Monte Carlo Estimation of $g(\mu)$ from Normally Distributed Data with Applications

Leonard A. Stefanski
Department of Statistics
North Carolina State University
Raleigh, NC 27695-8203, USA

Steven J. Novick
Schering-Plough Research Institute
Kenilworth, NJ, USA 07033-1300

Viswanath Devanarayan
Eli Lilly & Company
Indianapolis, IN 46285, USA

Institute of Statistics Mimeo Series No. 2569

Abstract We derive Monte Carlo-amenable solutions to the problem of unbiased estimation of a nonlinear function of the mean of a normal distribution. For most nonlinear functions the maximum likelihood estimator is biased. Our method yields a Monte Carlo approximation to the uniformly minimum variance unbiased estimator for a wide class of nonlinear functions. Applications to problems arising in the analysis of data measured with error and the secondary analysis of estimated data are described.

Keywords: Corrected Score, Deconvolution, Jackknife, Measurement Error, Nonlinear, Resampling, SIMEX, Spherical Uniform Distribution, UMVUE.

1 Introduction

We present simulation-based solutions to the problem of estimating $g(\mu)$ given an iid sample $W_1, \ldots, W_m \sim \mathrm{N}(\mu,\ \sigma^2)$ with both μ and σ^2 unknown, and describe applications to the analysis of data measured with error. This problem and closely related problems have a long history dating from the 1950s (Kolmogorov, 1950; Kitagawa, 1956; Washio et al., 1956; Neyman and Scott, 1960; Gray, Watkins and Schucany 1973; Woodward, 1977; Watkins, 1979), and solutions have been derived using a variety of approaches. The solution most relevant to our approach is that of Gray, et al. (1973) for which Fortran subroutines were presented by Woodward (1977).

Gray et al. (1973) used the generalized jackknife to derive the uniformly minimum variance unbiased estimator of $g(\mu)$ for polynomial $g()$. Then, noting the form of the solution, they generalized the estimator to the case in which $g()$ has an infinite power series representation, resulting in the estimator

$$\widehat{g}(\mu) = g(\overline{W}) + \sum_{k=1}^{\infty} \frac{(-1)^k\ \Gamma((m-1)/2)\ g^{(2k)}(\overline{W})}{k!\ \Gamma((m-1)/2+k)}\ \left(\frac{\widehat{\sigma}^2}{4}\right)^k.$$

Like Gray et al. (1973) our estimator is derived from initial consideration of generalized jackknife estimators. However, the path we take differs from that of Gray et al. (1973) and leads to characterizations of the estimator as expectations of relatively simple functions with respect to certain uniform distributions. This in turn suggests a simple Monte Carlo estimation procedure. A significant theoretical and practical advantage of our method relative to that of Gray et al. (1973) is that our method also produces an unbiased estimator of the variance of the Monte Carlo estimator of $g(\mu)$.

Our interest in the problem lies in its relevance to problems arising in the analysis of data measured with error. Specifically, the construction of corrected score functions (Nakamura, 1990; Stefanski, 1989; Novick and Stefanski, 2002) and the identification

of deconvoluting kernels (Carroll and Hall, 1988; Stefanski and Carroll, 1990; Fan, 1991c) when the measurement error variance is unknown but replicate measurements are available. In such applications a common design is $m = 2$ replicates and we examine this case in detail. A very useful and significant advantage of the new method for dealing with measurement error is the automatic accommodation of heteroscedastic measurement errors. The new method is an alternative to the empirical simulation extrapolation method (simex) for replicate measurements studied by Devanarayan and Stefanski (2002). Main results on the fundamental estimation problem are developed in Section 2. Applications to measurement error problems are described in Section 4 and certain technical details are given in Section 5.

2 Main Results

Our assumptions are that the components of $\boldsymbol{W} = (W_1, \ldots, W_m)^T$ are iid $\mathrm{N}(\mu,\ \sigma^2)$, and $\theta = g(\mu)$ is the parameter of interest. The estimators we propose are closely related to, and motivated by, jackknife estimators and this is where we start.

2.1 The Jackknife Connection

Consider the jackknife approach to reducing bias in the maximum likelihood estimator of $g(\mu)$, $\widetilde{\theta} = g\left(\overline{W}\right)$. The leave-$b$-out version of $\widetilde{\theta}$ is

$$\widetilde{\theta}_{(-b)} = \frac{1}{\binom{m}{b}} \sum g\left(\overline{W}_{(i_1,\ldots,i_b)}\right), \qquad 1 \le b < m, \tag{1}$$

where $\overline{W}_{(i_1,\ldots,i_b)}$ is the data sample mean with $W_{i_1}, \ldots, W_{i_b}$ removed. By definition, the leave-0-out estimator is $\widetilde{\theta}$. The first-order, reduced-bias jackknife estimator, $m\widetilde{\theta} - (m-1)\widetilde{\theta}_{(-1)}$, is the intercept of the line fit to the points $\left\{(a_0,\ \widetilde{\theta}),\ (a_1,\ \widetilde{\theta}_{(-1)})\right\}$, where $a_b = 1/(m-b)$. Higher-order jackknife corrections are obtained by fitting b^{th} -order

polynomials to the pairs $\left\{(a_0,\ \widetilde{\theta}),\ (a_1,\ \widetilde{\theta}_{(-1)}),\ldots,(a_b,\ \widetilde{\theta}_{(-b)})\right\}$. The intercept again provides the reduced-bias estimator.

Because $0 < a_0 < a_1 < \cdots$, the intercepts of the polynomials used in jackknife estimation lie outside the range of the data used to fit the polynomials. Thus the jackknife estimators are obtained by extrapolating the polynomials to the origin, which can be identified as $a_{-\infty}$ in the sense that $a_{-\infty} = \lim_{b\to-\infty} a_b = 0$.

