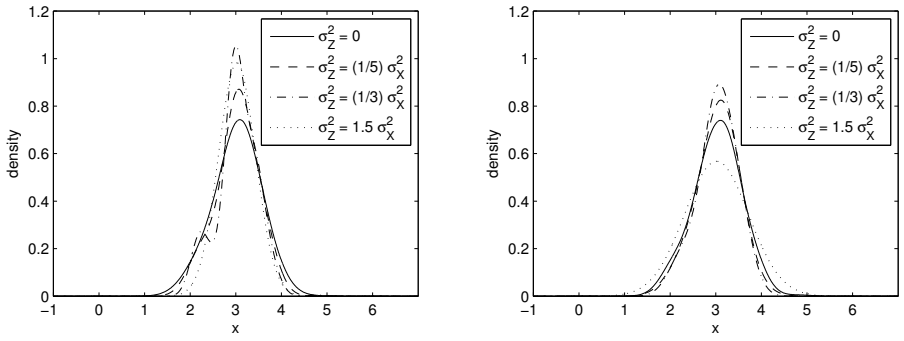

Figure 11. Minimum distance estimates for log daily saturated fat assuming normal (left) and Laplace error densities (right) when $\sigma_Z^2 = 0$, $1/5\sigma_X^2$, $1/3\sigma_X^2$, and $1.5\sigma_X^2$.

A comparison of the two plots in Figure 11 suggests that when $\sigma_Z^2 = 0$, $\sigma_Z^2 = (1/5)\,\sigma_X^2$, and $\sigma_Z^2 = (1/3)\,\sigma_X^2$, the resulting density estimates do not substantially change by assuming different symmetric error densities. In addition, the estimates do not appear to be very sensitive to fairly large error variances, although when $\sigma_Z^2 = 1.5\sigma_X^2$, the estimates are more sensitive to the underlying error density. However, caution should be exercised before accepting any of the estimates as the "true" underlying density of long-term logarithm of daily saturated fat intake. Since the fat intake was self-reported, it is very possible that fat intake was systematically under-reported. This could make the error density skewed or bimodal, causing the density estimates under the assumption of symmetric error densities to be invalid. Compared with Delaigle and Gijbels (2004) and Stefanski and Carroll (1990), the density estimates for these data do not appear to suffer from the "wiggly" effects in the tails, which often result from methods that rely on the ratio of two characteristic functions.

CHAPTER III

DECONVOLUTION IN LOCATION RANDOM EFFECTS MODEL

3.1 Minimum Distance Method

For the location random effects (LRE) model, the data consist of observations X_{ij}, where

$$X_{ij} = \alpha_i + \gamma Z_{ij}, \quad i = 1, \ldots, p, \quad j = 1, \ldots, n.$$

The random variable Z_{ij} is assumed to come from an unknown probability density f_Z having mean 0 and standard deviation 1. Furthermore, the random variable α_i is assumed to come from an unknown density f_α. The variables Z_{ij} and α_i are assumed to be independent. As in the measurement error model, Z_{ij} is not observed directly, so ordinary density estimation techniques do not apply when estimating f_Z. In addition, the LRE model can be viewed as a measurement error model with repeated measurements on the variable α_i. Since there are multiple measurements for each α_i, it is not necessary to assume that the density of α_i is known. Note that the LRE model reduces to the ordinary measurement error model when $n = 1$. However, in that case, to ensure identifiability of f_Z it must be assumed that the density of α_i is known.

The main idea for estimating the density f_Z is analogous to estimation in the measurement error model. However, in comparison to the previous model, there are many more ways of implementing this. For example, both f_α and f_Z may be identified by directly considering the multivariate density of the n-dimensional vector $(X_{i1}, X_{i2}, \ldots, X_{in})$, $i = 1, \ldots, p$. However, even for a moderate size n, working with an n-dimensional joint density is computationally impractical. It then becomes necessary

to consider some lower dimensional transformation of the original n-dimensional data $(X_{i1}, X_{i2}, \ldots, X_{in})$, $i = 1, \ldots, p$. For example, the density of $(X_i, X_j),\ 1 \leq i < j \leq n$, may be used to identify f_α and f_Z. Perhaps the univariate data $(X_{ij} - \bar{X}_{i\cdot}),\ i = 1, \ldots, p,\ j = 1, \ldots, n$, or the differences $(X_i - X_j, X_i - X_k),\ 1 \leq i < j < k \leq n$, could be used to only identify f_Z since these two transformations of the original data are free of α_i. Transformed data that are free of the location are more useful since f_Z is of primary interest. It is worth noting, however, that in principle both f_α and f_Z could simultaneously be estimated using minimum distance as several authors have done using other methods. A straightforward way of doing this would be to apply minimum distance to first estimate f_Z using observed data that is location free. The density estimate of f_Z could then be treated as known and minimum distance could be reapplied to estimate f_α.

For generality, let $\mathbf{Y}_i = (Y_{i1}, \ldots, Y_{im}),\ 1 \leq i \leq q,\ 1 \leq m \leq n$, be an m-dimensional transformed data set of size q computed from the original data X_{ij}, $i = 1, \ldots, p,\ j = 1, \ldots, n$. This will be referred to as *observable* data since it is not necessarily the original data, but is a function of it. Furthermore, assume that the density of $(Y_{i1}, \ldots, Y_{im})$, denoted by $f_{\mathbf{Y}}$, does not depend on the distribution of α_i and that f_Z is identifiable from $f_{\mathbf{Y}}$. Note that the distribution of $\mathbf{Y}$ is observable since it can be estimated in a straightforward way. That is, compute an $m-$dimensional density estimate, $\hat{f}_{\mathbf{Y}}$, using the transformed data $(Y_{i1}, \ldots, Y_{im}),\ 1 \leq i \leq q,\ 1 \leq m \leq n$. A candidate error density $\widetilde{f}_Z$ is used to calculate the induced density $\widetilde{f}_{\mathbf{Y}}$ of $(Y_1, \ldots, Y_m)$, which is then compared with the density estimate $\hat{f}_{\mathbf{Y}}$. As in the measurement error model, the intuition is that the closer $\widetilde{f}_{\mathbf{Y}}$ is to the density estimate $\hat{f}_{\mathbf{Y}}$, then the closer $\widetilde{f}_Z$ is to f_Z.

Unlike in the measurement error model, the density $\widetilde{f}_{\mathbf{Y}}$ in many cases does not have an explicit form and may need to be calculated by simulation or numerical

integration if an alternative method cannot be found. An estimate of f_Z is then obtained by choosing a suitable density $\widetilde{f}_Z$ that minimizes the density-based metric given by

$$D^2(\hat{f}_{\mathbf{Y}}, \widetilde{f}_{\mathbf{Y}}) = \int \cdots \int_{\mathbb{R}^m} |\hat{f}_{\mathbf{Y}}(\mathbf{y}) - \widetilde{f}_{\mathbf{Y}}(\mathbf{y})|^2 d\mathbf{y}, \tag{3.1}$$

where $\widetilde{f}_{\mathbf{Y}} = G(\widetilde{f}_Z)$ for some functional G.

An important consideration is under what conditions the error density f_Z is identifiable from the density $f_{\mathbf{Y}}$. This is investigated in the next section.

3.2 Identifiability of Error Density

3.2.1 Existing Identifiability Results

The are several useful identifiability results that can be used in various ways to identify f_Z in the LRE model. The problem of identifiability in the LRE model dates back at least as early as Reiersøl (1950) in which the LRE model is a special case. In this paper, Reiersøl shows that the distributions of α_i and Z_{ij} are identifiable even for a sample size as small as $n = 2$. Apparently, this is not a well known result as many authors that have studied the LRE model do not make use of it. When $n = 2$, the distributions of α_i and Z_{ij} are identifiable from the distribution of (X_{i1}, X_{i2}) as long as the characteristic functions of α_i and Z_{ij} to have only discrete zeros.

Another useful identifiability result that is applicable to this model is proven by Kotlarski (1967) and discussed in Rao (1992). Define $U_1 = V_1 - V_3$ and $U_2 = V_2 - V_3$, and assume that the characteristic function of (U_1, U_2) does not vanish. Then the joint distribution of (U_1, U_2) determines the distributions of V_1, V_2, and V_3 up to a change in location. While this result requires the stronger condition that the characteristic function of (U_1, U_2) is nonvanishing everywhere, it can in principle be used to identify three distributions. This result also has the advantage that the

characteristic functions corresponding to V_1, V_2, and V_3 can be explicitly determined. Several authors, including Li and Vuong (1998), use these explicit expressions for the characteristic functions to form the basis for their nonparametric estimators.

Thus far, both of the identifiability results rely on using a bivariate distribution to identify both f_α and f_Z. However, Hart and Cañette (2009) prove that for i.i.d. random variables $W_1, \ldots, W_n$ with mean 0, the distribution of W is identifiable from the distribution of the univariate difference $(W_1 - \bar{W})$, where $\bar{W} = \sum_{i=2}^{n} W_i/(n-1)$. This result holds when $n \geq 3$ provided that the characteristic function of W does not vanish throughout an interval. Since the primary interest is in the error density f_Z, this result proves very useful in the next section.

3.2.2 Identifiability from Residuals

Without loss of generality, let $\gamma = 1$ in the LRE model for the remaining discussion. In practice, γ can easily be estimated by the unbiased estimator

$$\hat{\gamma} = \left(\frac{1}{p(n-1)} \sum\nolimits_{i=1}^{p} \sum\nolimits_{j=1}^{n} (X_{ij} - \bar{X}_{i\cdot})^2 \right)^{1/2},$$

where $\bar{X}_{i\cdot} = \frac{1}{n} \sum_{j=1}^{n} X_{ij}$. This estimator is $\sqrt{p}$ consistent provided that the first two moments of the error exist. Now, define ϵ_{ij0} as

$$\epsilon_{ij0} = X_{ij} - \bar{X}_{i(j)} = Z_{ij} - \bar{Z}_{i(j)},$$

where $\bar{X}_{i(j)} = \frac{1}{n-1}(\sum_{k=1}^{n} X_{ik} - X_{ij})$, $i = 1, \ldots, p, \quad j = 1, \ldots, n$. Furthermore, define ϵ_{ijk1} and ϵ_{ijk2} as

$$\epsilon_{ijk1} = X_{ij} - \bar{X}_{i(jk)} = Z_{ij} - \bar{Z}_{i(jk)}$$

and

$$\epsilon_{ijk2} = X_{ik} - \bar{X}_{i(jk)} = Z_{ik} - \bar{Z}_{i(jk)},$$

where $\bar{X}_{i(jk)} = \frac{1}{n-2}(\sum_{\ell=1}^{n} X_{i\ell} - X_{ij} - X_{ik})$, $i = 1, \ldots, p$, $j = 1, \ldots, n$, $k = 1, \ldots, n$, $j \neq k$. Two cases are considered in which the distribution of Z_{ij} is identifiable from

(1) The univariate distribution of ϵ_0 and

(2) The bivariate distribution of (ϵ_1, ϵ_2).

By applying the identifiability result from Hart and Cañette (2009), the error distribution is identifiable from the f_{ϵ_0} provided $n \geq 3$. However, when n is small, it can be difficult in practice to identify the true target density from the density of ϵ_0. This is a consequence of the ill-posed nature of the problem, which is illustrated in Figure 12.

Note that the residual densities shown in Figures 12, 13, 14, and 15 can easily be computed using Fourier inversion or simulation. That is, to compute the density of ϵ_0 using Fourier inversion, the characteristic function of ϵ_0 given by

$$\psi_{\epsilon_0}(t) = \psi_Z(t)\psi_Z^{n-1}(-t/(n-1))$$

can be used to obtain f_{ϵ_0} via the inversion formula

$$f_{\epsilon_0}(x) = \frac{1}{2\pi}\int_{-\infty}^{\infty} \exp\{-itx\}\psi_{\epsilon_0}(t)dt.$$

Furthermore, to compute the density of ϵ_0 using simulation, first generate as many realizations from f_Z as is computationally feasible. A kernel density estimate can then be computed from the residuals $\epsilon_{ij0} = Z_{ij} - Z_{i(j)}$, $i = 1, \ldots, p$, $j = 1, \ldots, n$, which are calculated using the simulated data from f_Z. Provided that enough data is generated, the resulting density estimate can be taken as the true residual density for practical purposes.

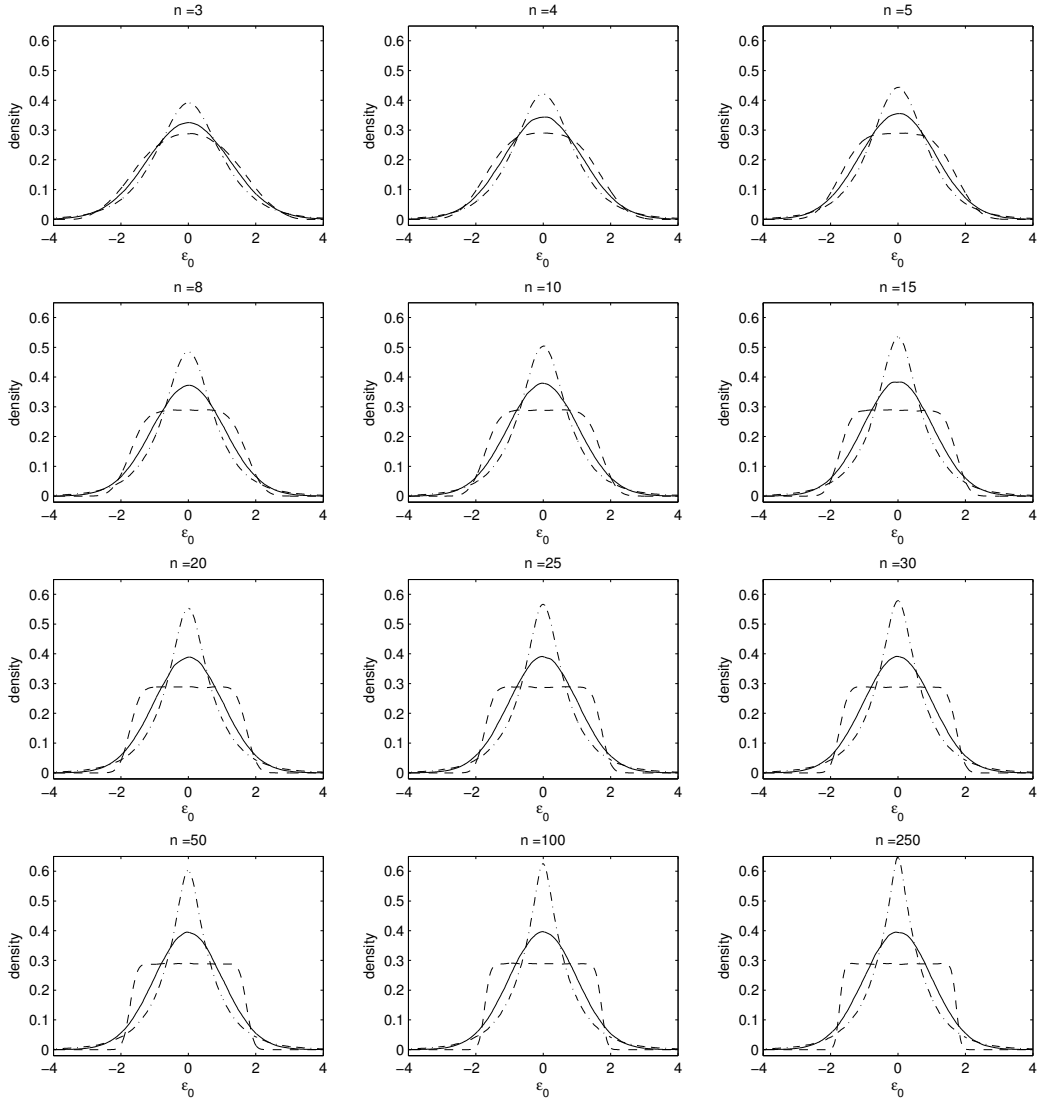

Figure 12. Residual densities for different values of n when error density is N(0,1) (solid), Laplace$(0, \sqrt{1/2})$ *(dotted-dash) and Unif*$(-\sqrt{3}, \sqrt{3})$ *(dashed).*

As can be seen from Figure 12, when $n = 3$, there is little difference between the density of ϵ_0 when the underlying error density f_Z is a normal, Laplace, or uniform density. However, as n increases, the residual density f_{ϵ_0} begins to look more like the error density since $\bar{Z}_{i(j)} \to 0$ as $n \to \infty$. Figures 13, 14, and 15 show the density of ϵ_0 in comparison with the underlying error density for different values of n. As before,

the normal, Laplace, and uniform error densities are used.