Note that in (1), $\overline{W}_{(i_1,\ldots,i_b)} = \overline{W} + \left(\overline{W}_{(i_1,\ldots,i_b)} - \overline{W}\right)$ where $\left(\overline{W}_{(i_1,\ldots,i_b)} - \overline{W}\right)$ is an unbiased estimator of 0, and thus is a contrast in $\boldsymbol{W} = (W_1,\ldots,W_m)^T$. After normalizing these contrasts to unit norm, (1) can be rewritten as

$$\widetilde{\theta}_{(-b)} = E\left\{g\left(\overline{W} + \lambda_b^{1/2} m^{-1/2} \boldsymbol{T}_b^T \boldsymbol{W}\right) \mid \boldsymbol{W}\right\} \tag{2}$$

where $\lambda_b = ma_b - 1$ and $\boldsymbol{T}_b$ is a random $m \times 1$ vector, independent of $\boldsymbol{W}$, distributed uniformly on the set of normalized contrasts $\mathcal{C}_b$ consisting of the $\binom{m}{b}$ vectors with b entries equal to $-\gamma m^{-1}$ and $(m-b)$ entries equal to $\gamma\{(m-b)^{-1} - m^{-1}\}$. The normalizing constant is $\gamma = \{m(m-b)/b\}^{1/2}$. The contrasts in $\mathcal{C}_b$ have norm 1, but are not orthogonal. Normalization of the contrasts in $\mathcal{C}_b$ essentially eliminates their dependence on a_b (although $\mathcal{C}_b$ does depend on b both in cardinality of the set and the composition of the contrasts), and thus (2) reveals that $\widehat{\theta}_{(-b)}$ is essentially a function of a_b only through $\lambda_b = ma_b - 1$. Note that $\lambda_b = -1$ when $a_b = 0$. The higher-order jackknife can now be described as fitting a b^{th} -order polynomial to the points $\left\{(\lambda_0,\ \widetilde{\theta}),\ (\lambda_1,\ \widetilde{\theta}_{(-1)}),\ldots,(\lambda_b,\ \widetilde{\theta}_{(-b)})\right\}$ and extrapolating to $\lambda_{-\infty} = -1$.

Why bother fitting polynomials and extrapolating? Equation (2) explicitly reveals the dependence of $\widehat{\theta}_{(-b)}$ on λ_b. Evaluation at $\lambda_b = -1$ results in what we call the *exact-extrapolant, jackknife "estimator,"*

$$\widetilde{\widetilde{\theta}}_{(-b)} = E\left\{g\left(\overline{W} + im^{-1/2} \boldsymbol{T}_b^T \boldsymbol{W}\right) \mid \boldsymbol{W}\right\}, \tag{3}$$

where $i = \sqrt{-1}$ (the quotation marks indicate that $\widetilde{\widetilde{\theta}}_{(-b)}$ is not real-valued and thus is not an estimator in the strict sense). The expectation of $\widetilde{\widetilde{\theta}}_{(-b)}$ is

$$E\left(\widetilde{\widetilde{\theta}}_{(-b)}\right) = \quad E\left\{g\left(\overline{W} + im^{-1/2}\boldsymbol{T}_b^T\boldsymbol{W}\right)\right\} = E\left[E\left\{g\left(\overline{W} + im^{-1/2}\boldsymbol{T}_b^T\boldsymbol{W}\right) \mid \boldsymbol{T}_b\right\}\right].$$

Consider the inner conditional expectation. Because $\boldsymbol{T}_b$ is a normalized contrast vector, the argument of the function $g()$

$$\eta \quad = \quad \overline{W} + im^{-1/2}\boldsymbol{T}_b^T\boldsymbol{W} = \mu + \left(\overline{W} - \mu\right) + im^{-1/2}\boldsymbol{T}_b^T\boldsymbol{W} = \mu + \varepsilon_1^* + i\varepsilon_2^*,$$

where $\varepsilon_1^* = \overline{W} - \mu$ and $\varepsilon_2^* = m^{-1/2}\boldsymbol{T}_b^T\boldsymbol{W}$ are conditionally independent and identically distributed $\mathrm{N}\left(0,\ \sigma^2/m\right)$ given $\boldsymbol{T}_b$. Note that the conditional distribution of ε_1^* and ε_2^* does not depend on $\boldsymbol{T}_b$. Suppose now that $g()$ is an entire function, so that we can write

$$g(\eta) = g(\mu + \varepsilon_1^* + i\varepsilon_2^*) = \sum_{k=0}^{\infty} \frac{g^{(k)}(\mu)\left\{\varepsilon_1^* + i\varepsilon_2^*\right\}^k}{k!}.$$

Assuming that the interchange of summation and expectation are justified, an appeal to Lemma 5.0.1 shows that

$$E\left\{g(\eta) \mid \boldsymbol{T}_b\right\} = \sum_{k=0}^{\infty} \frac{g^{(k)}(\mu)E\left\{\left(\varepsilon_1^* + i\varepsilon_2^*\right)^k\right\}}{k!} = g(\mu).$$

It follows that the unconditional expectation of $g(\eta)$ equals $g(\mu)$, that is, $E\left(\widetilde{\widetilde{\theta}}_{(-b)}\right) = g(\mu)$. In other words, the complex-valued, exact-extrapolant, jackknife "estimator" $\widetilde{\widetilde{\theta}}$ is unbiased for θ.

There are two problems with this "estimator." The first, that it is complex-valued, is easily solved by dropping the imaginary part, which is an unbiased estimator of 0 as the analysis above shows. The second is lack of efficiency resulting from the particular averaging in (2). Except for special cases $\mathrm{Re}\left\{\widetilde{\widetilde{\theta}}_{(-b)}\right\}$ is not a function of the data only via $\left\{\overline{W},\ \sum_j \left(W_j - \overline{W}\right)^2\right\}$, and thus need not be fully efficient. Examination of

(3) provides a clue to increasing efficiency. The only properties of the contrasts $\boldsymbol{T}_b$ used in proving unbiasedness were that they are contrasts. In other words, there is nothing special about the set of contrasts $\mathcal{C}_b$. One might expect that variation can be reduced by averaging over the largest possible set of normalized contrasts. The theorem in the next section shows this to be the case.

2.2 The First Form on the Estimator

The estimators we describe are defined in terms of uniformly distributed random contrast vectors. Such vectors can be defined, and generated, in terms of standard normal random variables. Let $Z_1, \ldots, Z_m$ be independent and identically distributed N(0, 1) independently of $W_1, \ldots, W_m$. Define

$$T_j = \frac{Z_j - \overline{Z}}{\sqrt{\sum_{j=1}^{m} \left(Z_j - \overline{Z}\right)^2}}, \qquad j = 1, \ldots, m, \tag{4}$$

where $\overline{Z}$ is the sample mean of $Z_1, \ldots, Z_m$. Then $\boldsymbol{T} = (T_1, \ldots, T_m)^T$ is a random vector uniformly distributed on $\mathcal{C}$, the set of $m \times 1$ normalized, contrast vectors.