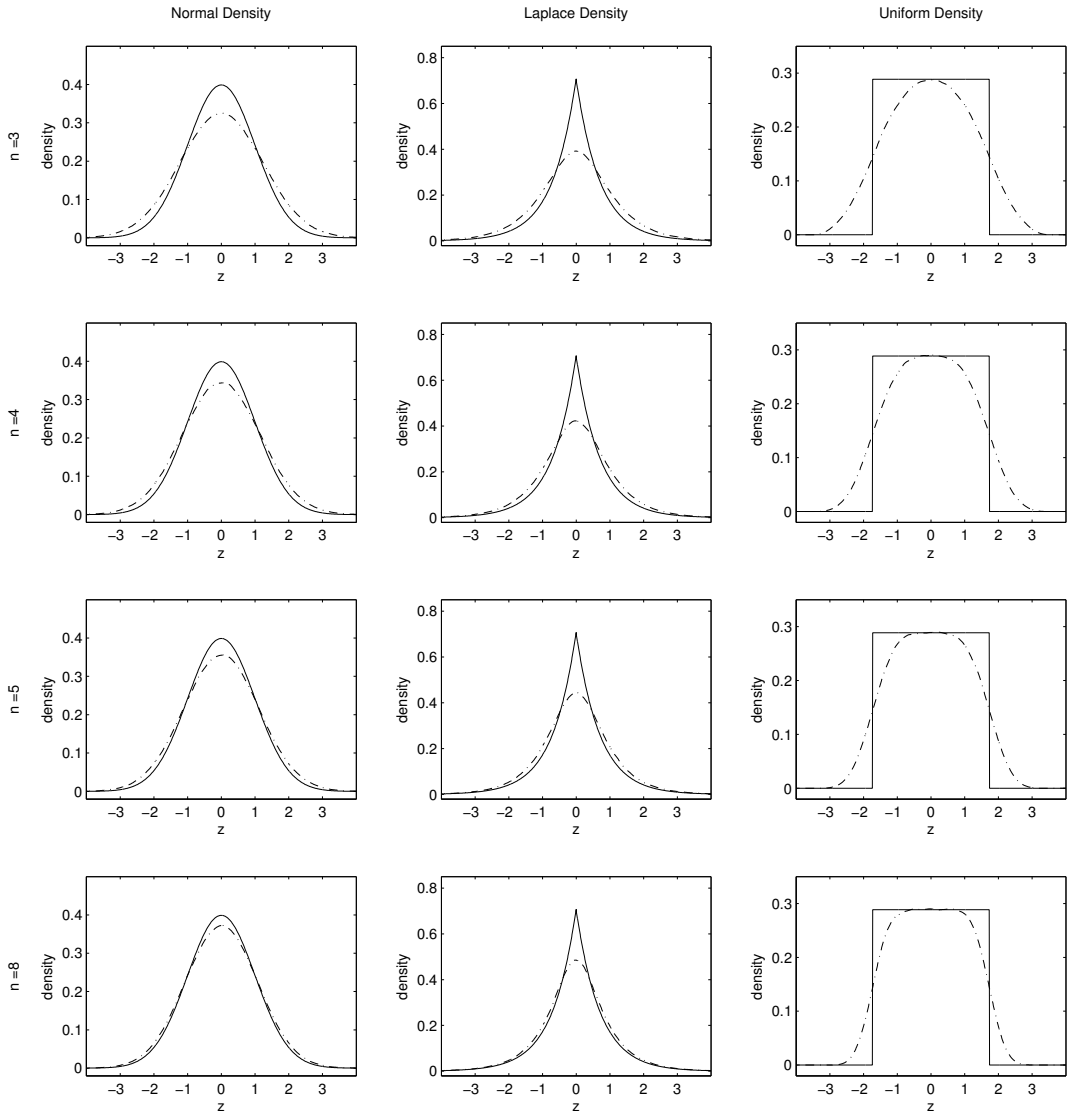

Figure 13. Residual densities for n = 3, 4, 5, and 8 compared with their respective underlying error density.

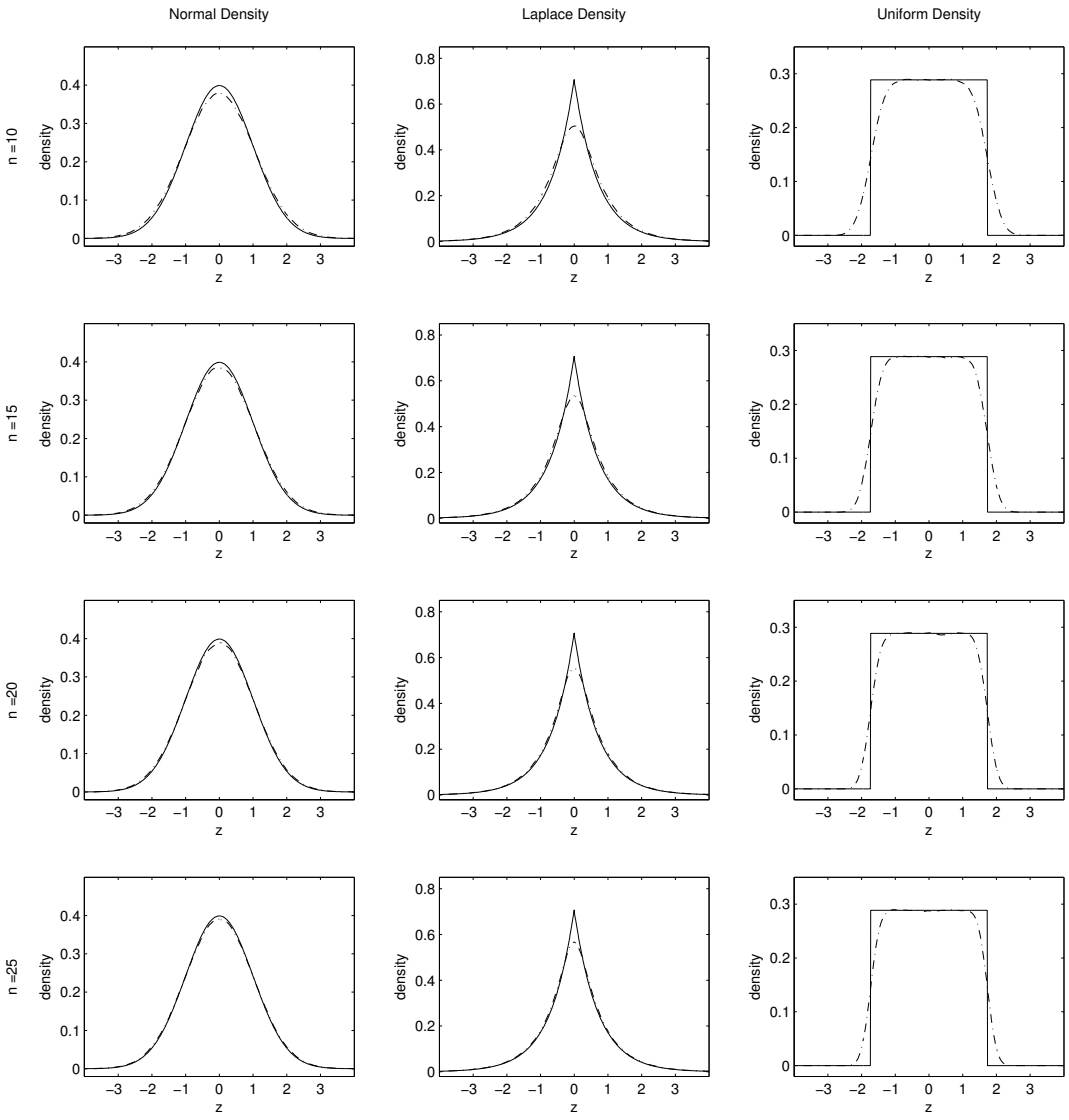

Figure 14. Residual densities for n = 10, 15, 20, and 25 compared with their respective underlying error density.

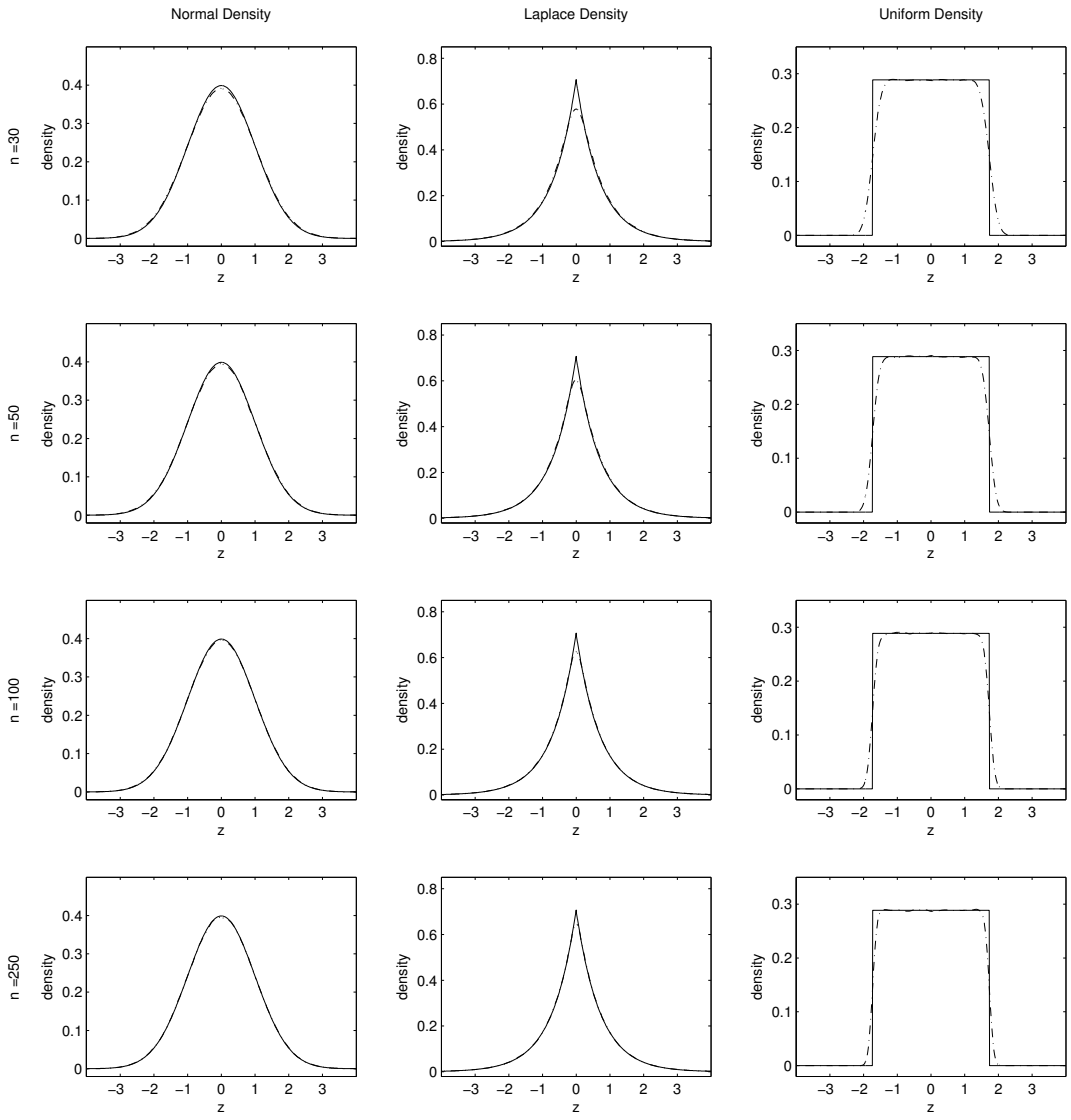

Figure 15. Residual densities for n = 30, 50, 100, and 250 compared with their respective underlying error density.

The value of n for which the residual density f_{ϵ_0} begins to look like the error density depends on the particular density. For example, when the underlying error density is normal, the residual density looks very similar to the error density when $n = 4$ or 5. On the other hand, the value of n must be somewhat larger for the residual density to adequately capture the spike of the Laplace density and the discontinuities of the uniform density.

By applying the result from Reiersøl (1950), the distribution of Z_{ij} can be identified from the bivariate distribution of (ϵ_1, ϵ_2) as long as $n \geq 3$. This is because the residuals ϵ_{ijk1} and ϵ_{ijk2}, where $i = 1, \ldots, p,\ j = 1, \ldots, n,\ k = 1, \ldots, n,\ j \neq k$, are a special case of the LRE model. That is, the common variable $\bar{Z}_{i(jk)}$ can be viewed as α_i, while Z_{ij} and Z_{ik} are the error terms. Furthermore, it is also possible to identify the distributions of α_i and Z_{ij} from the bivariate distribution of $(X_i, X_j),\ i = 1, \ldots, n,\ j = 1, \ldots, n,\ i \neq j$, when $n \geq 2$. Attention is restricted to using (ϵ_1, ϵ_2) since estimation of f_Z is of primary interest.

A natural question then becomes "Is there an advantage to using the distribution of ϵ_0 versus (ϵ_1, ϵ_2)?" Intuition says that a bivariate distribution is harder to estimate than a univariate one, but there may be some payoff in terms of identifiability of the error distribution when using a bivariate distribution. While this question is beyond the scope of this dissertation, it is an interesting one that remains open.

3.3 Approximating Minimum Distance Estimate Using Normal Mixtures

The methodology for the LRE model is analogous to that of the measurement error model. The density function of the error f_Z is approximated by a normal mixture with L components centered at $\mu_1, \ldots, \mu_L$ having common scale parameter

σ. That is, f_Z is assumed to be well approximated by

$$\widetilde{f}_Z(z) = \sum_{j=1}^{L} p_j \phi_j \left(\frac{z - \mu_j}{\sigma} \right), \tag{3.2}$$

where ϕ_j denotes a standard normal density, and p_j denotes the mixing proportion of ϕ_j. The means $\mu_1, \ldots, \mu_L$ are assumed to be equally spaced. Instead of computing the distance metric D based on the observed data as in the measurement error model, ϵ_0 and (ϵ_1, ϵ_2) are calculated according to the definitions in the previous section. This computed data can then be used in the same way as the observed data in the measurement error model to infer the target density f_Z.

3.3.1 Minimum Distance Using Univariate Distribution of ϵ_0

Recall that ϵ_{ij0} was previously defined as

$$\epsilon_{ij0} = X_{ij} - \bar{X}_{i(j)},$$

where $\bar{X}_{i(j)} = \frac{1}{n-1}(\sum_{k=1}^{n} X_{ik} - X_{ij})$, $i = 1, \ldots, p, \quad j = 1, \ldots, n$. A standard kernel density estimate $\hat{f}_{\epsilon_0}(x)$ is computed from ϵ_{ij0}, $i = 1, \ldots, p$, $j = 1, \ldots, n$, which is given by

$$\hat{f}_{\epsilon_0}(x) = \frac{1}{nph} \sum_{i=1}^{p} \sum_{j=1}^{n} \phi \left(\frac{x - \epsilon_{ij0}}{h} \right),$$

where ϕ is a standard normal density, and h is the bandwidth. The characteristic function corresponding to $\hat{f}_{\epsilon_0}(x)$ is then

$$\hat{\psi}_{\epsilon_0}(t) = \exp\{-h^2 t^2/2\} \hat{\psi}(t), \tag{3.3}$$

where

$$\hat{\psi}(t) = \frac{1}{np} \sum_{j=1}^{p} \sum_{k=1}^{n} \exp\{\epsilon_{jk0} it\}$$

is the empirical characteristic function of ϵ_{ij0}, $i = 1, \ldots, p$, $j = 1, \ldots, n$. Note that $\hat{\psi}(t)$ is a $\sqrt{p}$ consistent estimator of $\psi(t)$. See Feuerverger and Mureika (1977) for more information on convergence of empirical characteristic functions. The characteristic function corresponding to the induced density $\widetilde{f}_{\epsilon_0}(x)$ from assuming a normal mixture representation of f_Z is

$$\widetilde{\psi}_{\epsilon_0}(t) = \widetilde{\psi}_Z(t)\widetilde{\psi}_Z^{n-1}(-t/(n-1)), \tag{3.4}$$

where

$$\widetilde{\psi}_Z(t) = \sum_{j=1}^{L} p_j \exp\{-\sigma^2 t^2/2\} \exp\{\mu_j i t\}$$

is the characteristic function of the mixture of normals. By using the expressions in Equations (3.3) and (3.4), the discrepancy metric $D(\hat{f}_{\epsilon_0}, \widetilde{f}_{\epsilon_0})$ can be written as

$$D^2(\hat{f}_{\epsilon_0}, \widetilde{f}_{\epsilon_0}) = \frac{1}{2\pi} \int_{-\infty}^{\infty} |\hat{\psi}_{\epsilon_0}(t) - \widetilde{\psi}_{\epsilon_0}(t)|^2 dt.$$

Note that this expression does not require truncation since both characteristic functions in the integrand involve a normal mixture. However, because this integral can be algebraically complicated, truncating this expression in a way similar to Equation (2.9) may also be appropriate.