Theorem 2.2.1 *Assume that $W_1, \ldots, W_m$ are independent and identically distributed $N(\mu,\ \sigma^2)$ with both μ and σ^2 unknown. Suppose that $g()$ is an entire function over the complex plane that also satisfies certain integrability conditions that will be made clear in the proof. Then the estimator,*

$$\widehat{\theta} = E\left\{g\left(\overline{W} + im^{-1/2}\boldsymbol{T}^T\boldsymbol{W}\right) \mid \boldsymbol{W}\right\}, \tag{5}$$

is real-valued, unbiased for $\theta = g(\mu)$, and depends on $\boldsymbol{W}$ only through the statistics $\left\{\overline{W},\ \sum_j \left(W_j - \overline{W}\right)^2\right\}$. Consequently, if $\widehat{\theta}$ has finite variance it is the uniformly minimum variance unbiased estimator of θ.

Proof. Note that the random vector $\boldsymbol{T}$ is independent of $\boldsymbol{W}$. The proof uses the facts that, almost surely,

$$\boldsymbol{T}^T\mathbf{1} = \sum T_j = 0, \qquad \text{and} \qquad \boldsymbol{T}^T\boldsymbol{T} = \sum T_j^2 = 1. \tag{6}$$

Upon taking expectations via conditioning,

$$E\left(\widehat{\theta}\right) \;=\; E\left\{g\left(\overline{W} + im^{-1/2}\boldsymbol{T}^T\boldsymbol{W}\right)\right\} = E\left[E\left\{g\left(\overline{W} + im^{-1/2}\boldsymbol{T}^T\boldsymbol{W}\right) \mid \boldsymbol{T}\right\}\right].$$

Consider the inner conditional expectation. Invoking the constraints in (6) shows that the argument of the function $g()$

$$\eta \;=\; \overline{W} + im^{-1/2}\boldsymbol{T}^T\boldsymbol{W} = \mu + \left(\overline{W} - \mu\right) + im^{-1/2}\boldsymbol{T}^T\boldsymbol{W} = \mu + \varepsilon_1^* + i\varepsilon_2^*,$$

where $\varepsilon_1^* = \overline{W} - \mu$ and $\varepsilon_2^* = m^{-1/2}\boldsymbol{T}^T\boldsymbol{W}$ are conditionally independent and identically distributed $\mathrm{N}\left(0,\ \sigma^2/m\right)$ given $\boldsymbol{T}$.

By assumption $g(\eta) = g(\mu+\varepsilon_1^*+i\varepsilon_2^*) = \sum_{k=0}^{\infty} g^{(k)}(\mu)\left\{\varepsilon_1^* + i\varepsilon_2^*\right\}^k/k!$. Assuming that the interchange of summation and expectation are justified, an appeal to Lemma 5.0.1 shows that

$$E\left\{g(\eta)|\ \boldsymbol{T}\right\} = \sum_{k=0}^{\infty} g^{(k)}(\mu)E\left\{\left(\varepsilon_1^* + i\varepsilon_2^*\right)^k\right\}/k! = g(\mu). \tag{7}$$

It follows that $E\left\{g(\eta)\right\} = g(\mu)$ and thus $E\left(\widehat{\theta}\right) = g(\mu)$.

It remains to show that $\widehat{\theta}$ is the uniformly minimum variance unbiased estimator. Again by assumption on $g()$

$$g(\eta) = g\left(\overline{W} + im^{-1/2}\boldsymbol{T}^T\boldsymbol{W}\right) = \sum_{k=0}^{\infty} \frac{g^{(k)}(\overline{W})\left\{im^{-1/2}\boldsymbol{T}^T\boldsymbol{W}\right\}^k}{k!},$$

so that

$$\widehat{\theta} \;=\; E\left\{g(\eta)|\ \boldsymbol{W}\right\} = \sum_{k=0}^{\infty} \frac{g^{(k)}(\overline{W})i^k m^{-k/2} E\left\{\left(\boldsymbol{T}^T\boldsymbol{W}\right)^k \mid \boldsymbol{W}\right\}}{k!}. \tag{8}$$

Appealing to Lemma 5.0.2 with $p = 1$ and $\boldsymbol{P_x} = m^{-1}\mathbf{1}\mathbf{1}^T$ shows that for odd k, $E\left\{(\boldsymbol{T}^T\boldsymbol{W})^k \mid \boldsymbol{W}\right\} = 0$, from which it follows that $\widehat{\theta}$ is real-valued; and for even k $E\left\{(\boldsymbol{T}^T\boldsymbol{W})^k \mid \boldsymbol{W}\right\} = Q_{m,k}\left(\sum_j \left(W_j - \overline{W}\right)^2\right)$, for some real-valued function $Q_{m,k}()$ that does not depend on $\boldsymbol{W}$, from which it follows that $\widehat{\theta}$ is a function of $\overline{W}$ and $\sum_j \left(W_j - \overline{W}\right)^2$. Completeness of the N($\mu,\ \sigma^2$) family justifies the efficiency assertion. ♠

Sufficient conditions for the interchanges of summation and expectation in (7) and (8) can be readily derived (Stefanski, 1989).

2.2.1 Monte Carlo Estimation

From equation (8) it follows that

$$\widehat{\theta} \;=\; E\left\{g(\eta)|\ \boldsymbol{W}\right\} = \sum_{k=0}^{\infty} \frac{g^{(2k)}(\overline{W})(-1)^k E\left\{(\boldsymbol{T}^T\boldsymbol{W})^{2k} \mid \boldsymbol{W}\right\}}{m^k(2k)!}. \tag{9}$$

For polynomials the sum in (9) is finite, otherwise it is generally infinite; and even for simple functions, closed form expressions, when they can be determined, generally involve special functions. For example, when $g(\mu) = \exp(\mu)$ the uniformly minimum variance unbiased estimator estimator (Neyman and Scott, 1960) is

$$\exp\left(\overline{W}\right)\Gamma\left(\frac{m-1}{2}\right) J_{(m-1)/2-1}\left(\widehat{\sigma}\sqrt{\frac{m-1}{m}}\right)\left(\frac{\widehat{\sigma}\sqrt{\frac{m-1}{m}}}{2}\right)^{1-(m-1)/2},$$

where $J_\nu(z)$ is the Bessel function, $J_\nu(z) = \sum_{k=0}^{\infty} \frac{(-1)^k}{k!\ \Gamma(\nu+k+1)}\left(\frac{z}{2}\right)^{\nu+2k}$.