3.3.2 Minimum Distance Using Bivariate Distribution of (ϵ_1, ϵ_2)

Recall that ϵ_{ijk1} and ϵ_{ijk2} are defined by

$$\epsilon_{ijk1} = X_{ij} - \bar{X}_{i(jk)} \quad \text{and} \quad \epsilon_{ijk2} = X_{ik} - \bar{X}_{i(jk)},$$

where $\bar{X}_{i(jk)} = \frac{1}{n-2}(\sum_{\ell=1}^{n} X_{i\ell} - X_{ij} - X_{ik})$, $i = 1, \ldots, p$, $j = 1, \ldots, n$, $k = 1, \ldots, n$, $j \neq k$. A bivariate kernel density estimate $\hat{f}_{(\epsilon_1,\epsilon_2)}(x, y)$ is computed from $(\epsilon_{ijk1}, \epsilon_{ijk2})$,

$i = 1, \ldots, p,\ j = 1, \ldots, n,\ k = 1, \ldots, n,\ j \neq k$, which is given by

$$\hat{f}_{(\epsilon_1,\epsilon_2)}(x, y) = \frac{1}{pn(n-1)h^2} \sum_{i=1}^{p} \sum_{j=1}^{n} \sum_{k \neq j} \phi\left(\frac{x - \epsilon_{ijk1}}{h}\right) \phi\left(\frac{y - \epsilon_{ijk2}}{h}\right),$$

where ϕ is a standard normal density, and h is the bandwidth. The characteristic function corresponding to $\hat{f}_{(\epsilon_1,\epsilon_2)}(x, y)$ is

$$\hat{\psi}_{(\epsilon_1,\epsilon_2)}(s, t) = \exp\{-h^2(s^2 + t^2)/2\}\hat{\psi}(s, t), \tag{3.5}$$

where

$$\hat{\psi}(s, t) = \frac{1}{pn(n-1)} \sum_{j=1}^{p} \sum_{k=1}^{n} \sum_{\ell \neq k} \exp\{\epsilon_{jk\ell 1} is + \epsilon_{jk\ell 2} it\}$$

is the empirical characteristic function of $(\epsilon_{ijk1}, \epsilon_{ijk2}),\ i = 1, \ldots, p,\ j = 1, \ldots, n,\ k = 1, \ldots, n,\ j \neq k$. As mentioned in the previous section, $\hat{\psi}(s, t)$ is a $\sqrt{p}$ consistent estimator of $\psi(s, t)$. Furthermore, the characteristic function corresponding to the induced density $\widetilde{f}_{(\epsilon_1,\epsilon_2)}(x, y)$ is

$$\widetilde{\psi}_{(\epsilon_1,\epsilon_2)}(s, t) = \widetilde{\psi}_Z(s)\widetilde{\psi}_Z(t)\widetilde{\psi}_Z^{n-2}(-(s + t)/(n - 2)), \tag{3.6}$$

where

$$\widetilde{\psi}_Z(t) = \sum_{j=1}^{L} p_j \exp\{-\sigma^2 t^2/2\} \exp\{\mu_j it\}$$

is the characteristic function of the mixture of normals. By using the expressions in Equations (3.5) and (3.6), the discrepancy metric $D(\hat{f}_{(\epsilon_1,\epsilon_2)}, \widetilde{f}_{(\epsilon_1,\epsilon_2)})$ can be written as

$$D^2(\hat{f}_{(\epsilon_1,\epsilon_2)}, \widetilde{f}_{(\epsilon_1,\epsilon_2)}) = \frac{1}{4\pi^2} \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} |\hat{\psi}_{(\epsilon_1,\epsilon_2)}(s, t) - \widetilde{\psi}_{(\epsilon_1,\epsilon_2)}(s, t)|^2 ds dt.$$

3.3.3 Estimating Mixing Proportions

The mixing proportions are estimated in the LRE model in a way similar to that used for the measurement error model. Since it is not possible to explicitly solve for

the mixing proportions $\mathbf{p}$ while enforcing $\sum_{j=1}^{L} p_j = 1$ and $p_j \geq 0$, $j = 1, \ldots, L$, $\mathbf{p}$ is approximated numerically using an optimization routine. Unlike in the measurement error model, it is assumed that the random variable Z_{ij} has mean 0 and standard deviation 1. In principle, these constraints should also be enforced when estimating $\mathbf{p}$. While it is relatively easy to enforce the linear constraint that $\sum_{j=1}^{L} p_j \mu_j = 0$, it is sometimes computationally problematic to enforce the nonlinear constraint that the variance equals 1. However, enforcing this variance constraint does not seem to affect the resulting estimator. As a result, this constraint is not enforced when estimating $\mathbf{p}$.

3.3.4 Estimating σ and Number of Components L

The method for estimating σ in the LRE model is the same as in the measurement error model. That is, cross-validation is used to estimate σ at a large fixed value of L. Once this is done, the number and location of the mixture components can be chosen as described below.

To select the means of the normal mixture components μ, a rough estimate for the support of Z_{ij} must be obtained. Thus, it is necessary to find an interval that covers most of the mass of the probability density function of Z_{ij}. Since it is known that the variance of Z_{ij} is 1, a conservative way of doing this is by making use of Chebyshev's inequality. Recall that Chebyshev's inequality says that

$$P(|Z - \mu_Z| > M) \leq \frac{\sigma_Z^2}{M^2} \tag{3.7}$$

for any real number $M > 0$, and so it puts an upper bound on the mass of f_Z in the tails of the distribution. Since $\sigma_Z = 1$ and $\mu_Z = 0$ by assumption, some reasonable value for M, such as $M = 5$ or 8, can be selected. This would insure that $P(|Z| > 5)$

and $P(|Z| > 8)$ are at most 1/25 and 1/64, respectively. However, since Equation (3.7) is conservative, these probabilities are likely much smaller. As previously discussed, there is essentially no penalty in the algorithm for placing the normals on too wide an interval since the algorithm assigns identically zero or near zero mixing proportions to normal mixture components whose locations are too extreme.

For a fixed L, one way to select the location of the components is

$$\mu_j = -M + j\left(\frac{2M}{L+1}\right), \quad j = 1, \ldots, L. \tag{3.8}$$

As before, if L^* normals are sufficient to represent f_Z, then for a given σ, the estimate of f_Z that results from using L^* normal components does not appreciably change as the number of mixture components increase beyond this point. Once a good value of σ is selected in the previous step, the number of mixture components can then be chosen by looking at the value of the metric computed for increasing values of L. The number of normals at which the value of the metric stops changing by some specified amount (i.e., two or three percent) is selected as the number of components to use for the mixture.

A main difference between the LRE model and the measurement error model is that the target density in the LRE model is assumed to have mean 0. However, by defining the locations of the mixture components according to Equation (3.8), there is not a mixture component centered at 0 when L is even. This can be an inconvenience, especially when the target density is symmetric and unimodal. As a result, the algorithm will often favor a larger L than is necessary. A way around this problem is to either devise an alternative method for placing the mixture components or use the scheme in Equation (3.8) for odd L only. The latter method will be used in the simulations and real data example.

3.3.5 Outline for Showing Consistency of Estimator

A rigorous mathematical proof of consistency for the minimum distance estimator in the LRE model will not be pursued in this dissertation. However, the proof for the measurement error model in Appendix A only needs to be modified in the parts that make use of the particular structure of the distance metric. Otherwise, the argument remains the same. That is, it must be shown that as $n \to \infty$ for a fixed L and σ, the estimator approaches some approximation to the target density. It must then hold that the approximation approaches the target density as $L \to \infty$ and $\sigma \to 0$. A "ratchet" argument can then be used to say that as n and L go to infinity and σ goes to 0, the estimator approaches the true target density. As before, the convergence argument applies to characteristic functions, so conditions for which this implies convergence of density functions must be addressed.

3.4 Simulated Examples

Simulations using the minimum distance methodology are implemented to assess its performance against other methods. In particular, the minimum distance method is compared to the explicit characteristic function inversion method from Hall and Yao (2003). The normal and exponential densities are chosen since these are used as part of the simulation study in Hall and Yao. In addition, the chi-square and normal mixture distributions are chosen to see how the method performs for skewed and bimodal data, respectively. The method is performed using $n = 3$ and $p = 100$, 250, 500, and 1000. Although Hall and Yao's method only requires $n \geq 2$, this value of n is used because it is the minimum value such that the residual distribution can identify the error distribution under the assumptions previously discussed. The four error densities considered for f_Z are shown in Figure 16 and defined by

(1) Density 1: $Z \sim N(0,1)$;

(2) Density 2: $Z \sim \chi^2(3)$;

(3) Density 3: $Z \sim \exp(1)$;

(4) Density 4: $Z \sim \frac{1}{2}N(-2,1) + \frac{1}{2}N(2,1)$.

Densities 2 and 3 are shifted to have mean 0. Also, densities 2 and 4 do not have a unit variance. Even though this is an assumption of the underlying LRE model, it does not present a problem since any density can be scaled to have a variance 1. After estimation, the density estimate can be transformed back to its original scale.

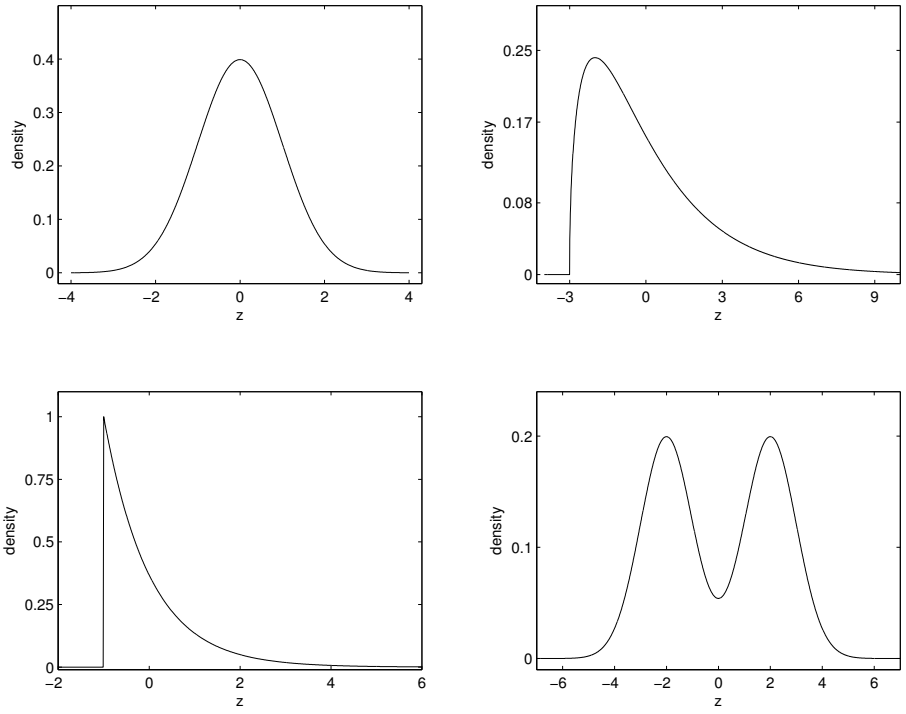

Figure 16. Target error densities of increasing difficulty to estimate; density 1 (top left), density 2 (top right), density 3 (bottom left), density 4 (bottom right).

To evaluate the performance of the estimator, 100 samples are generated for each of the four error densities. Note that residuals under the LRE model when using the minimum distance method do not depend on the distribution of the random effects because of the differencing when computing the residuals. However, Hall and Yao's method is not a location free method, and so data from the random effects distribution must also be generated. Thus, following Hall and Yao, data from a $N(0,1)$ distribution are used for the random effects. After data is generated from the error and random effects distributions, LRE data is generated according to $X_{ij} = \alpha_i + \gamma Z_{ij}$, $i = 1, \ldots, p$, $j = 1, 2, 3$. An estimate of the error density is then computed using Hall and Yao's explicit characteristic function inversion method for each sample. This is used as a basis for comparing the performance of the minimum distance method using both the univariate and bivariate distribution of residuals. The density estimate $\hat{f}_Z$ is computed for each sample and compared with the true density using the ISE criterion given by

$$ISE(\hat{f}_Z, f_Z) = \int (\hat{f}_Z(z) - f_Z(z))^2 dz.$$

3.4.1 Hall and Yao's Explicit Characteristic Function Inversion Method

Table 5 summarizes the ISE when estimating the four densities previously described using the method by Hall and Yao. As in the simulation results for the measurement error model, the medians and interquartile ranges of the 100 values of ISE are reported for each target density and sample size combination.

Table 5. Median ISE of density estimates using Hall and Yao's method. The numbers in parentheses are interquartile ranges of ISE for 100 replications.

	Target Density			
p	*1* - $N(0,1)$	*2* - $\chi^2(3)$	*3* - $\exp(1)$	*4* - *Normal Mix*
100	0.2452 (0.0412)	0.0906 (0.0219)	0.3991 (0.2710)	0.0978 (0.0104)
250	0.0727 (0.0106)	0.0843 (0.0122)	0.3910 (0.1801)	0.0970 (0.0100)
500	0.0563 (0.0042)	0.0853 (0.0094)	0.3843 (0.1529)	0.0970 (0.0104)
1000	0.0427 (0.0038)	0.0767 (0.0062)	0.2644 (0.0934)	0.0713 (0.0056)

In terms of ISE, clearly the exponential density is very difficult to estimate. As pointed out by Hall and Yao, this is because the estimator is constrained to be a smooth curve. While the estimates are able to capture skewness when estimating the exponential density, the discontinuity makes the problem difficult. Figures 17, 18, 19, and 20 show the resulting estimates for each of the four target densities.

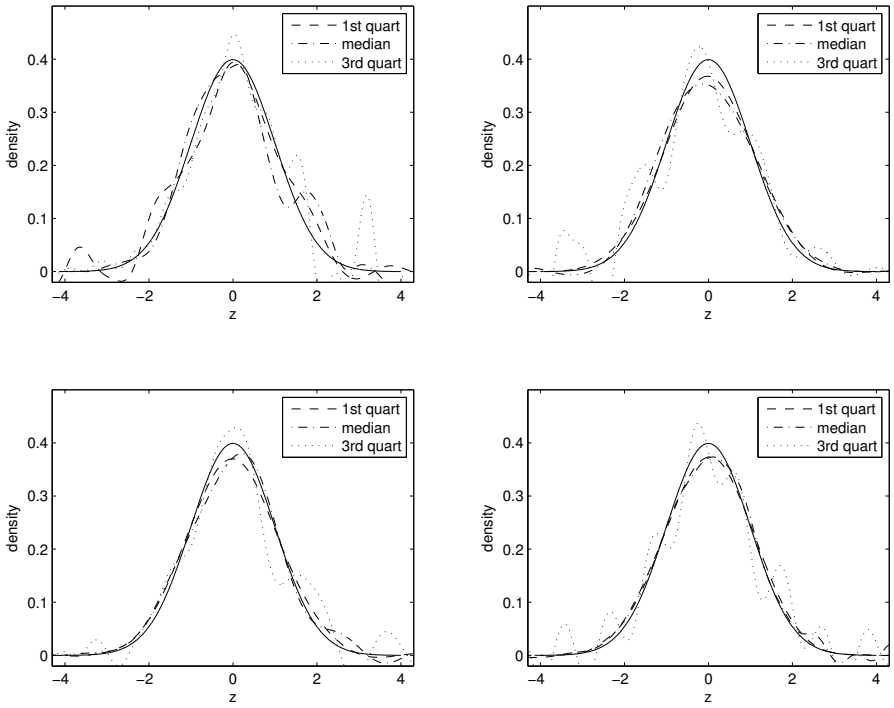

Figure 17. Density estimates for N(0,1) error density (density 1) using Hall and Yao's method at sample size p=100 (top left), p=250 (top right), p=500 (bottom left), p=1000 (bottom right).

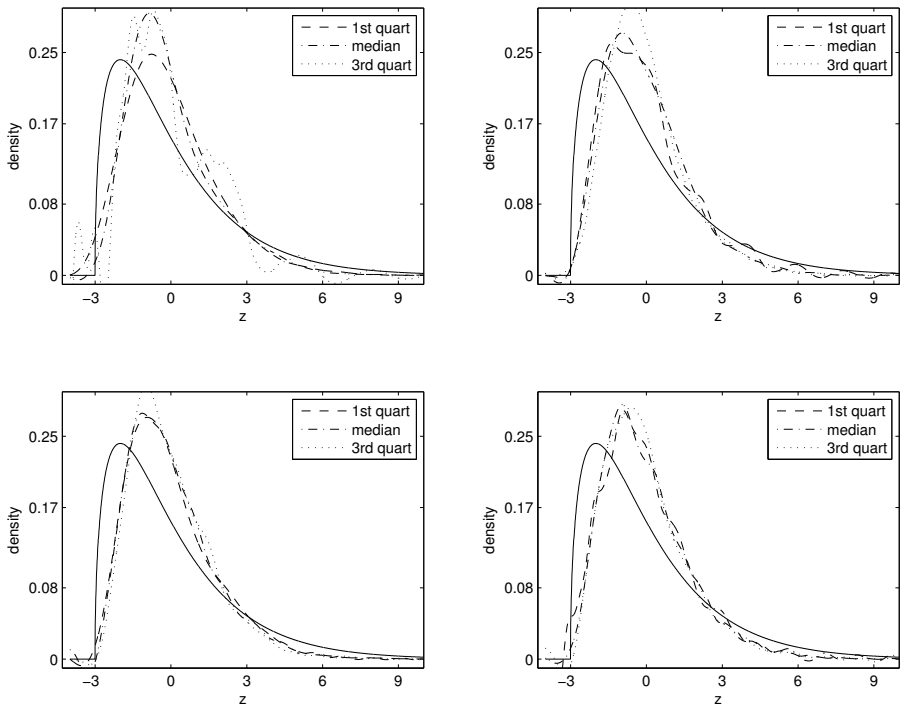

Figure 18. Density estimates for $\chi^2(3)$ error density (density 2) using Hall and Yao's method at sample size p=100 (top left), p=250 (top right), p=500 (bottom left), p=1000 (bottom right).

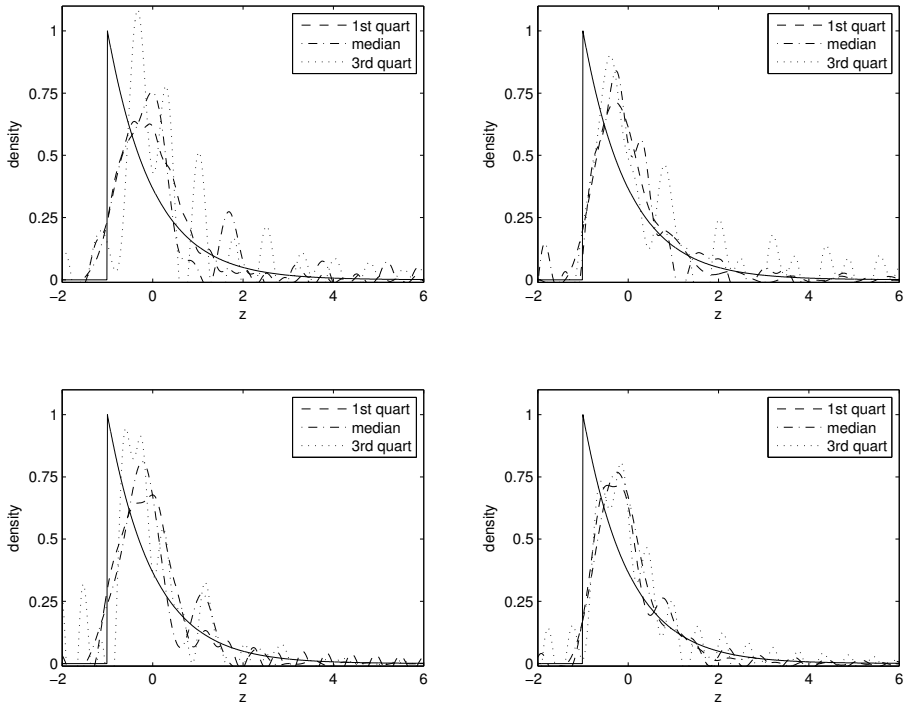

Figure 19. Density estimates for exp(1) error density (density 3) using Hall and Yao's method at sample size p=100 (top left), p=250 (top right), p=500 (bottom left), p=1000 (bottom right).

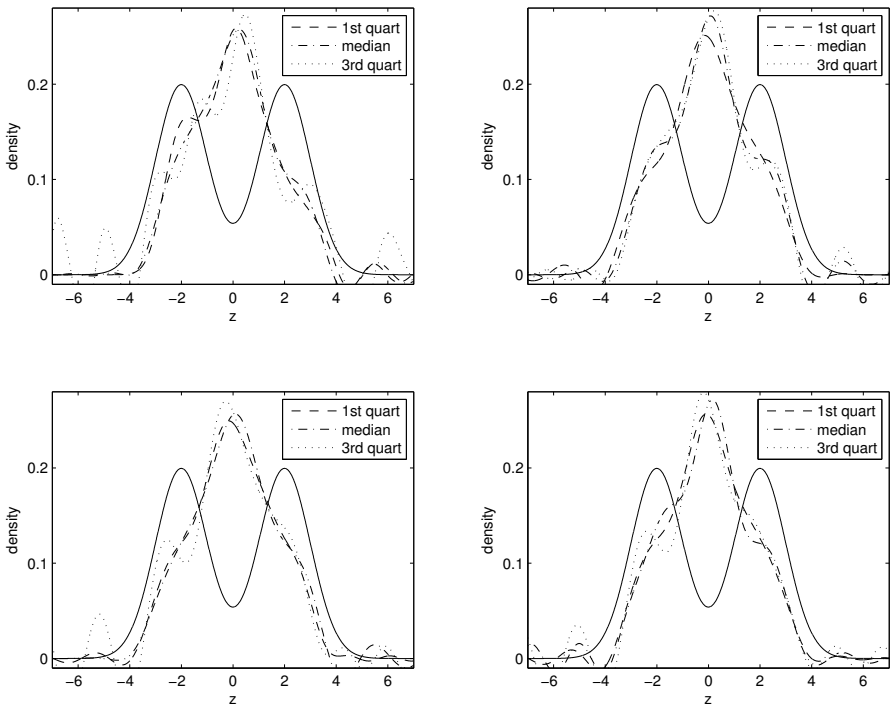

Figure 20. Density estimates for normal mixture error density (density 4) using Hall and Yao's method at sample size p=100 (top left), p=250 (top right), p=500 (bottom left), p=1000 (bottom right).

As is evident in these plots, one fundamental problem with Hall and Yao's estimator is that there can be substantial oscillations in the estimated density function itself. This is in general a common feature of density estimators obtained through a Fourier inversion method, and in some cases these effects can result in a density estimate that takes negative values. While their estimator performs reasonably well for the normal and chi-square target densities, estimation of the exponential and normal mixture densities is problematic. In both the chi-square and exponential densities, the estimator can capture skewness of the target density reasonably well. However, the estimator is poor at capturing the bimodal features of the normal mixture. Overall, the results indicate that the method may be useful when estimating smooth unimodal densities that are symmetric or possess mild skewness. The method may be too crude or impractical when it is suspected that the target density may be discontinuous or multimodal.

3.4.2 Minimum Distance Method Using a Normal Mixture

Previously, it was assumed that Z_{ij} has variance 1. However, the $\chi^2(3)$ and normal mixture density do not have unit variance. Thus, in the actual implementation of the minimum distance methodology, the residuals are first divided by an estimate of the scale parameter γ. Minimum distance is then applied to the scaled data with variance 1. The range for the μ_j's is selected by setting $M = 5$ to insure that $P(|Z| > 5) \leq 1/25$, which holds by Chebyshev's inequality. The location of the mixture components are evenly spaced on this interval by defining

$$\mu_j = -5 + j\left(\frac{10}{L+1}\right), \quad j = 1, \ldots L$$

according to Equation (3.8). The mixing proportions $\mathbf{p}$, common scale σ, and number of components L, are estimated by the methods previously described. After a density

estimate is obtained, the density estimate is then rescaled by $\hat{\gamma}$. Tables 6 and 7 summarize the ISE when estimating the four densities using the univariate and bivariate methods, respectively. The medians and interquartile ranges of the 100 values of ISE are reported for each target density and sample size combination.

Table 6. Median ISE of minimum distance estimates for error densities using univariate distribution of residuals. The numbers in parentheses are interquartile ranges of ISE for 100 replications.

	Target Density			
p	*1* - $N(0,1)$	*2* - $\chi^2(3)$	*3* - $\exp(1)$	*4* - *Normal Mix*
100	0.0186 (0.0134)	0.0827 (0.0251)	0.3079 (0.0181)	0.0276 (0.0080)
250	0.0173 (0.0141)	0.0731 (0.0214)	0.2813 (0.0145)	0.0234 (0.0090)
500	0.0158 (0.0112)	0.0676 (0.0194)	0.2771 (0.0132)	0.0207 (0.0057)
1000	0.0123 (0.0108)	0.0592 (0.0188)	0.2548 (0.0142)	0.0181 (0.0048)

From the tables, it is evident that there is no real advantage in this model to using the bivariate versus univariate distribution. A disadvantage to using the bivariate distribution is that the algorithm requires more computational time to minimize the metric and compute the bandwidth of the bivariate kernel density estimate. See Sain et al. (1994) for a discussion of bandwidth selection methods for multivariate densities. It may be the case that when using a higher dimensional distribution, more error distributions become identifiable for a given model. However, it remains an open question as to whether or not there is an advantage to using an observable distribution of higher dimension.

Table 7. Median ISE of minimum distance estimates for error densities using bivariate distribution of residuals. The numbers in parentheses are interquartile ranges of ISE for 100 replications.

	Target Density			
p	*1* - $N(0,1)$	*2* - $\chi^2(3)$	*3* - $\exp(1)$	*4* - *Normal Mix*
100	0.0192 (0.0122)	0.0791 (0.0281)	0.3295 (0.0173)	0.0256 (0.0099)
250	0.0179 (0.0119)	0.0717 (0.0202)	0.3050 (0.0157)	0.0219 (0.0085)
500	0.0133 (0.0106)	0.0682 (0.0189)	0.2840 (0.0112)	0.0193 (0.0068)
1000	0.0107 (0.0101)	0.0572 (0.0163)	0.2441 (0.0109)	0.0185 (0.0051)

As with the method from Hall and Yao (2003), the exponential density appears to be difficult to estimate due to the marked discontinuity. In addition, the performance of the minimum distance estimator outperforms Hall and Yao's estimator for all four target densities in terms of median ISE. When compared with the Hall and Yao method, the interquartile ranges using the minimum distance estimator are larger for the normal and chi-square densities and smaller for the exponential and normal mixture densities. This phenomenon is also observed when considering the minimum distance estimator in the measurement error context. Figures 21, 22, 23, and 24 show the resulting estimates for each of the four target densities.

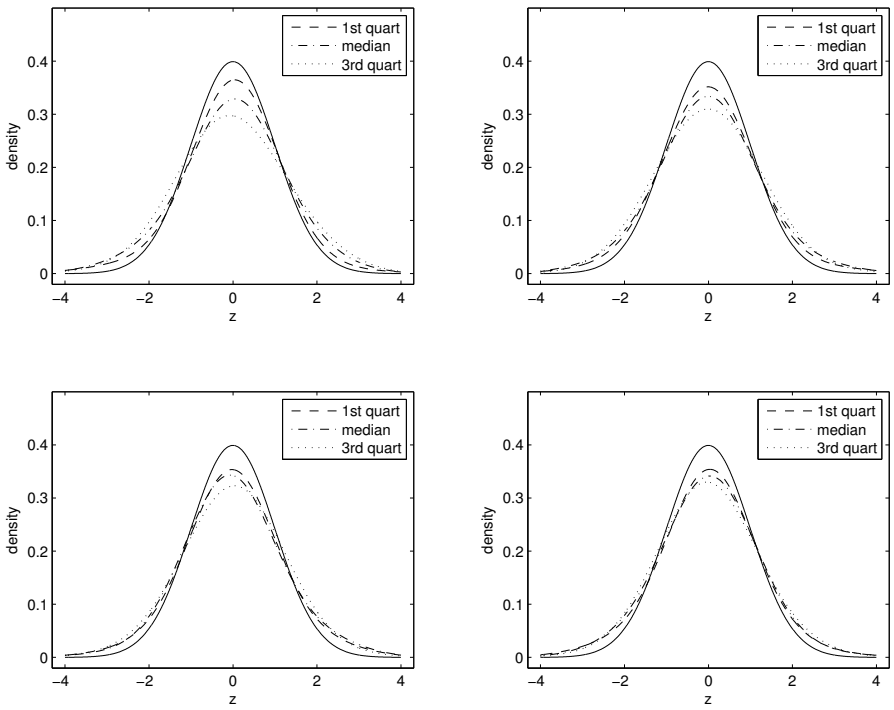

Figure 21. Minimum distance estimates for N(0,1) error density (density 1) at sample size p=100 (top left), p=250 (top right), p=500 (bottom left), p=1000 (bottom right).

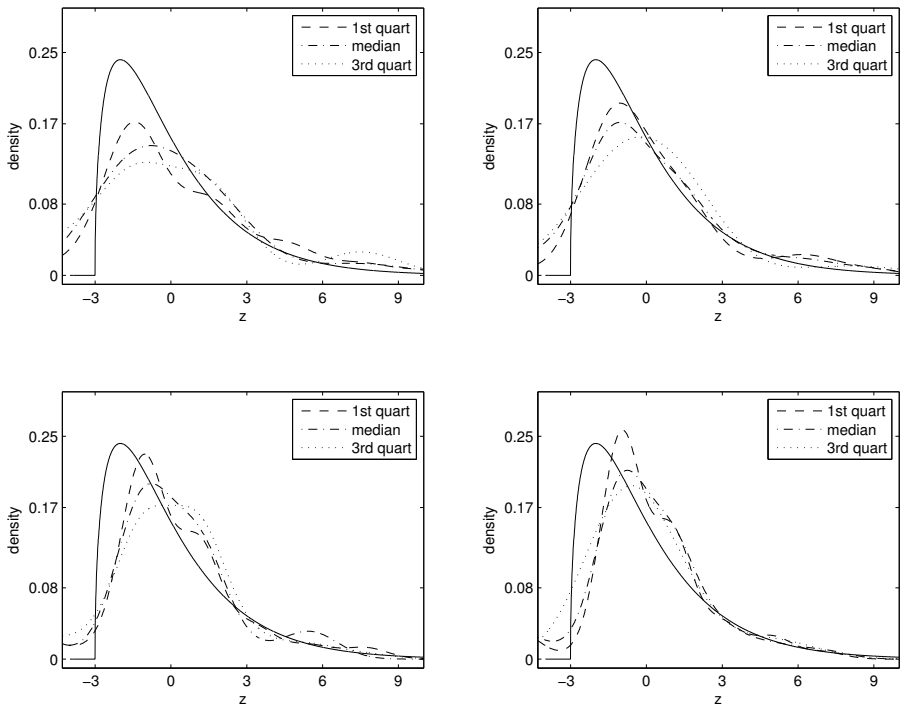

Figure 22. Minimum distance estimates for $\chi^2(3)$ error density (density 2) at sample size p=100 (top left), p=250 (top right), p=500 (bottom left), p=1000 (bottom right).

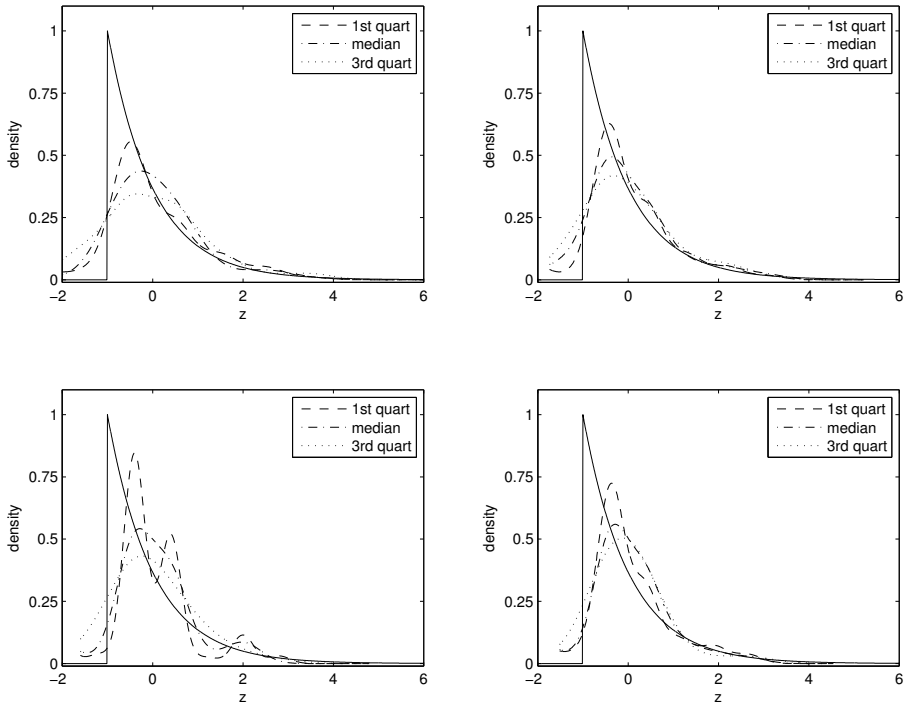

Figure 23. Minimum distance estimates for exp(1) error density (density 3) at sample size p=100 (top left), p=250 (top right), p=500 (bottom left), p=1000 (bottom right).