The technical details of calculating $\widehat{\theta}$ can be avoided by replacing the conditional expectation in (5) by a Monte Carlo mean. Define

$$\widehat{\theta}_B = \frac{1}{B}\sum_{b=1}^{B} \mathrm{Re}\left\{g\left(\overline{W} + im^{-1/2}\boldsymbol{T}_b^T\boldsymbol{W}\right)\right\} \tag{10}$$

where Re() "real part" and the vectors $\boldsymbol{T}_b = (T_{b,1}, \ldots, T_{b,m})^T$ are independent and identically distributed Monte Carlo replicates of $\boldsymbol{T}$. These are easily constructed

from pseudo-random, standard normal variables using the relationships in (4). As indicated in the proof of the theorem the imaginary part of $g\left(\overline{W} + im^{-1/2}\boldsymbol{T}_b^T\boldsymbol{W}\right)$ has expectation 0, so that taking the real part in (10) does not affect bias and yields in a real-valued estimator.

The estimator simplifies in the special case $m = 2$. For then $T_2 = -T_1$, $\Pr(T_1 = 1/\sqrt{2}) = 1/2 = \Pr(T_1 = -1/\sqrt{2})$, and $T_1^2 = 1/2$ almost surely. In this case

$$\begin{aligned}
\widehat{\theta}_B &= \frac{1}{B}\sum_{b=1}^{B} \mathrm{Re}\left\{g\left(\overline{W} + i2^{-1/2}\left(T_{b,1}W_1 + T_{b,2}W_2\right)\right)\right\} \\
&= \frac{1}{B}\sum_{b=1}^{B} \mathrm{Re}\left\{\sum_{k=0}^{\infty} \frac{g^{(2k)}(\overline{W})}{(2k)!}\left\{im^{-1/2}T_{b,1}\left(W_1 - W_2\right)\right\}^{2k}\right\} \\
&= \frac{1}{B}\sum_{b=1}^{B} \mathrm{Re}\left\{\sum_{k=0}^{\infty} \frac{(-1)^k g^{(2k)}(\overline{W})}{4^k(2k)!}\left(W_1 - W_2\right)^{2k}\right\} \\
&= \mathrm{Re}\left\{\sum_{k=0}^{\infty} \frac{(-1)^k g^{(2k)}(\overline{W})}{4^k(2k)!}\left(W_1 - W_2\right)^{2k}\right\} \\
&= \mathrm{Re}\left\{g\left(\overline{W} + (i/2)\left(W_1 - W_2\right)\right)\right\}
\end{aligned}$$

independent of B. In other words, when $m = 2$ the uniformly minimum variance unbiased estimator can be obtained exactly in terms of $\mathrm{Re}\left\{g()\right\}$.

For example, the uniformly minimum variance unbiased estimator of $g(\mu) = \exp(\mu)$ is $\widehat{\theta} = \exp(\overline{W})\cos((W_1 - W_2)/2)$, and the uniformly minimum variance unbiased estimator of $g(\mu) = \mu^r$ is

$$\begin{aligned}
\widehat{\theta}_r &= \mathrm{Re}\left[\left\{\overline{W} + (i/2)\left(W_1 - W_2\right)\right\}^r\right] \\
&= \mathrm{Re}\left\{\sum_{k=0}^{r}\binom{r}{k}\overline{W}^{r-k}(i^k/2^k)\left(W_1 - W_2\right)^k\right\} \\
&= \sum_{k=0}^{[r/2]}\binom{r}{2k}\overline{W}^{r-2k}\frac{(-1)^k}{4^k}\left(W_1 - W_2\right)^{2k}.
\end{aligned}$$

For $r = 2,\ 3$ and 4, the estimators are $\widehat{\theta}_2 = \overline{W}^2 - (1/4)\left(W_1 - W_2\right)^2$, $\widehat{\theta}_3 = \overline{W}^3 - (3/4)\overline{W}\left(W_1 - W_2\right)^2$, and $\widehat{\theta}_4 = \overline{W}^4 - (3/2)\overline{W}^2\left(W_1 - W_2\right)^2 + (1/16)\left(W_1 - W_2\right)^4$, and their unbiasedness can be verified directly.

Note that in the case of $m = 2$, the only possible jackknife estimator (1) is the linear jackknife with $b = 1$. In this case, the set of contrasts (for $b = 1$) described in Section 2.1 is $\mathcal{C}_1 = \left\{(1/\sqrt{2},\ -1/\sqrt{2}),\ (-1/\sqrt{2},\ 1/\sqrt{2})\right\}$. So for $m = 2$, $\mathcal{C}_1$ is the largest set of possible contrasts over which to average, and thus the exact extrapolant, jackknife estimators are fully efficient for $m = 2$.

2.2.2 Extensions to Regression Models

Now assume that $\boldsymbol{W}_{m\times 1} \sim \mathrm{N}\left(\boldsymbol{X}\boldsymbol{\beta},\ \sigma^2\boldsymbol{I}_m\right)$ where $\boldsymbol{X}$ is a fixed $m \times p$ design matrix of rank p with $m \geq 2p$ and both $\boldsymbol{\beta}$ and σ^2 unknown. Suppose that the parameter of interest is a nonlinear function of a linear function of $\boldsymbol{\beta}$, say $\theta = g\left(\boldsymbol{a}^T\boldsymbol{\beta}\right)$ for some fixed, non-random $\boldsymbol{a}$. Let $\widehat{\boldsymbol{\beta}} = \left(\boldsymbol{X}^T\boldsymbol{X}\right)^{-1}\boldsymbol{X}^T\boldsymbol{W}$. Then the generalization of (5) to the regression case results in the uniformly minimum variance unbiased estimator

$$\widehat{\theta} = E\left[g\left(\boldsymbol{a}^T\widehat{\boldsymbol{\beta}} + i\boldsymbol{a}^T\left(\boldsymbol{X}^T\boldsymbol{X}\right)^{-1/2}\boldsymbol{T}^T\boldsymbol{W}\right) \mid \boldsymbol{W}\right]$$

where the random $m \times p$ matrix $\boldsymbol{T}$ can be defined (and hence generated) in terms of standard normal random variables. Specifically,

$$\boldsymbol{T} = \left(\boldsymbol{I}_m - \boldsymbol{P}_{\boldsymbol{X}}\right)\boldsymbol{Z}\left\{\boldsymbol{Z}^T\left(\boldsymbol{I}_m - \boldsymbol{P}_{\boldsymbol{X}}\right)\boldsymbol{Z}\right\}^{-1/2} \tag{11}$$

where the matrix $\boldsymbol{Z}_{m\times p}$ has iid entries $Z_{i,j} \sim \mathrm{N}(0,\ 1)$, and $\boldsymbol{P}_{\boldsymbol{X}}$ is the projection matrix of $\boldsymbol{X}$. It is shown in Lemma 5.0.2 that the matrix $\left\{\boldsymbol{Z}^T\left(\boldsymbol{I}_m - \boldsymbol{P}_{\boldsymbol{X}}\right)\boldsymbol{Z}\right\}$ is positive definite almost surely when $m \geq 2p$ so that its negative square root exists almost surely. Note also that $\boldsymbol{T}$ is such that

$$\boldsymbol{T}^T\boldsymbol{X} = \boldsymbol{0} \qquad \text{and} \qquad \boldsymbol{T}^T\boldsymbol{T} = \boldsymbol{I}_p. \tag{12}$$

The proof is virtually identical to the proof of Theorem 2.2.1.