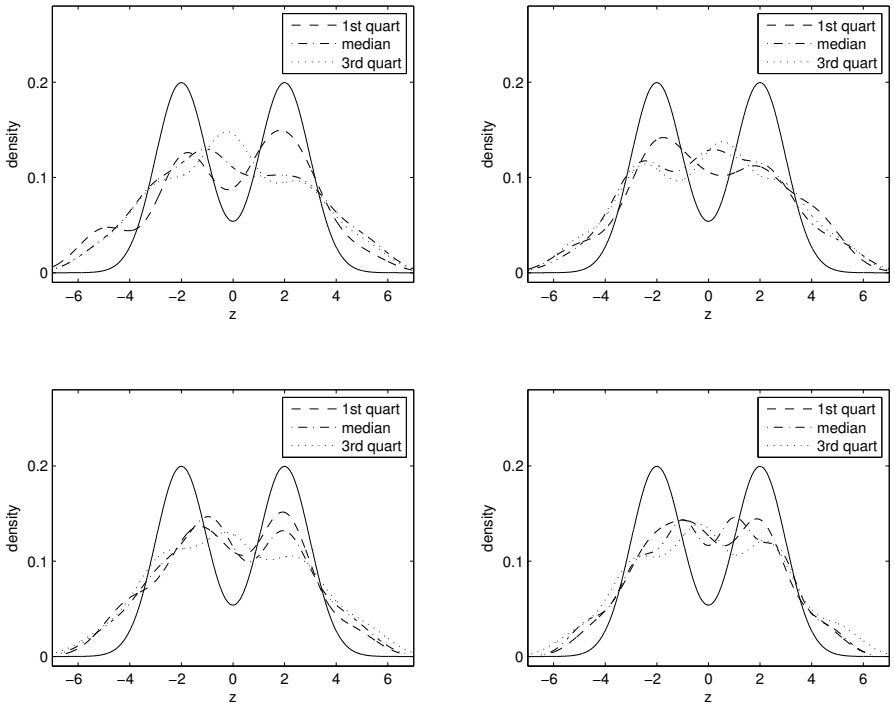

Figure 24. Minimum distance estimates for normal mixture error density (density 4) at sample size p=100 (top left), p=250 (top right), p=500 (bottom left), p=1000 (bottom right).

From the previous figures, it is clear that the estimator tends to smooth out the peaks of the target density. This is especially apparent when estimating the density that is the mixture of two normals. However, most of the estimates detect the bimodal features of the normal mixture fairly well for larger sample sizes, although the modes are not as pronounced as they should be. In addition, the estimator does reasonably well at estimating the normal density and at capturing the skewness of the chi-square and exponential densities.

Interestingly, the same problems do not arise for the LRE model as in the measurement error model from Chapter II where it is often the case that a poor estimate results from a small choice of σ. While a small value of σ does occasionally occur in the simulations, it is more often the case that an oversmoothed estimate occurs.

Overall, the minimum distance approach produces better estimates when compared with the explicit characteristic function inversion method from Hall and Yao. Their method tends to produce density estimates that have oscillations, which sometimes result in negative values for the density estimate. In addition, the estimator erroneously gives one mode when estimating the mixture of two normals. While the minimum distance estimator tends to smooth out the peaks of the target density, it does not seem to suffer from the drawbacks of Hall and Yao's method. For one, the estimator always gives a proper density estimate since the values of the density estimate are always nonnegative. Furthermore, the minimum distance approach in this model does not suffer from oscillations in the density estimates that are common to estimators that use Fourier inversion techniques. Minimum distance is better able to capture the general features of the underlying density, which can readily be seen when estimating the mixture of two normals.

3.5 Extension to Location-Scale Random Effects Model

A possible extension to the LRE model is one in which the scale is allowed to vary between small data sets. That is, the location-scale random effects (LSRE) model is defined as

$$X_{ij} = \alpha_i + \gamma_i Z_{ij}, \quad i = 1, \ldots, p, \quad j = 1, \ldots, n.$$

The random variable Z_{ij} is still assumed to come from an unknown probability density f_Z having mean 0 and standard deviation 1, and (α_i, γ_i) is assumed to come from an unknown bivariate density $f_{(\alpha,\gamma)}$. The variables Z_{ij} and (α_i, γ_i) are assumed to be independent. Hart and Cañette (2009) also consider this model by implementing their minimum distance estimator to nonparametrically estimate the distribution of the scale random effects and the error distribution. This model is both computationally and theoretically challenging. As with the LRE model, there are many transformations of the original data that can be useful for a variety of reasons. For example, in some situations, it is useful to have an idea as to how the ordinary residuals defined by

$$\epsilon_{ij} = \frac{X_{ij} - \bar{X}_i}{s_i} = \frac{Z_{ij} - \bar{Z}_i}{s_{Zi}}, \quad i = 1, \ldots, p, \quad j = 1, \ldots, n \tag{3.9}$$

are distributed. In this expression, $\bar{X}_i$ and s_i are the sample mean and standard deviation, respectively. Furthermore, consider defining the residuals by

$$\epsilon^*_{ij} = \frac{X_{ij} - \bar{X}_{iJ}}{s_{iJ^*}} = \frac{Z_{ij} - \bar{Z}_J}{s_{ZJ^*}}, \quad i = 1, \ldots, p, \quad j = 1, \ldots, n, \tag{3.10}$$

where J and J^* are index sets such that $J \cap J^* = \emptyset$ and $J \cup J^* = \{1, \ldots, j-1, j+1, \ldots, n\}$. The sample mean $\bar{X}_{iJ}$ is the mean of $X_{ij},\ j \in J$, for the ith data set, and s_{iJ^*} is the standard deviation of $X_{iJ^*},\ j \in J^*$, also for the ith data set. The residuals defined in Equation (3.10) are analogous to ϵ_{ij0} defined in the LRE model. They have the advantage that the three terms X_{ij}, $\bar{X}_{iJ}$, and s_{iJ^*} are independent, which

is useful when trying to establish identifiability results or implementing the method in an efficient way. It may also be advantageous to consider a transformation such as $(X_{i1} - X_{i2})/(X_{i3} - X_{i4})$ for computational reasons when n is small. While there are many location and scale free transformations, minimum distance estimation in the measurement error model and LRE model is computationally much easier because of the way the metric is computed. Recall that in these two models, the metric is computed using characteristic functions instead of density functions. This is because there is an explicit form for the characteristic function of f_Y in the measurement error model and the univariate and bivariate densities f_{ϵ_0} and $f_{(\epsilon_1,\epsilon_2)}$ in the LRE model. In general, this is not the case for the LSRE model. Since the primary interest is to estimate only the error density f_Z, some transformation of the observed data that does not depend on the location and scale of each small data set must be found. The expression for the characteristic function of the observable data then usually always involves a ratio so that the quantities are free of the scale γ_i. As a result, the characteristic function is given by an integral equation and can be difficult to manipulate.

This computational challenge also leads to a theoretical one. Showing identifiability of the target density from the density of observable data (i.e., observed data in measurement error model or residuals in LRE model) often depends on having an explicit tractable form for the characteristic function of the observable data. Since the characteristic function of the residuals (or other quantities) can be complicated under the LSRE model, it is especially challenging to establish identifiability results. Intuitively, one might expect identification of the error density from the univariate residual density to hold under mild conditions when $n \geq 4$, since a similar result holds for $n \geq 3$ in the LRE model. While this is still an open question, identification

of the error density can be shown to hold when $n \geq 7$ from the bivariate density of

$$\left(\frac{X_{i1} - \bar{X}_{i(23)}}{X_{i4} - X_{i5}}, \frac{X_{i1} - \bar{X}_{i(23)}}{X_{i6} - X_{i7}} \right), \quad i = 1, \ldots, p,$$

where $\bar{X}_{i(23)} = \frac{1}{2}(X_{i2} + X_{i3})$. See Appendix B for a detailed proof. This can also be written as

$$\left(\frac{Z_{i1} - \bar{Z}_{i(23)}}{Z_{i4} - Z_{i5}}, \frac{Z_{i1} - \bar{Z}_{i(23)}}{Z_{i6} - Z_{i7}} \right), \quad i = 1, \ldots, p,$$

which does not depend on α_i or γ_i. The result holds provided that the characteristic functions of $Z_{i1} - \bar{Z}_{i(23)}$, $Z_{i4} - Z_{i5}$, and $Z_{i6} - Z_{i7}$ do not vanish.

3.6 Applications

For this example, the same microarray data that are used in Hart and Cañette (2009) are considered. The data were originally collected by Robert Chapkin and his coworkers from Texas A&M University and analyzed in Davidson et al. (2004). The data consist of the logarithm of the expression level for the ith gene and jth rat denoted by X_{ij}, $i = 1, \ldots, 8038,\ j = 1, \ldots, 5$. While these data are only a subset of the original data, it corresponds to the actual data used in Hart and Cañette (2009).

The LRE and LSRE models are both considered as potential models for the rat data. The difference between the two models is that in the LRE model, it is assumed that a common scale exists across all genes, while the LSRE model allows for there to be scale differences between genes. The residuals under the LRE model are computed by

$$\epsilon_{ij} = \frac{X_{ij} - \bar{X}_i}{s}, \quad i = 1, \ldots, 8038, \quad j = 1, \ldots, 5,$$

where $\bar{X}_i = \frac{1}{5}\sum_{j=1}^{5} X_{ij}$ and $s = \left(\frac{1}{8038(4)} \sum_{i=1}^{8038} \sum_{j=1}^{5} (X_{ij} - \bar{X}_i)^2 \right)^{1/2}$. The residuals

under the LSRE model are computed by

$$\epsilon_{ij} = \frac{X_{ij} - \bar{X}_i}{s_i}, \quad i = 1, \ldots, 8038, \qquad j = 1, \ldots, 5,$$

where $s_i = \left(\sum_{j=1}^{5}(X_{ij} - \bar{X}_i)^2\right)^{1/2}$, $i = 1, \ldots, 8038$. To get some idea as to how these residuals are distributed, kernel density estimates are computed for each set of residuals, which are shown in Figure 25. It can be seen that the residuals under the LRE model have heavier tails than the residuals under the LSRE model.

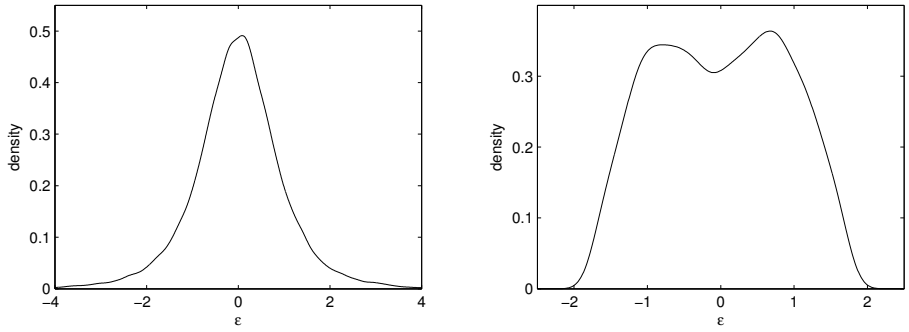

Figure 25. Kernel density estimates for residual density under LRE model (left) and LSRE model (right).

The univariate residual distribution under the LRE model is used in the minimum distance method to estimate the error density as previously described. Figure 26 shows estimates of the error density in the LRE model for different values of L.

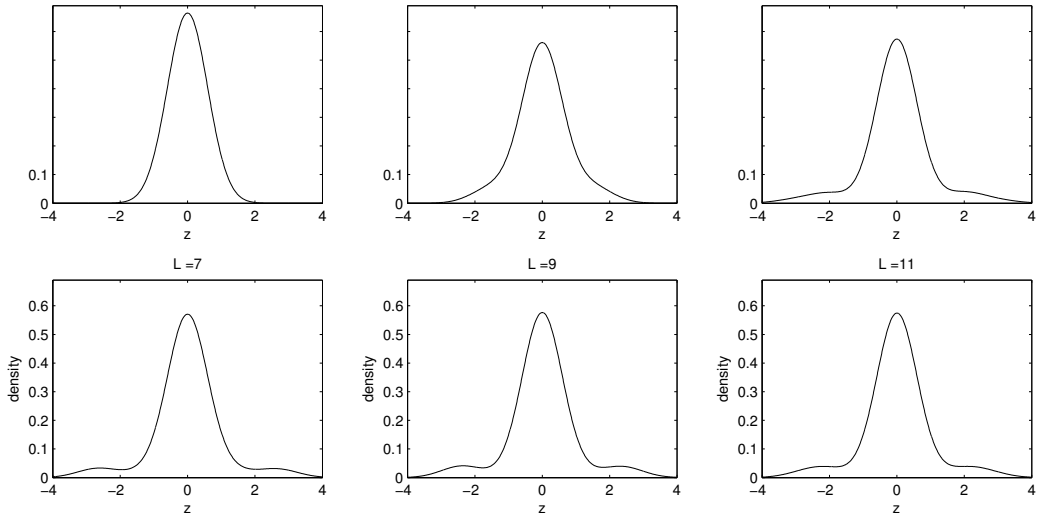

Figure 26. Minimum distance estimates of error density under LRE model when the normal mixture has L=1, 3, 5, 7, 9, and 11 components.

The estimate corresponding to $L = 7$ mixture components is selected as the density estimate of the error, although several of these components have a zero proportion assigned to them. The common scale parameter is estimated to be $\hat{\sigma} = 0.63$.

While the residual distribution is used for minimum distance in the LRE model, using the residuals in the LSRE model is problematic because of the difficulty in obtaining an expression for the characteristic function of $\epsilon_{ij} = \frac{X_{ij} - \bar{X}_i}{s_i}$. In addition, since $n = 5$, the identifiability result in the LSRE model that the error density is identifiable when $n \geq 7$ does not apply. While there are no known identifiability results pertaining to the LSRE model when $n \leq 5$, the actual implementation of minimum distance for estimating the error density in this model makes use of the

quantities

$$W_i = \frac{X_{i1} - \bar{X}_{i(23)}}{X_{i4} - X_{i5}} = \frac{Z_{i1} - \bar{Z}_{i(23)}}{Z_{i4} - Z_{i5}} = \frac{U_i}{V_i}, \quad i = 1, \ldots, p,$$

instead of the residuals. The characteristic function of W corresponding to a normal mixture approximation of f_Z can be written as

$$\begin{aligned} \widetilde{\psi}_W(t) &= E(E(\exp\{itU/V\}|V = v)) \\ &= E(E(\exp\left\{it(Z_{i1} - \bar{Z}_{i(23)})/V\right\}|V = v)) \\ &= E(\widetilde{\psi}_Z(t/V)\,\widetilde{\psi}_Z^2(-t/(2V))), \end{aligned} \tag{3.11}$$

where $\widetilde{f}_Z$ and $\widetilde{\psi}_Z$ are the density and characteristic functions of the normal mixture, respectively. Furthermore, the density function of V under the normal mixture is obtained by the convolution equation

$$\widetilde{f}_V(v) = \int_{-\infty}^{\infty} \widetilde{f}_Z(v + \xi)\widetilde{f}_Z(\xi)d\xi,$$

which is used to compute the expectation in Equation (3.11) to yield

$$\widetilde{\psi}_W(t) = \int_{-\infty}^{\infty} \widetilde{\psi}_Z(t/v)\,\widetilde{\psi}_Z^2(-t/(2v))\,\widetilde{f}_V(v)dv. \tag{3.12}$$

While minimizing

$$D^2(\hat{f}_W, \widetilde{f}_W) = \frac{1}{2\pi}\int_{-\infty}^{\infty} |\hat{\psi}_W(t) - \widetilde{\psi}_W(t)|dt$$

is computationally challenging, a discretized version of the metric can be computed for a given normal mixture $\widetilde{f}_Z$ as follows:

(1) Compute $\widetilde{f}_V(v)$ on an appropriately chosen grid of points $[v_1, \ldots, v_{m_1}]$ to obtain $[\widetilde{f}_V(v_1), \ldots, \widetilde{f}_V(v_{m_1})]$;

(2) Compute $\widetilde{\psi}_W(t)$ on an grid of points $[t_1, \ldots, t_{m_2}]$ by

$$\widetilde{\psi}_W(t_i) = \sum_{j=1}^{m_1} \widetilde{\psi}_Z(t_i/v_j)\, \widetilde{\psi}_Z^2(-t_i/(2v_j))\, \widetilde{f}_V(v_j)\Delta v_j$$

to obtain $[\widetilde{\psi}_W(t_1), \ldots, \widetilde{\psi}_W(t_{m_2})]$, where $\Delta v_j = v_j - v_{j-1}$;

(3) Compute

$$D^2(\hat{f}_W, \widetilde{f}_W) \approx \frac{1}{2\pi} \sum_{i=1}^{m_2} |\hat{\psi}_W(t_i) - \widetilde{\psi}_W(t_i)|\Delta t_i$$

to obtain a numerical approximation to $D^2(\hat{f}_W, \widetilde{f}_W)$, where $\Delta t_i = t_i - t_{i-1}$;

The metric is then minimized subject to the linear constraint that the error density has a mean 0. The variance constraint was not enforced since this had little effect on the resulting estimator and because the nonlinear constraint could sometimes lead to numerical instabilities. Minimum distance estimates using a normal mixture are computed for the error density under the LSRE model for different values of L. The estimates are shown in Figure 27.