2.3 The Second Form on the Estimator

The form of the estimator in Theorem 2.2.1 makes apparent its close relationship to jackknife estimation. However, it hides the dependence on $\overline{W}$ and $\sum_j \left(W_j - \overline{W}\right)^2$. The following result gives a simpler and more general form of the estimator.

Theorem 2.3.1 *Suppose that $\widehat{\mu}$ and $\widehat{\sigma}^2$ are independent random variables such that $\widehat{\mu} \sim N(\mu,\ \tau\sigma^2)$ and $(d\widehat{\sigma}^2/\sigma^2) \sim$ Chi-Squared(d), where τ and d are known. Define the estimator,*

$$\widehat{\theta} = E\left\{ g\left(\widehat{\mu} + i(\tau d)^{1/2}\widehat{\sigma}\left(\frac{Z_1}{\sqrt{Z_1^2 + \cdots + Z_d^2}} \right) \right) \,\middle|\, \widehat{\mu}, \widehat{\sigma}^2 \right\}, \tag{13}$$

where the components of $\boldsymbol{Z} = (Z_1, \ldots, Z_d)^T$ are independent and identically distributed $N(0,\ 1)$ independent of $\widehat{\mu}$ and $\widehat{\sigma}^2$. Then $E\left(\widehat{\theta}\right) = g(\mu)$.

Proof. Define $\boldsymbol{T} = \boldsymbol{Z}/||\boldsymbol{Z}||$. The random vector $\boldsymbol{T}$ is not a contrast but is uniformly distributed on the d-dimensional unit ball ($\boldsymbol{T}^T\boldsymbol{T} = 1$, almost surely). Let the components of $\boldsymbol{W}_* = (W_{*1}, \ldots, W_{*d})^T$ be iid N$(0,\ \sigma^2)$ random variables independent of $\widehat{\mu}$ and $\boldsymbol{Z}$. Then $(\boldsymbol{T}^T,\ d^{-1}\boldsymbol{W}_*^T\boldsymbol{W}_*,\ \widehat{\mu})^T$ and $(\boldsymbol{T}^T,\ \widehat{\sigma}^2,\ \widehat{\mu})^T$ have the same distribution. Thus it is sufficient to prove that

$$E\left\{ g\left(\widehat{\mu} + i\tau^{1/2}||\boldsymbol{W}_*||\boldsymbol{T}^T\boldsymbol{e}_1 \right) \mid \widehat{\mu},\ \boldsymbol{W}_*^T\boldsymbol{W}_* \right\}$$

has expectation $g(\mu)$, where $\boldsymbol{e}_1$ is the $d \times 1$ unit vector with a 1 in the first position.

The distribution of $\boldsymbol{T}^T\boldsymbol{v}$ is the same for any $\boldsymbol{v}$ with $||\boldsymbol{v}|| = 1$ (Watson, 1983; Lemma 5.0.2), thus $\boldsymbol{T}^T\boldsymbol{e}_1$ and $\boldsymbol{T}^T\boldsymbol{W}_*/||\boldsymbol{W}_*||$ have the same distribution, and it is sufficient to show that $\widehat{\theta}_* = E\left\{ g\left(\widehat{\mu} + i\tau^{1/2}\boldsymbol{T}^T\boldsymbol{W}_* \right) \mid \widehat{\mu},\ \boldsymbol{W}_*^T\boldsymbol{W}_* \right\}$ has mean $g(\mu)$.

Upon taking expectations and then conditioning,

$$E\left(\widehat{\theta}_*\right) \;=\; E\left\{ g\left(\widehat{\mu} + i\tau^{1/2}\boldsymbol{T}^T\boldsymbol{W}_* \right) \right\} = E\left[E\left\{ g\left(\widehat{\mu} + i\tau^{1/2}\boldsymbol{T}^T\boldsymbol{W}_* \right) \mid \boldsymbol{T} \right\} \right].$$

The argument of the function $g()$, $\eta = \widehat{\mu} + i\tau^{1/2}\boldsymbol{T}^T\boldsymbol{W}_* = \mu + \varepsilon_1^* + i\varepsilon_2^*$, where $\varepsilon_1^* = \widehat{\mu} - \mu$ and $\varepsilon_2^* = \tau^{1/2}\boldsymbol{T}^T\boldsymbol{W}_*$ are conditionally iid $\mathrm{N}(0,\ \tau\sigma^2)$ given $\boldsymbol{T}$. The proof concludes in the same fashion as the proof of Theorem 2.2.1. ♠

Note that under sufficient regularity conditions on $g()$, as $d \to \infty$ and $\widehat{\sigma} \to \sigma$ the estimator in (13) converges to $\widehat{\theta} = E\left\{ g\left(\widehat{\mu} + i\tau^{1/2}\sigma Z_1\right) \,\middle|\, \widehat{\mu} \right\}$.

2.3.1 Monte Carlo Estimation

The Monte Carlo version of (13) is simpler than (10). It is

$$\widehat{\theta}_B = \frac{1}{B}\sum_{b=1}^{B} \mathrm{Re}\left\{ g\left(\widehat{\mu} + i(\tau d)^{1/2}\widehat{\sigma}T_b\right)\right\}, \tag{14}$$

where $T_1, \ldots, T_B$ are iid replicates of the random variable $T = Z_1/\sqrt{Z_1^2 + \cdots + Z_d^2}$, and $Z_1, \ldots, Z_d$ are independent and identically distributed $\mathrm{N}(0,\ 1)$. When $d = 1$, $T^2 = 1$ almost surely and (14) is equal to $\mathrm{Re}\left\{ g\left(\widehat{\mu} + i\tau^{1/2}\widehat{\sigma}\right)\right\}$ independent of B.