The estimate corresponding to $L = 9$ components with common scale parameter $\hat{\sigma} = 0.56$ is selected as the density estimate of the error under the LSRE model.

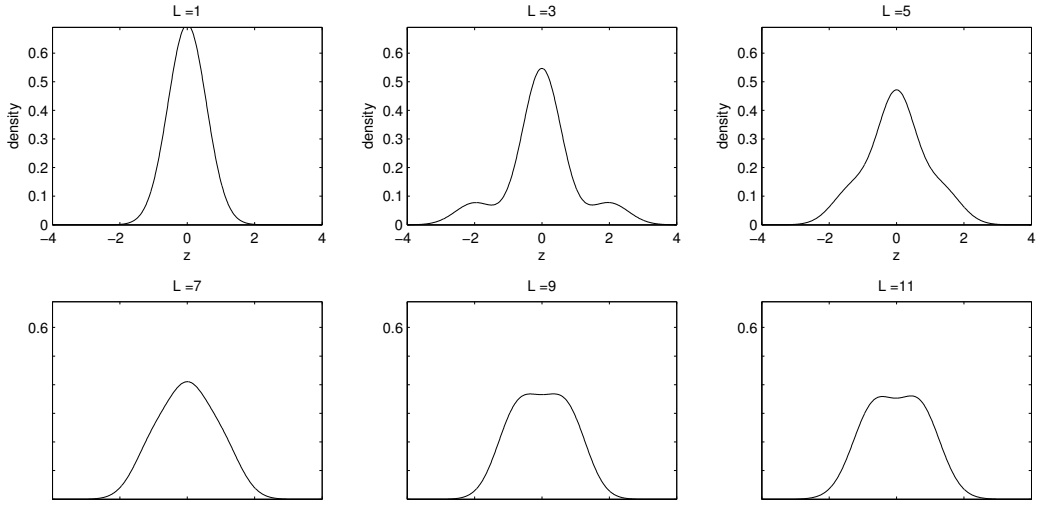

Figure 27. Minimum distance estimates of error density under LSRE model when the normal mixture has L=1, 3, 5, 7, 9, and 11 components.

A comparison of the two estimates for the error density in the LRE and LSRE models are shown in Figure 28.

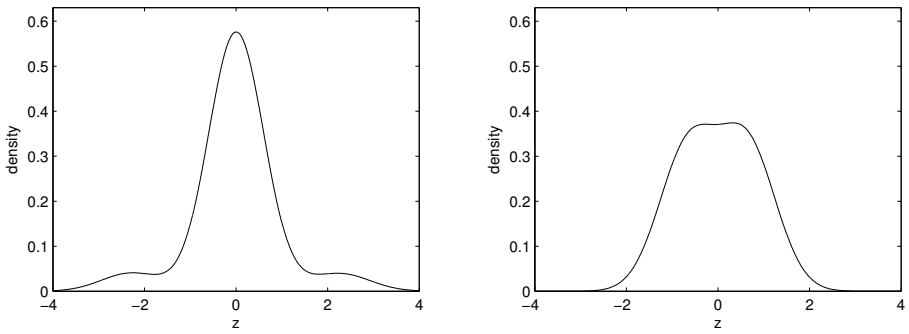

Figure 28. Density estimates for error density under LRE model (left) and LSRE model (right).

Interestingly, the error densities under the two models are quite different. It is also reassuring that these estimates are similar to those obtained in Hart and Cañette (2009), who use a different method. While both are symmetric, the error density under the LRE model has heavier tails than the density under the LSRE model. As noted in Hart and Cañette, the estimates indicate that there are scale differences, and these are inflating the tails of the LRE estimates. Depending on the setting, the differences in error densities can lead to very different conclusions.

CHAPTER IV

CONCLUSION

4.1 Summary

In this dissertation, a minimum distance methodology is developed for density estimation in the large-p, small-n setting when the variable of interest is not directly observed. The methodology is adapted to estimate the target density in the measurement error and the LRE model. A main advantage to a minimum distance approach is that it can be applied to a much more general class of problems than most of the estimators found in the deconvolution literature. Many of these estimators require that the characteristic function of the density of interest can be explicitly expressed in terms of a characteristic function of observable data. However, it can often be very difficult to come up with such an expression.

While there are many ways to approach this problem, it is assumed that the underlying density of interest could be well approximated by a mixture of normals. The means are assumed to be equally spaced, and a common scale parameter is used for the normal mixture components. The mixing proportions $\mathbf{p}$ could then be estimated by minimizing a density-based metric. This metric compares a density estimate of the observed data to the density induced from assuming a mixture of normals representation of the target density. In both models, the minimum distance methodology is compared with another deconvolution method. In the measurement error model, the results indicate that the proposed estimator performs as well as the deconvoluting kernel estimator of Stefanski and Carroll (1990) using different bandwidth selection methods. However, a shortcoming of the algorithm in this model is in selecting the

scale parameter σ. Unlike in the measurement error model, estimating the scale parameter of the normal mixture in the LRE model is not problematic. Furthermore, there seems to be quite an advantage to using the minimum distance estimator in the LRE model compared with the methodology of Hall and Yao (2003). Identifiability of the target densities in both models is considered. In addition, consistency of the estimator is addressed.

4.2 Further Research

Using univariate and bivariate distributions of the residuals in the LRE model when estimating the target density is discussed. It is still an open question as to whether or not there is an advantage to using a higher dimensional distribution of the observable data to identify the target density. As previously discussed, even though strict identifiability may hold, there may be some advantage to using a higher dimensional distribution in terms of mitigating near identifiability problems. It would also be of interest to devise some way of measuring near identifiability in a problem for a given model and transformation of the observed data.

Several theoretical aspects of the problem should be investigated in greater detail. For example, identifiability in the LSRE model as well as other models should be further studied. Mathematically, this is equivalent to finding conditions for which a particular integral equation is unique. Furthermore, consistency results for the minimum distance estimator should be strengthened. While a weak form of consistency is shown to hold, it is done rather indirectly and rates of convergence could not be established. In addition, relative efficiency of the minimum distance estimator compared with other estimators is an open question, although Hart and Cañette (2009) mention that their simulation results suggest that minimum distance is more efficient

than plug-in type estimators. The nature of deconvolution and the way minimum distance is implemented makes the question of efficiency a difficult one.

Finally, several extensions to the mixture approach could also be considered. For example, while a mixture of normals is used to represent the underlying density of interest, this could be extended to include a broader class of distributions for each mixture component. Also, instead of equally spacing the means of the mixture components on an interval, the locations could be assumed to follow an arbitrary, but known, distribution. It would also be of interest to reverse the problem and fix the mixing proportions to be $1/K$ and estimate the location of the mixture components. This approach is analogous to that of Hart and Cañette (2009).

REFERENCES

Beran, R. and Millar, P. W. (1994), "Minimum Distance in Random Coefficient Regression Models," *Annals of Statistics*, 22, 1976-1992.

Butucea, C. (2004), "Deconvolution of Supersmooth Densities with Smooth Noise," *The Canadian Journal of Statistics*, 32, 181-192.

Carroll, R. J. and Hall, P. (1988), "Optimal Rates of Convergence for Deconvolving a Density," *Journal of the American Statistical Association*, 83, 1184-1186.

Carroll, R. J. and Hall, P. (2004), "Low Order Approximations in Deconvolution and Regression with Errors in Variables," *Journal of the Royal Statistical Society. Series B*, 66, 31-46.

Cordy, C. B. and Thomas, D. R. (1997), "Deconvolution of a Distribution Function," *Journal of the American Statistical Association*, 92, 1459-1465.

Cox, D. R. and Hall, P. (2002), "Estimation in a Simple Random Effects Model with Nonnormal Distributions," *Biometrika*, 89, 832-840.

Davidson, L. A., Nguyen, D. V., Hokanson, R. M., Callaway, E. S., Isett, R. B., Turner, N. D., Dougherty, E. R., Wang, N., Lupton, J. R., Carroll, R. J., and Chapkin, R. S. (2004), "Chemopreventive $n-3$ Polyunsaturated Fatty Acids Reprogram Genetic Signatures During Colon Cancer Initiation and Progression in the Rat," *Cancer Research*, 64. 6797-6804.

Delaigle, A. and Gijbels, I. (2004), "Comparison of Data-Driven Bandwidth Selection Procedures in Deconvolution Kernel Density Estimation," *Computational Statistics & Data Analysis,* 45, 249-267.

Delaigle, A., Hall, P., and Meister, A. (2008), "On Deconvolution with Repeated Measurements," *The Annals of Statistics*, 36, 665-685.

Devroye, L. (1989), "Consistent Deconvolution in Density Estimation," *The Canadian Journal of Statistics*, 17, 235-239.

Diggle, P. and Hall, P. (1993), "A Fourier Approach to Nonparametric Deconvolution of a Density Estimate," *Journal of the Royal Statistical Society. Series B*, 55, 523-531.

Fan, J. (1991), "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.

Feller, W. (1971), *An Introduction to Probability Theory and Its Applications, Volume 2*, New York: Wiley.

Feuerverger, A. and Mureika, R. A. (1977), "The Empirical Characteristic Function and Its Applications," *Annals of Statistics*, 5, 88-97.

Hall, P. and Lahiri, S. N. (2008), "Estimation of Distributions, Moments and Quantiles in Deconvolution Problems," *The Annals of Statistics*, 36, 2110-2134.

Hall, P. and Yao, Q. (2003), "Inference in Components of Variance Models with Low Replication," *Annals of Statistics*, 31, 414-441.

Hart, J. D. and Cañette, I. (2009), "Nonparametric Estimation of Distributions in a Large-p, Small-n Setting," Technical Report, Texas A&M University.

Hesse, C. H. (1999), "Data-Driven Deconvolution," *Nonparametric Statistics*, 10, 343-373.

Horowitz, J. L. and Markatou, M. (1996), "Semiparametric Estimation of Regression Models for Panel Data," *The Review of Economic Studies*, 63, 145-168.

Jones, D. Y., Schatzkin, A., Green, S. B., Block, G., Brinton, L. A., Ziegler, R. G., Hoover, R., and Taylor, P. R. (1987), "Dietary Fat and Breast Cancer in the National Health and Nutritional Survey I: Epidemiologic Follow-up Study," *Journal of the National Cancer Institute*, 79, 465-471.

Jones, M. C., Marron, J. S., and Sheather, S. J. (1996), "A Brief Survey of Bandwidth Selection for Density Estimation," *Journal of the American Statistical Association*, 91, 401-407.

Kotlarski, I. I. (1967), "On Characterizing the Gamma and the Normal Distributions," *Pacific Journal of Mathematics*, 20, 69-76.

Li, T. and Vuong, Q. (1998), "Nonparametric Estimation of the Measurement Error Model Using Multiple Indicators," *Journal of Multivariate Analysis*, 65, 139-165.

Liu, M. C. and Taylor, R. L. (1989), "A Consistent Nonparametric Density Estimator for the Deconvolution Problem," *The Canadian Journal of Statistics*, 17, 427-438.

Lukacs, E. (1970), *Characteristic Functions*, London: Griffin.

Neumann, M. H. (2007), "Deconvolution from Panel Data with Unknown Error Distribution," *Journal of Multivariate Analysis*, 98, 1955-1968.

Rao, B. L. S. P. (1983), *Nonparametric Functional Estimation*, Orlando, Florida: Academic Press.

Rao, B. L. S. P. (1992), *Identifiability in Stochastic Models: Characterization of Probability Distributions*, Cambridge, Massachusetts: Academic Press.

Reiersøl, O. (1950), "Identifiability of a Linear Relationship Between Variables which are Subject to Error," *Econometrica*, 18, 375-389.

Sain, S. R., Baggerly, K. A., and Scott, D. W. (1994), "Cross-validation of Multivariate Densities," *Journal of the American Statistical Association*, 89, 807-817.

Stefanski, L. and Carroll, R. J. (1990), "Deconvoluting Kernel Density Estimators," *Statistics*, 21, 169-184.

Susko, E. and Nadon, R. (2002), "Estimation of a Residual Distribution with Small Numbers of Repeated Measurements," *The Canadian Journal of Statistics*, 30, 383-400.

Ushakov, N. G. (1999), *Selected Topics in Characteristic Functions*, Utrecht, The Netherlands: VSP VB.

Wolfowitz, J. (1957), "The Minimum Distance Method," *The Annals of Mathematical Statistics*, 28, 75-88.

APPENDIX A

CONSISTENCY IN MEASUREMENT ERROR MODEL

Assumptions:

1. $\psi_Z(t)$ does not vanish throughout an interval, and $\psi_Z(t)$ is square integrable;
2. $\psi_X(t)$, $\psi_X^{'}(t)$, and $\psi_X^{''}(t)$ are absolutely integrable;
3. f_X is everywhere continuous, f_X is bounded, and $|f_X^{\prime}|$ is bounded and integrable;
4. $b_n = n^{-1/4}$, $K(\cdot)$ is a pdf such that $\int_{-\infty}^{\infty} |u| K(u)\, du < \infty$, $K(\cdot)$ is of bounded variation, and $f_X * f_Z$ and $(f_X * f_Z)^{(1)}$ are bounded;
5. The means of the normal mixtures are evenly spaced on $(-M, M)$;
6. $\sigma \to 0$, $\frac{M^2}{L\sigma} \to 0$, $\frac{M}{L\sigma^2} \to 0$, $M\sigma \to \infty$, and $\left(\frac{L}{2MS_L} - 1\right)\frac{1}{\sigma} \to 0$, where $S_L = \sum_{i=1}^{L} f_X(\mu_i)$;

To develop notation, represent $f_X(x)$ by

$$f_X(x; \mathbf{w}, \mu, \sigma) = \frac{1}{\sigma} \sum_{j=1}^{L} w_j \phi\left(\frac{x - \mu_j}{\sigma}\right), \tag{A.1}$$

where $\mathbf{w} = (w_1, \ldots, w_{L-1})$, $w_L = 1 - \sum_{j=1}^{L-1} w_j$, and $\mu = (\mu_1, \ldots, \mu_L)$. Define $f_Y(y; \mathbf{w}, \mu, \sigma)$ by the convolution equation

$$f_Y(y; \mathbf{w}, \mu, \sigma) = \int_{-\infty}^{\infty} f_X(y - u; \mathbf{w}, \mu, \sigma) f_Z(u) du = (f_X * f_Z)(y), \quad \text{for all } y.$$

Let $\hat{f}_b$ be a kernel estimate of $f_X * f_Z$ based on the observed data having bandwidth b. Define, for an arbitrary density h,

$$D^2(f_Y(\cdot; \mathbf{w}, \mu, \sigma), h) = \int_{-\infty}^{\infty} (f_Y(y; \mathbf{w}, \mu, \sigma) - h(y))^2 dy, \tag{A.2}$$

where it is assumed that $\mathbf{w} \in \Theta_L = \{\theta : \theta_j \geq 0, j = 1, \ldots, L-1, \theta_1 + \cdots + \theta_{L-1} \leq 1\}$. For a given μ and σ, define $\widetilde{\mathbf{p}}_L$ by

$$D^2(f_Y(\cdot; \widetilde{\mathbf{p}}_L, \mu, \sigma), f_X * f_Z) = \inf_{\mathbf{w} \in \Theta_L} D^2(f_Y(\cdot; \mathbf{w}, \mu, \sigma), f_X * f_Z).$$