3 Variance and Variance Estimation

We now derive variance expressions and variance estimators for the second form of the estimator (13) which we write as

$$\widehat{\theta} = E\left\{ g\left(\widehat{\mu} + i\left(\tau d\right)^{1/2}\widehat{\sigma}T\right) \mid \widehat{\mu},\ \widehat{\sigma}\right\}, \tag{15}$$

where $T = Z_1/\sqrt{Z_1^2 + \cdots + Z_d^2}$.

Let Y_1, Y_2 and Y_3 be real random variables. For the complex random variable, $Y_1 + iY_2$, define $\mathrm{Var}_*(Y_1 + iY_2) = E\left\{(Y_1 + iY_2)^2\right\} - \left\{E(Y_1 + iY_2)\right\}^2$ and $\mathrm{Var}_*(Y_1 + iY_2 |\, Y_3) = E\left\{(Y_1 + iY_2)^2 |\, Y_3\right\} - \left\{E(Y_1 + iY_2 |\, Y_3)\right\}^2$. It can be verified that the familiar variance decomposition holds for Var_*,

$$\mathrm{Var}_*(Y_1 + iY_2) = \mathrm{Var}_*\left\{E\left(Y_1 + iY_2 |\, Y_3\right)\right\} + E\left\{\mathrm{Var}_*(Y_1 + iY_2 |\, Y_3)\right\}. \tag{16}$$

When $Y_2 = 0$ almost surely Var_* is the usual variance operator.

When $g()$ is an entire function so is $g^2()$. Thus, subject to integrability conditions, not only does $E\left\{g\left(\widehat{\mu} + i\,(\tau d)^{1/2}\,\widehat{\sigma}T\right)\right\} = g(\mu)$, but $E\left\{g^2\left(\widehat{\mu} + i\,(\tau d)^{1/2}\,\widehat{\sigma}T\right)\right\} = g^2(\mu)$ as well. It follows that

$$\begin{aligned}\mathrm{Var}_*\left\{g\left(\widehat{\mu} + i\,(\tau d)^{1/2}\,\widehat{\sigma}T\right)\right\} &= E\left\{g^2\left(\widehat{\mu} + i\,(\tau d)^{1/2}\,\widehat{\sigma}T\right)\right\} \\ &\quad - \left[E\left\{g\left(\widehat{\mu} + i\,(\tau d)^{1/2}\,\widehat{\sigma}T\right)\right\}\right]^2 \\ &= g^2(\mu) - g^2(\mu) = 0. \end{aligned} \tag{17}$$

Combining (16), (17) and the fact that $\widehat{\theta} = E\left\{g\left(\widehat{\mu} + i\,(\tau d)^{1/2}\,\widehat{\sigma}T\right) \mid \widehat{\mu},\ \widehat{\sigma}\right\}$ is real-valued leads to the identity

$$\begin{aligned}\mathrm{Var}\left(\widehat{\theta}\right) &= -E\left[\mathrm{Var}_*\left\{g\left(\widehat{\mu} + i\,(\tau d)^{1/2}\,\widehat{\sigma}T\right) \mid \widehat{\mu}, \widehat{\sigma}^2\right\}\right] \\ &= -E\left[E\left\{g^2\left(\widehat{\mu} + i\,(\tau d)^{1/2}\,\widehat{\sigma}T\right) \mid \widehat{\mu}, \widehat{\sigma}^2\right\} - \widehat{\theta}^2\right]. \end{aligned} \tag{18}$$

Because both $\widehat{\theta}$ and $E\left\{g^2\left(\widehat{\mu} + i\,(\tau d)^{1/2}\,\widehat{\sigma}T\right) \mid \widehat{\mu}, \widehat{\sigma}^2\right\}$ are real-valued functions of $\widehat{\mu}$ and $\widehat{\sigma}^2$, it follows that

$$\widehat{\mathrm{Var}}\left(\widehat{\theta}\right) = \widehat{\theta}^2 - E\left\{g^2\left(\widehat{\mu} + i\,(\tau d)^{1/2}\,\widehat{\sigma}T\right) \mid \widehat{\mu}, \widehat{\sigma}^2\right\} \tag{19}$$

is an unbiased estimator of $\mathrm{Var}\left(\widehat{\theta}\right)$. A Monte Carlo version of this is obtained as

$$\widetilde{\mathrm{Var}}_B\left(\widehat{\theta}\right) = -\mathrm{Re}\left[\frac{1}{B-1}\sum_{b=1}^{B}\left\{g\left(\widehat{\mu} + i\,(\tau d)^{1/2}\,\widehat{\sigma}T_b\right) - \overline{g}_\bullet\right\}^2\right] \tag{20}$$

where $\overline{g}_\bullet = B^{-1}\sum_{b=1}^{B} g\left(\widehat{\mu} + i\,(\tau d)^{1/2}\,\widehat{\sigma}T_b\right)$. To see that $\widetilde{\mathrm{Var}}_B\left(\widehat{\theta}\right)$ is unbiased note that

$$\begin{aligned}E\left\{\widetilde{\mathrm{Var}}_B\left(\widehat{\theta}\right) \mid \widehat{\mu}, \widehat{\sigma}^2\right\} &= -\mathrm{Re}\left(E\left[\frac{1}{B-1}\sum_{b=1}^{B}\left\{g\left(\widehat{\mu} + i\,(\tau d)^{1/2}\,\widehat{\sigma}T_b\right) - \overline{g}_\bullet\right\}^2 \mid \widehat{\mu}, \widehat{\sigma}^2\right]\right) \\ &= -\mathrm{Re}\left[\mathrm{Var}_*\left\{g\left(\widehat{\mu} + i\,(\tau d)^{1/2}\,\widehat{\sigma}T\right) \mid \widehat{\mu}, \widehat{\sigma}^2\right\}\right], \end{aligned}$$

and thus

$$\begin{aligned} E\left\{\widetilde{\mathrm{Var}}_B\left(\widehat{\theta}\right)\right\} &= E\left(-\mathrm{Re}\left[\mathrm{Var}_*\left\{g\left(\widehat{\mu}+i\left(\tau d\right)^{1/2}\widehat{\sigma}T\right)\mid \widehat{\mu},\widehat{\sigma}^2\right\}\right]\right) \\ &= -E\left[\mathrm{Var}_*\left\{g\left(\widehat{\mu}+i\left(\tau d\right)^{1/2}\widehat{\sigma}T\right)\mid \widehat{\mu},\widehat{\sigma}^2\right\}\right] \\ &= \mathrm{Var}\left(\widehat{\theta}\right). \end{aligned}$$