Estimate $\widetilde{\mathbf{p}}_L$ by $\hat{\mathbf{p}}_{Ln}$, where

$$D^2(f_Y(\cdot; \hat{\mathbf{p}}_{Ln}, \mu, \sigma), \hat{f}_b) = \inf_{\mathbf{w} \in \Theta_L} D^2(f_Y(\cdot; \mathbf{w}, \mu, \sigma), \hat{f}_b).$$

Define,

$$\begin{aligned} D_1^2(\mathbf{w}) &\equiv D^2(f_Y(\cdot; \mathbf{w}, \mu, \sigma), \hat{f}_b) \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \left| \psi_Z(t)\, e^{-\sigma^2 t^2/2} \sum\nolimits_{j=1}^{L} w_j e^{it\mu_j} - \psi_K(bt)\, \psi_n(t) \right|^2 dt, \end{aligned} \tag{A.3}$$

where $\psi_Z(t)$ and $\psi_K(t)$ are the characteristic functions of f_Z and kernel K, respectively, and

$$\psi_n(t) = \frac{1}{n} \sum_{j=1}^{n} e^{itX_j}.$$

Here, the assumption is used that $\psi_Z(t)$ does not vanish throughout an interval. Likewise, define

$$\begin{aligned} D_2^2(\mathbf{w}) &\equiv D^2(f_Y(\cdot; \mathbf{w}, \mu, \sigma), f_X * f_Z) \\ &= \frac{1}{2\pi} \int_{-\infty}^{\infty} \left| \psi_Z(t)\, e^{-\sigma^2 t^2/2} \sum\nolimits_{j=1}^{L} w_j e^{it\mu_j} - \psi_Z(t)\psi_X(t) \right|^2 dt. \end{aligned} \tag{A.4}$$

The proof proceeds by first showing that for fixed L and σ, $\hat{\mathbf{p}}_{Ln} \rightarrow_{a.s.} \widetilde{\mathbf{p}}_L$ as $n \rightarrow \infty$. This is followed by showing that $e^{-\sigma^2 t^2/2} \sum_{j=1}^{L} \widetilde{p}_{jL} e^{it\mu_j} \rightarrow_{a.s.} \psi_X(t)$ as $L \rightarrow \infty$ and $\sigma \rightarrow 0$, where $\widetilde{\mathbf{p}}_L = [\ \widetilde{p}_{1L} \ \ \widetilde{p}_{2L} \ \ \cdots \ \ \widetilde{p}_{LL}\ ]$. Conditions are then established for $f_X(x; \widetilde{\mathbf{p}}_L, \mu, \sigma) \rightarrow_{a.s.} f_X(x)$ as $L \rightarrow \infty$ and $\sigma \rightarrow 0$. Finally, conclude that $f_X(x; \hat{\mathbf{p}}_{Ln}, \mu, \sigma) \rightarrow_{a.s.} f_X(x)$ as $n \rightarrow \infty$, $L \rightarrow \infty$, and $\sigma \rightarrow 0$.

(1) To show that $\hat{\mathbf{p}}_{Ln} \rightarrow_{a.s.} \widetilde{\mathbf{p}}_L$ as $n \rightarrow \infty$ for fixed L and σ, the following

sufficient conditions must be satisfied:

(i) $D_2^2(\mathbf{w})$ is uniquely minimized at $\widetilde{\mathbf{p}}_L$;

(ii) $D_2^2(\mathbf{w})$ is continuous in $\mathbf{w}$;

(iii) the parameter space Θ_L is compact;

(iv) $D_1^2(\mathbf{w})$ converges uniformly to $D_2^2(\mathbf{w})$ almost surely;

Each of these conditions will now be verified. First observe that since $D_2^2(\mathbf{w})$ previously defined is a quadratic form and a continuous function in $\mathbf{w}$, the minimizer $\widetilde{\mathbf{p}}_L$ is unique. Hence, (i) and (ii) hold. In addition, (iii) holds by the Heine–Borel theorem. To show that

$$\sup_{\mathbf{w}\in\Theta_L} |D_1^2(\mathbf{w}) - D_2^2(\mathbf{w})| \to_{a.s.} 0, \tag{A.5}$$

the shorter notation is used, $c_n(t) = \psi_Z(t)e^{-\sigma^2t^2/2}\sum_{j=1}^{L} w_j e^{it\mu_j}$, $d_n(t) = \psi_K(bt)\psi_n(t)$, and $e_n(t) = \psi_Z(t)\psi_X(t)$. Then $|D_1^2(\mathbf{w}) - D_2^2(\mathbf{w})|$ can be written as

$$\left|\frac{1}{2\pi}\int_{-\infty}^{\infty}|c_n - d_n|^2dt - \frac{1}{2\pi}\int_{-\infty}^{\infty}|c_n - e_n|^2dt\right|.$$

It is then true that

$$\left|\frac{1}{2\pi}\int_{-\infty}^{\infty}|c_n - d_n|^2dt - \frac{1}{2\pi}\int_{-\infty}^{\infty}|c_n - e_n|^2dt\right| \leq \frac{1}{2\pi}\int_{-\infty}^{\infty}\left||c_n - d_n|^2 - |c_n - e_n|^2\right|dt. \tag{A.6}$$

Working with the right-hand side of Equation (A.6) and ignoring the constant term, it can be seen that

$$\begin{aligned}\int_{-\infty}^{\infty}\left||c_n-d_n|^2-|c_n-e_n|^2\right|dt &= \int_{-\infty}^{\infty}|(c_n-d_n)(\bar{c}_n-\bar{d}_n)-(c_n-e_n)(\bar{c}_n-\bar{e}_n)|dt\\ &= \int_{-\infty}^{\infty}|d_n\bar{d}_n-e_n\bar{e}_n+c_n(\bar{e}_n-\bar{d}_n)+\bar{c}_n(e_n-d_n)|dt\\ &\leq \int_{-\infty}^{\infty}(|d_n\bar{d}_n-e_n\bar{e}_n|+|c_n(\bar{e}_n-\bar{d}_n)|\\ &\qquad +|\bar{c}_n(e_n-d_n)|)dt\\ &\leq \int_{-\infty}^{\infty}|d_n\bar{d}_n-e_n\bar{e}_n|dt+2\int_{-\infty}^{\infty}e^{-\sigma^2t^2/2}|e_n-d_n|dt.\end{aligned}$$

The last inequality results from the fact that $|c_n(t)|$ can be bounded by $e^{-\sigma^2t^2/2}$. To show that $\int_{-\infty}^{\infty}|d_n\bar{d}_n-e_n\bar{e}_n|dt \rightarrow_{a.s.} 0$, by Parseval's formula, $\int_{-\infty}^{\infty}\hat{f}_b^2dx \rightarrow_{a.s.} \int_{-\infty}^{\infty}(f_X*f_Z)^2dx$ implies that $\int_{-\infty}^{\infty}|d_n\bar{d}_n-e_n\bar{e}_n|dt \rightarrow_{a.s.} 0$. By Theorem 4.4.1 of Rao (1983), if the assumptions in 4 hold, then $\int_{-\infty}^{\infty}\hat{f}_b^2dx \rightarrow_{a.s.} \int_{-\infty}^{\infty}(f_X*f_Z)^2dx$. Furthermore, to show that the second term in the above expression goes to 0 as $n\rightarrow\infty$, recall (a) $e^{-\sigma^2t^2/2}|(e_n-d_n)|\leq 2e^{-\sigma^2t^2/2}$ a.e., (b) $\int_{-\infty}^{\infty}e^{-\sigma^2t^2/2}dt<\infty$ for $\sigma>0$, and (c) $|(e_n-d_n)|\rightarrow 0$ a.e. Then by dominated convergence, it is true that

$$2\int_{-\infty}^{\infty}e^{-\sigma^2t^2/2}|(e_n-d_n)|dt\rightarrow_{a.s.}0$$

as $n\rightarrow\infty$. Note that this holds for all $\mathbf{w}\in\Theta_L$ and thus $\sup_{\mathbf{w}\in\Theta_L}|D_1^2(\mathbf{w})-D_2^2(\mathbf{w})|\rightarrow_{a.s.}$ 0.

Using (i), (ii), (iii), and (iv), it can now shown that $\hat{\mathbf{p}}_{Ln}\rightarrow_{a.s.}\widetilde{\mathbf{p}}_L$. For any $\varepsilon>0$, almost surely (a) $D_1^2(\widetilde{\mathbf{p}}_L)-\varepsilon/3<D_1^2(\hat{\mathbf{p}}_{Ln})$ since $\hat{\mathbf{p}}_{Ln}$ minimizes D_1^2, (b) $D_2^2(\widetilde{\mathbf{p}}_L)-\varepsilon/3<D_1^2(\widetilde{\mathbf{p}}_L)$ and (c) $D_1^2(\hat{\mathbf{p}}_{Ln})-\varepsilon/3<D_2^2(\hat{\mathbf{p}}_{Ln})$ since $D_1^2(\mathbf{w})$ converges uniformly to $D_2^2(\mathbf{w})$ almost surely. Combining (a), (b), and (c), it holds that almost surely

$$D_2^2(\widetilde{\mathbf{p}}_L)-\varepsilon<D_1^2(\widetilde{\mathbf{p}}_L)-2\varepsilon/3<D_1^2(\hat{\mathbf{p}}_{Ln})-\varepsilon/3<D_2^2(\hat{\mathbf{p}}_{Ln}),$$

which implies that for any $\varepsilon > 0$,

$$D_2^2\left(\widetilde{\mathbf{p}}_L\right) - \varepsilon < D_2^2\left(\hat{\mathbf{p}}_{Ln}\right). \tag{A.7}$$

Let $\mathcal{S}$ be an arbitrary open subset of the parameter space Θ_L that contains $\widetilde{\mathbf{p}}_L$. Since $\Theta_L \cap \mathcal{S}^c$ is compact, and $D_2^2\left(\mathbf{w}\right)$ is continuous and uniquely minimized at $\widetilde{\mathbf{p}}_L$, it can be seen that

$$\sup_{\mathbf{w}\in\Theta_L\cap\mathcal{S}^c} D_2^2(\mathbf{w}) \equiv D_2^2(\widetilde{\mathbf{p}}_L^*) < D_2^2(\widetilde{\mathbf{p}}_L).$$

By choosing $\varepsilon = D_2^2(\widetilde{\mathbf{p}}_L) - D_2^2(\widetilde{\mathbf{p}}_L^*)$ and using Equation (A.7), $D_2^2(\widetilde{\mathbf{p}}_L^*) < D_2^2(\hat{\mathbf{p}}_{Ln})$ almost surely. Therefore, it follows that $\hat{\mathbf{p}}_{Ln} \in \mathcal{S}$ which implies that $\hat{\mathbf{p}}_{Ln} \rightarrow_{a.s.} \widetilde{\mathbf{p}}_L$ as $n \rightarrow \infty$.

(2) To show $\widetilde{f}_L(x) = f_X(x; \widetilde{\mathbf{p}}_L, \mu, \sigma) \rightarrow_{a.s.} f_X(x)$ as $L \rightarrow \infty$ and $\sigma \rightarrow 0$, first find a solution with respect to the mixing proportions that satisfies the natural constraints on the proportions. This explicit solution is useful since the actual proportions are estimated numerically without an explicit form. Let $u_j = (j - 1/2)/L,\ j = 1, \ldots, L$, and set $\mu_j = M(2u_j - 1),\ j = 1, \ldots, L$. Define

$$p_{jL} = \frac{f_X(\mu_j)}{\sum_{i=1}^{L} f_X(\mu_i)} = \frac{f_X(\mu_j)}{S_L}, \quad j = 1, \ldots, L, \tag{A.8}$$

and write $\mathbf{p}_L = (p_{1L}, \ldots, p_{(L-1)L})$. It is worth mentioning that μ_j's can be taken to be the quantiles $\widetilde{Q}$ of an arbitrary distribution $\widetilde{G}$ other than uniform on $(-M, M)$, but this generalization is not considered here. It can be seen that $p_{jL},\ j = 1, \ldots, L$, from Equation (A.8) are nonnegative and sum to one. Note that $f_X(x; \mathbf{p}_L, L, \sigma)$ can

be written as

$$f_X(x;\mathbf{p}_L,L,\sigma)=\sum_{j=1}^{L}\frac{1}{\sigma}\cdot\frac{\frac{2M}{L}f_X(\mu_j)}{\frac{2M}{L}\sum_{i=1}^{L}f_X(\mu_i)}\phi\left(\frac{x-\widetilde{Q}((j-1/2)/L)}{\sigma}\right)$$
$$\approx\int_{-\infty}^{\infty}\frac{1}{\sigma}\cdot f_X(\mu)\,\phi\left(\frac{x-\mu}{\sigma}\right)d\mu$$

for large L and M. This is because $\int_{-\infty}^{\infty}f_X(x)$ can be written as

$$\begin{aligned}\int_{-\infty}^{\infty}f_X(x)&=\int_{-M}^{M}f_X(x)dx+\int_{|x|>M}f_X(x)dx\\&=\sum_{j=1}^{L}\int_{M(2(j-1)/L-1)}^{M(2j/L-1)}f_X(x)dx+\int_{|x|>M}f_X(x)dx\\&=\sum_{j=1}^{L}\frac{2M}{L}f_X(\mu_j^*)+\int_{|x|>M}f_X(x)dx,\end{aligned}$$

where $M(2(j-1)/L-1)\le\mu_j^*\le M(2j/L-1),\ j=1,\ldots,L$. By assumption 2, $f_X(x)\sim o\left(\frac{1}{x^2}\right)$ as $x\to\infty$, and hence $\int_{|x|>M}f_X(x)\,dx\to 0$, which implies that $\sum_{j=1}^{L}\frac{2M}{L}f_X(\mu_j^*)\to 1$ as $M\to\infty$ and $L\to\infty$. It can be seen that

$$\int_{-\infty}^{\infty}\frac{1}{\sigma}\cdot f_X(\mu)\phi\left(\frac{x-\mu}{\sigma}\right)d\mu\to f_X(x)\quad\text{as }\sigma\to 0.$$

Therefore, $f_X(x;\mathbf{p}_L,L,\sigma)\to f_X(x)$ as $L\to\infty$ and $\sigma\to 0$, which implies that $\psi(t;\mathbf{p}_L,\mu,\sigma)\to_{a.s.}\psi_X(t)$ as $L\to\infty$ and $\sigma\to 0$. Then $\mathbf{p}_L\in\Theta_L$, and so $D_2^2(\widetilde{\mathbf{p}}_L)\le D_2^2(\mathbf{p}_L)$. To show that $D_2^2(\mathbf{p}_L)\to 0$ as $L\to\infty$ and $\sigma\to 0$, let $\psi_{M_L}(t)$ be short for $\psi(t;\mathbf{p}_L,\mu,\sigma)$, which is the characteristic function corresponding to the mixture of normals. Then

$$\begin{aligned}2\pi\cdot D_2^2(\mathbf{p}_L)&=\int_{-\infty}^{\infty}|\psi_Z(t)|^2|\psi_{M_L}(t)-\psi_X(t)|^2dt\\&\le 4\int_{-\infty}^{\infty}|\psi_Z(t)|^2dt<\infty.\end{aligned}$$

Using the fact that $\psi_{M_L}(t)\to_{a.s.}\psi_X(t)$ as $L\to\infty$ and $\sigma\to 0$, by dominated conver-

gence $\lim_{L\to\infty,\sigma\to 0} D_2^2(\mathbf{p}_L) = 0$, which implies that $\lim_{L\to\infty,\sigma\to 0} D_2^2(\widetilde{\mathbf{p}}_L) = 0$. Rewriting D_2^2 as

$$D_2^2(\widetilde{\mathbf{p}}_L) = \frac{1}{2\pi}\int_{-\infty}^{\infty} |\psi_Z(t)|^2 \left| e^{-\sigma^2 t^2/2} \sum\nolimits_{j=1}^{L} \widetilde{p}_{jL} e^{it\mu_j} - \psi_X(t) \right|^2 dt,$$

it can be seen that the only way $D_2^2(\widetilde{\mathbf{p}}_L)$ can go to 0 is for $e^{-\sigma^2 t^2/2}\sum_{j=1}^{L} \widetilde{p}_{jL} e^{it\mu_j}$ to converge pointwise to $\psi_X(t)$ a.e. since $\psi_Z(t)$ was assumed not to vanish throughout an interval. Thus,

$$e^{-\sigma^2 t^2/2} \sum_{j=1}^{L} \widetilde{p}_{jL} e^{it\mu_j} \to_{a.s.} \psi_X(t) \quad \text{as } L \to \infty \text{ and } \sigma \to 0. \tag{A.9}$$

In order for Equation (A.9) to imply that $f_X(x; \widetilde{\mathbf{p}}_L, \mu, \sigma) \to_{a.s.} f_X(x)$ as $L \to \infty$ and $\sigma \to 0$, it is required that $\int_{-\infty}^{\infty} |\psi_X(t; \widetilde{\mathbf{p}}_L, \mu, \sigma) - \psi_X(t)| dt \to_{a.s.} 0$ as $L \to \infty$ and $\sigma \to 0$. Ideally, the convergence of $f_X(x; \widetilde{\mathbf{p}}_L, \mu, \sigma)$ that minimizes the L_2 metric should be investigated. However, conditions for which $\int_{-\infty}^{\infty} |\psi_X(t; \mathbf{p}_L, \mu, \sigma) - \psi_X(t)| dt \to_{a.s.} 0$ as $L \to \infty$ and $\sigma \to 0$ using the mixing proportions defined in Equation (A.8) are given. Define

$$\hat{\psi}_X(t) = e^{-\sigma^2 t^2/2} \sum_{j=1}^{L} p_{jL} e^{it\mu_j} = e^{-\sigma^2 t^2/2} \widetilde{\psi}_X(t),$$

where $\widetilde{\psi}_X(t) = \sum_{j=1}^{L} p_{jL} e^{it\mu_j}$. Then

$$\begin{aligned} \int |\hat{\psi}_X(t) - \psi_X(t)| dt \leq & \int e^{-\sigma^2 t^2/2} |\widetilde{\psi}_X(t) - \psi_X(t)| dt \\ & + \int (1 - e^{-\sigma^2 t^2/2}) |\psi_X(t)| dt. \end{aligned} \tag{A.10}$$

Since $\psi_X(t)$ is integrable, it follows from dominated convergence that the second term on the right hand side of Equation (A.10) tends to 0 as $\sigma \to 0$. Then $\psi_X(t)$ can

be written as

$$\begin{aligned}\psi_X(t) &= \int_{-M}^{M} f_X(x)e^{itx}dx + \int_{|x|>M} f_X(x)e^{itx}dx \\ &= \sum_{j=1}^{L}\int_{M(2(j-1)/L-1)}^{M(2j/L-1)} f_X(x)e^{itx}dx + \int_{|x|>M} f_X(x)e^{itx}dx \\ &= \frac{2M}{L}\sum_{j=1}^{L} f_X(x_j)e^{itx_j} + \int_{|x|>M} f_X(x)e^{itx}dx,\end{aligned}$$

where $M(2(j-1)/L-1) \le x_j \le M(2j/L-1),\ j=1,\ldots,L$. Furthermore,

$$\widetilde{\psi}_X(t) = \frac{2M}{L}\sum_{j=1}^{L} f_X(\mu_j)e^{it\mu_j} + \delta_{ML}(t),$$

where $\delta_{ML}(t) = \left(\frac{L}{2MS_L} - 1\right)\frac{2M}{L}\sum_{j=1}^{L} f_X(\mu_j)e^{it\mu_j}$ and $\sum_{i=1}^{L} f_X(\mu_i)$, and therefore

$$|\widetilde{\psi}_X(t) - \psi_X(t)| \le \frac{2M}{L}\sum_{j=1}^{L}|f_X(\mu_j)e^{it\mu_j} - f_X(x_j)e^{itx_j}| + |\delta_{ML}(t)| + \int_{|x|>M} f_X(x)dx.$$

Using differentiability of f_X,

$$\begin{aligned}|f_X(\mu_j)e^{it\mu_j} - f_X(x_j)e^{itx_j}| &= |f_X(\mu_j) - f_X(x_j)e^{it\mu_j} + f_X(x_j)(e^{it\mu_j} - e^{itx_j})| \\ &\le \frac{2M}{L}|f_X'(\eta_j)| + \frac{4M}{L}f_X(x_j)|t|,\end{aligned}$$

where η_j is between μ_j and x_j, $j=1,\ldots,L$. In addition,

$$\begin{aligned}e^{-\sigma^2t^2/2}|\widetilde{\psi}_X(t) - \psi_X(t)| \le e^{-\sigma^2t^2/2}\Bigg(&\frac{4M^2}{L^2}\sum\nolimits_{j=1}^{L}|f_X'(\eta_j)| + \sum\nolimits_{j=1}^{L}\frac{8M^2}{L^2}f_X(x_j)|t| \\ &+|\delta_{ML}(t)| + \int_{|x|>M} f_X(x)dx\Bigg). \qquad \text{(A.11)}\end{aligned}$$

By assumption 3, $\sum_{j=1}^{L}|f_X'(\eta_j)| \sim O(L)$ and $\frac{1}{L}\sum_{j=1}^{L}|f_X'(\eta_j)| \sim O(1)$, and the integral of the first term in Equation (A.11) becomes

$$\frac{2M^2}{L}O(1)\int e^{-\sigma^2t^2/2}dt = \frac{2\sqrt{2\pi}M^2}{L\sigma}O(1).$$

The integral of the second term in Equation (A.11) is

$$\frac{2M}{L}\int e^{-\sigma^2t^2/2}\left(\sum_{j=1}^{L}\frac{4M}{L}f_X(x_j)|t|\right)dt = \frac{4M}{L}\left(\sum_{j=1}^{L}\frac{2M}{L}f_X(x_j)\right)\int |t|e^{-\sigma^2t^2/2}dt$$
$$= \frac{8M}{L\sigma^2}\left(\sum_{j=1}^{L}\frac{2M}{L}f_X(x_j)\right).$$

Note from a previous argument that $\sum_{j=1}^{L}\frac{2M}{L}f_X(x_j) \to 1$ as $M \to \infty$ and $L \to \infty$. The integral of the third term in Equation (A.11) yields

$$|\delta_{ML}(t)|\int e^{-\sigma^2t^2/2}dt = \frac{\sqrt{2\pi}|\delta_{ML}(t)|}{\sigma}.$$

Recall $\delta_{ML}(t) = \left(\frac{L}{2MS_L}-1\right)\frac{2M}{L}\sum_{j=1}^{L}f_X(\mu_j)e^{it\mu_j}$, and so $|\frac{2M}{L}\sum_{j=1}^{L}f_X(\mu_j)e^{it\mu_j}| \leq \frac{2M}{L}\sum_{j=1}^{L}f_X(\mu_j) \to 1$ as $M, L \to \infty$. Thus, by assumption 3, $\frac{2M}{L}\sum_{j=1}^{L}f_X(\mu_j) \leq \beta$, it is true that

$$|\delta_{ML}(t)| \leq \beta|\frac{L}{2MS_L}-1|.$$

Furthermore, the integral of the fourth term becomes

$$\int e^{-\sigma^2t^2/2}\int_{|x|>M}f_X(x)dxdt = \frac{\sqrt{2\pi}}{\sigma M},$$

since $f_X(x) = o\left(\frac{1}{|x|^2}\right)$. Thus, in order for $\int e^{-\sigma^2t^2/2}|\widetilde{\psi}_X(t)-\psi_X(t)|dt \to 0$, each of the four terms above must go to zero, which is true since by assumption 6, $\frac{M^2}{L\sigma} \to 0$, $\frac{M}{L\sigma^2} \to 0$, $M\sigma \to \infty$, and $\left(\frac{L}{2MS_L}-1\right)\frac{1}{\sigma} \to 0$.

(3) It has been shown that (1) $\hat{\mathbf{p}}_{Ln} \to_{a.s.} \widetilde{\mathbf{p}}_L$ as $n \to \infty$ and (2) $f_X(x;\widetilde{\mathbf{p}}_L,\mu,\sigma) \to_{a.s.} f_X(x)$ as $L \to \infty$ and $\sigma \to 0$. To show that $\hat{f}_n(x) = f_X(x;\hat{\mathbf{p}}_{Ln},\mu,\sigma) \to_p f_X(x)$ as $n \to \infty$, $L \to \infty$, and $\sigma \to 0$, write

$$P(|\hat{f}_n - f_X| > \epsilon) \leq P(2|\hat{f}_n - \widetilde{f}_L| > \epsilon) + P(2|\widetilde{f}_L - f_X| > \epsilon).$$

It is known that $\widetilde{f}_L \to_{a.s.} f_X$ as $L \to \infty$ and $\sigma \to 0$. It suffices then, to produce

an unbounded sequence $\{L_n\}$ with the property that $|\hat{f}_n - \widetilde{f}_{L_n}| \to_p 0$. Recall that $\hat{f}_n \to_p \widetilde{f}_L$. Thus, there exists a smallest n_L such that for all $n \geq n_L$, $P(|\hat{f}_n - \widetilde{f}_L| > \epsilon) \leq \frac{1}{L}$. Let $\{L_1, L_2, \ldots\}$ be an arbitrary increasing sequence of positive integers. Then $P(|\hat{f}_{n_{L_j}} - \widetilde{f}_{L_j}| > \epsilon) \leq \frac{1}{L_j}$ for all j, and so $P(|\hat{f}_{n_{L_j}} - \widetilde{f}_{L_j}| > \epsilon) \to 0$ as $j \to \infty$. Typically, $n_{L_1} < n_{L_2} < n_{L_3} < \cdots$. However, if it happens that $n_{L_j} \leq n_{L_{j-1}}$, then define $n_{L_j} = n_{L_{j-1}} + 1$. Define $L(n)$ as

$$L(n) = L_j, \quad n_{L_j} \leq n < n_{L_{j+1}}, \quad j = 1, 2, \ldots.$$

For the sequence $L(n)$ defined above,

$$P(|\hat{f}_n - \widetilde{f}_{L(n)}|) \leq \frac{1}{L(n)} \to 0 \quad \text{as } n \to \infty.$$

and therefore $\lim_{n\to\infty} P(|\hat{f}_n - f_X| > \epsilon) = 0$. $\square$

APPENDIX B

IDENTIFIABILITY OF ERROR DENSITY IN LOCATION-SCALE RANDOM EFFECTS MODEL

Let X have distribution F_X and density f_X. Define $U = \text{sign}(X)$ and $V = |X|$. Then

$$\begin{aligned} F_{V,U}(v, u = 1) &= P(|X| \leq v, U = 1) \\ &= P(X \leq v, U = 1) \\ &= F_X(v) - F_X(0). \end{aligned}$$

In addition,

$$\begin{aligned} F_{V,U}(v, u = -1) &= P(|X| \leq v, U = -1) \\ &= P(-X \leq v, U = -1) \\ &= P(X \geq -v, U = -1) \\ &= F_X(0) - F_X(-v). \end{aligned}$$

Then

$$f_{V,U}(v, u) = \begin{cases} f_X(v) & \text{for } U = 1 \\ f_X(-v) & \text{for } U = -1 \end{cases}$$

and therefore

$$f_{V|U}(v|u) = \begin{cases} f_X(v)/P(U = 1) & \text{for } U = 1 \\ f_X(-v)/P(U = -1) & \text{for } U = -1 \end{cases}.$$

Clearly by knowing $f_{V,U}$ or $f_{V|U}$, f_X can be uniquely identified. Let $\psi_1(t)$ and $\psi_{-1}(t)$

be the characteristic functions of $V|U=1$ and $V|U=-1$, respectively. Then

$$\psi_1(t) = \int_0^\infty e^{itv} f_X(v)/P(U=1)\,dv, \tag{B.1}$$

and

$$\psi_{-1}(t) = \int_0^\infty e^{itv} f_X(-v)/P(U=-1)\,dv. \tag{B.2}$$

Note that the characteristic function of X given by

$$\begin{aligned}\psi_X(t) &= \int_{-\infty}^\infty e^{itx} f_X(x)\,dx \\ &= \int_{-\infty}^0 e^{itx} f_X(x)\,dx + \int_0^\infty e^{itx} f_X(x)\,dx\end{aligned}$$

can be written in terms of $\psi_1(t)$ and $\psi_{-1}(t)$ as

$$\psi_X(t) = P(U=-1)\,\psi_{-1}(-t) + P(U=1)\,\psi_1(t). \tag{B.3}$$

Now, consider the location-scale random effects model

$$X_{ij} = \alpha_i + \gamma_i Z_{ij}, \quad i=1,\ldots,p,\ j=1,\ldots,n$$

for $n=7$. Define δ_{i1} and δ_{i2} as

$$\delta_{i1} = \frac{X_{i1} - \bar{X}_{i(23)}}{X_{i4} - X_{i5}} \quad \text{and} \quad \delta_{i2} = \frac{X_{i1} - \bar{X}_{i(23)}}{X_{i6} - X_{i7}},$$

where $\bar{X}_{i(23)} = (X_{i2} + X_{i3})/2$. Then δ_{i1} and δ_{i2} can be written as

$$\delta_{i1} = \frac{Z_{i1} - \bar{Z}_{i(23)}}{Z_{i4} - Z_{i5}} \quad \text{and} \quad \delta_{i2} = \frac{Z_{i1} - \bar{Z}_{i(23)}}{Z_{i6} - Z_{i7}}.$$

A basic identifiability result from Kotlarski (1967) and discussed in Rao (1992) says that the distributions of $Z_{i1} - \bar{Z}_{i(23)}$, $Z_{i4} - Z_{i5}$, and $Z_{i6} - Z_{i7}$ can be identified from the joint distribution of $(\delta_{i1}, \delta_{i2})$ if

1. the characteristic functions of $Z_{i1} - \bar{Z}_{i(23)}$, $Z_{i4} - Z_{i5}$, and $Z_{i6} - Z_{i7}$ do not vanish,

2. the distributions of $Z_{i1} - \bar{Z}_{i(23)}$, $Z_{i4} - Z_{i5}$, and $Z_{i6} - Z_{i7}$ are symmetric, and

3. $Z_{i1} - \bar{Z}_{i(23)}$, $Z_{i4} - Z_{i5}$, and $Z_{i6} - Z_{i7}$ are independent.

In the case under consideration, the distributions of $Z_{i4} - Z_{i5}$ and $Z_{i6} - Z_{i7}$ are obviously the same, but they do not need to be. Furthermore, since the distribution of $Z_{i1} - \bar{Z}_{i(23)}$ can be identified, the distribution of Z_{ij} can also be identified.

This result can be extended to relax the symmetric assumption on $Z_{i1} - \bar{Z}_{i(23)}$, which in turn allows for an asymmetric distribution of Z_{ij}. Note that the distribution of $Z_{i4} - Z_{i5}$ is symmetric even when the distribution of Z_{ij} is not symmetric. For notation, let $V_i = |Z_{i1} - \bar{Z}_{i(23)}|$ and $U_i = \text{sign}(Z_{i1} - \bar{Z}_{i(23)})$. Now define the index sets $\mathcal{I}^+ = \{i : Z_{i1} - \bar{Z}_{i(23)} \geq 0\}$ and $\mathcal{I}^- = \{i : Z_{i1} - \bar{Z}_{i(23)} < 0\}$, and let $(\delta_{i1}^+, \delta_{i2}^+) = (\delta_{i1}, \delta_{i2})$ if $i \in \mathcal{I}^+$ and $(\delta_{i1}^-, \delta_{i2}^-) = (\delta_{i1}, \delta_{i2})$ if $i \in \mathcal{I}^-$. Then by applying the above result from Kotlarski separately for $(\delta_{i1}^+, \delta_{i2}^+)$ and $(\delta_{i1}^-, \delta_{i2}^-)$, the distributions of $V_i|U_i = 1$ and $V_i|U_i = -1$ can be identified. From the above, the conditional distributions of $V_i|U_i = 1$ and $V_i|U_i = -1$ uniquely identify the distribution of $V_i = Z_{i1} - \bar{Z}_{i(23)}$. Since the distribution of $Z_{i1} - \bar{Z}_{i(23)}$ can be identified, the distribution of Z_{ij} can be identified from a result in Hart and Cañette (2009).

VITA

Nathaniel A. Litton was born in Houston, Texas. He graduated from Conroe High School in Conroe, Texas in 2000. He received a Bachelor of Science in operations research and industrial engineering from Cornell University in Ithaca, New York in 2004. After working for a year as an engineer, he joined the Department of Statistics at Texas A&M University in 2005. He received his Master of Science in statistics from Texas A&M University in 2007, and received his Doctor of Philosophy in statistics under the direction of Dr. Jeffrey D. Hart in August 2009. His mailing address is

Department of Statistics
Texas A&M University
3143 TAMU
College Station, TX 77843-3141
c/o Jeffrey D. Hart, Ph.D.