Note that $\widetilde{\mathrm{Var}}_B\left(\widehat{\theta}\right)$ is an estimator of $\mathrm{Var}\left(\widehat{\theta}\right)$ and not an estimator of $\mathrm{Var}\left(\widehat{\theta}_B\right)$ where $\widehat{\theta}_B$ is defined in (14). To obtain an estimator of the latter note that

$$\begin{aligned} \mathrm{Var}\left(\widehat{\theta}_B\right) &= \mathrm{Var}\left\{E\left(\widehat{\theta}_B|\ \widehat{\mu},\widehat{\sigma}^2\right)\right\}+E\left\{\mathrm{Var}\left(\widehat{\theta}_B|\ \widehat{\mu},\widehat{\sigma}^2\right)\right\} \\ &= \mathrm{Var}\left(\widehat{\theta}\right)+E\left\{\mathrm{Var}\left(\widehat{\theta}_B|\ \widehat{\mu},\widehat{\sigma}^2\right)\right\}. \end{aligned} \tag{21}$$

The second component in (21) is unbiasedly estimated by $B^{-1}S_G^2$ where

$$S_G^2=\frac{1}{B-1}\sum_{b=1}^{B}\left(G_b-\overline{G}\right)^2$$

with $G_b=\mathrm{Re}\left\{g\left(\widehat{\mu}+i(\tau d)^{1/2}\widehat{\sigma}T_b\right)\right\}$ and $\overline{G}=B^{-1}\sum_{b=1}^{B}G_b$. Thus an unbiased estimator of $\mathrm{Var}\left(\widehat{\theta}_B\right)$ is

$$\widehat{\mathrm{Var}}_B\left(\widehat{\theta}_B\right) = \widetilde{\mathrm{Var}}_B\left(\widehat{\theta}\right)+B^{-1}S_G^2 \tag{22}$$

A simple example helps understand these unusual formulas. Consider the function $g(\mu)=\mu^2$. The facts $E(T)=E(T^3)=0$, $E(T^2)=1/d$ and $E(T^4)=3/(d^2+2d)$ are used repeatedly in the following. The estimator defined in (15) is, $\widehat{\theta}=E\left\{\left(\widehat{\mu}+i\left(\tau d\right)^{1/2}\widehat{\sigma}T\right)^2\mid \widehat{\mu},\widehat{\sigma}^2\right\}=\widehat{\mu}^2-\tau\widehat{\sigma}^2$, and the Monte Carlo version (14) of the estimator is $\widehat{\theta}_B=\widehat{\mu}^2-\tau d\widehat{\sigma}^2\left(B^{-1}\sum_{b=1}^{B}T_b^2\right)$, From (18), (19) and (20) one finds $\mathrm{Var}(\widehat{\theta})=4\mu^2\tau\sigma^2+2\tau^2\sigma^4+2\tau^2\sigma^4/d$, $\widehat{\mathrm{Var}}(\widehat{\theta})=4\widehat{\mu}^2\tau\widehat{\sigma}^2+\tau^2\widehat{\sigma}^4-3d\tau^2\widehat{\sigma}^4/(d+2)$, and $\widetilde{\mathrm{Var}}_B(\widehat{\theta})=4\widehat{\mu}^2\tau d\widehat{\sigma}^2S_T^2-\tau^2d^2\widehat{\sigma}^4S_{T^2}^2$, where S_T^2 and $S_{T^2}^2$ are the sample variances of $\{T_1,\ldots,T_B\}$ and $\{T_1^2,\ldots,T_B^2\}$. Finally, from (21) and (22) one finds $\mathrm{Var}(\widehat{\theta}_B)=\mathrm{Var}(\widehat{\theta})+2(d-1)\tau^2\sigma^4/(Bd)$ and $\widehat{\mathrm{Var}}_B(\widehat{\theta}_B)=\widetilde{\mathrm{Var}}_B(\widehat{\theta})+\tau^2d^2\widehat{\sigma}^4S_{T^2}^2/B$.

4 Applications

We now present sample applications of the basic estimation method. A simulation study of the fundamental problem of estimating a nonlinear function of a normal mean is given in Section 4.1. In Sections 4.2 and 4.3 we describe the relevance of the method, and outline its application, to regression measurement error models and deconvolution respectively. Thorough development and study of the latter two applications is beyond the scope of the current paper and will be the topic of forthcoming papers. A key feature of the current method applied to measurement error problems is the automatic handling of heteroscedastic measurement error.

4.1 Estimating $g(\mu)$

We conducted a small simulation to illustrate the performance of the Monte Carlo estimator (14) for different values of B. The function chosen for the simulation is $g(\mu) = \mu \sin(\mu) - \mu(\mu - \pi)$. Other details of the simulation design and the results are reported in Table 1. The bias in the maximum likelihood estimator is evident as is the lack of bias in the Monte Carlo estimators. The variance estimator is from (22). For this particular example there is little advantage of taking $B > 32$.

4.2 Errors-in-Variables Regression

The corrected score method is a generally applicable method for the analysis of data measured with error (Stefanski, 1989; Nakamura, 1990; Carroll et al., 1995). The Monte Carlo estimators in (10) and (14) can be used to extend the method of Monte Carlo corrected score equations (Novick and Stefanski, 2002) to the case of replicate measurements with heteroscedastic errors. We describe the application in the context of regression modelling. The new method is an alternative to, and in a certain sense an exact implementation of, the empirical simulation extrapolation method (simex)

Estimator	MLE	$B=2$	$B=8$	$B=32$	$B=128$	$B=512$
Mean	-1.60	0.01	0.01	0.01	0.01	0.01
Std Err	0.02	0.02	0.02	0.02	0.02	0.02
t_{CALC}	-93.04	0.40	0.56	0.65	0.55	0.54
MC Variance	29.65	60.44	54.11	52.52	52.03	51.92
Mean $\widehat{\text{Var}}$		60.41	54.42	51.80	51.74	51.50
Std Err $\widehat{\text{Var}}$		0.79	0.51	0.42	0.43	0.44

Table 1: Simulation design: $g(\mu) = \mu \sin(\mu) - \mu(\mu - \pi)$; $\mu = \pi$; $\theta = g(\pi) = 0$; $\tau = 1$; $\sigma^2 = 1$; $d = 10$; $\widehat{\mu} \sim \text{N}(\mu,\ \tau\sigma^2)$; $\widehat{\sigma}^2 \sim \sigma^2\chi^2(d)/d$; Number of replications = 100,000. Estimator: MLE = $g(\widehat{\mu})$; $B = \bullet$, Monte Carlo estimator (14) with $B = k$. Mean, Std Err and t_{CALC}, Monte Carlo mean, standard error, and t-statistic for testing mean = 0. MC Variance, sample variance of 100,000 replicates; Mean $\widehat{\text{Var}}$, Mean of 100,000 variance estimates calculated using (22); Std Err $\widehat{\text{Var}}$, standard error of Mean $\widehat{\text{Var}}$.

studied by Devanarayan and Stefanski (2002).

Suppose that in the absence of measurement error the regression parameter $\boldsymbol{\beta}$ is consistently estimated by the m-estimator $\widehat{\boldsymbol{\beta}}$ solving

$$\sum_{r=1}^{n} \boldsymbol{\psi}\left(Y_r, \boldsymbol{Z}_r, X_r, \boldsymbol{\beta}\right) = \mathbf{0}, \tag{23}$$

where Y_r, $\boldsymbol{Z}_r$ and X_r are a response variable, a vector of predictors measured without error and a predictor subject to measurement error respectively. Suppose that iid replicate measurements $W_{r,1}, \ldots, W_{r,m_r}$ of X_r are available such that $W_{r,j} = X_r + U_{r,j}$ where $U_{r,1}, \ldots, U_{r,m_r}$ are independent and identically distributed $\text{N}\left(0,\ \sigma_r^2\right)$. Let $\overline{W}_r$ and $\widehat{\sigma}_r^2$ denote the sample mean and variance of the components of $\boldsymbol{W}_r = (W_{r,1}, \ldots, W_{r,m_r})^T$.

Define $T_{b,r} = Z_{b,1}/\sqrt{Z_{b,1}^2 + \cdots + Z_{b,m_r-1}^2}$, and the $Z_{b,j}$ are independent and identically distributed $\text{N}(0,\ 1)$. Then provided that the components of $\boldsymbol{\psi}()$ are entire functions of their third argument, the Monte Carlo score function

$$\boldsymbol{\psi}_B\left(Y_r, \boldsymbol{Z}_r, \boldsymbol{W}_r, \boldsymbol{\beta}\right) = \frac{1}{B}\sum_{b=1}^{B} \text{Re}\left\{\boldsymbol{\psi}\left(Y_r, \boldsymbol{Z}_r, \overline{W}_r + i(m_r - 1)^{1/2}\widehat{\sigma}_r T_{b,r}, \boldsymbol{\beta}\right)\right\},$$

has the property that

$$E\left\{\boldsymbol{\psi}_B\left(Y_r, \boldsymbol{Z}_r, \boldsymbol{W}_r, \boldsymbol{\beta}\right) \mid Y_r, \boldsymbol{Z}_r, X_r\right\} = \boldsymbol{\psi}\left(Y_r, \boldsymbol{Z}_r, X_r, \boldsymbol{\beta}\right). \tag{24}$$

This is the key property of a corrected score. It follows from (24) that the system of equations $\sum_{r=0}^{n} \boldsymbol{\psi}_B\left(Y_r, \boldsymbol{Z}_r, \boldsymbol{W}_r, \boldsymbol{\beta}\right) = \mathbf{0}$, admits a consistent sequence of m-estimators; and thus asymptotic inference is relatively straight forward (Novick and Stefanski, 2002).

For example, for Poisson regression the likelihood score function, $\psi(Y, X, \boldsymbol{\beta}) = \left(Y - \exp(\boldsymbol{\beta}_1 + \boldsymbol{\beta}_x X)\right)(1,\ X)^T$, satisfies the smoothness requirements, as do score functions for regression models that are polynomial in $\boldsymbol{Z}$ and X. An important advance over the methods in Novick and Stefanski (2002) is the accommodation of heteroscedastic measurement errors (σ_r^2) with a finite number (m_r) of replicates.

As an illustration of the method for polynomials we present results from an analysis of a subset of data from the Framingham Heart Study described in Carroll et al., (1995). The variable measured with error is systolic blood pressure (SBP), for which there are two measurements available, and the response variable in our analysis is the average of two measurements of serum cholesterol (SC). Preliminary analyses suggested that the relationship of SC to SBP is nonlinear and that the error in the SBP measurements is not homoscedastic (note that we did not transform SBP as is done by Carroll et al., 1995). The results from several quadratic regressions are presented in Table 2.

We fit three so-called naive quadratic models using as predictor variable: i) the first exam measurement W_1; ii) the second exam measurement W_2; and iii) the average of the two measurements $\overline{W}$. These allowed us to construct a crude jackknife corrected estimate, $2\widehat{\boldsymbol{\beta}}_{\overline{W}} - \left(\widehat{\boldsymbol{\beta}}_{W_1} + \widehat{\boldsymbol{\beta}}_{W_2}\right)/2$, with which to compare the Monte Carlo corrected score estimate. Attenuation in the naive estimates is apparent from the

	W_1	W_2	$\overline{W}$	JK	MCCS
Intercept	1.80	1.43	1.97	2.33	2.51
	(1.09)	(1.09)	(1.09)		(1.18)
Linear	0.31	0.30	0.37	0.44	0.47
	(0.06)	(0.06)	(0.07)		(0.08)
Quadratic	-0.0046	-0.0036	-0.0056	-0.0072	-0.0081
	(0.0013)	(0.0012)	(0.0015)		(0.0020)

Table 2: Coefficient estimates from the Framingham data quadratic regression model with standard errors in parentheses. W_1, regression on Exam I measurement; W_2, regression on Exam II measurement; $\overline{W}$, regression on average; JK, jackknife bias corrected; MCCS, Monte Carlo corrected score.

trend in the latter three ($\overline{W}$, JK, MCCS) sets of regression coefficients. The Monte Carlo estimates are consistent with the crude jackknife estimates in light of their approximate (conservative) bias corrections.

4.3 An Application of Estimating $\{\sin(\mu)/\mu\}^{2k}$

We now show that our results can be used for estimating a density function from replicate measurements. The estimator is a generalization of the deconvoluting-kernel density estimator for normal measurement error studied by Carroll and Hall (1988), Fan (1991a,b,c, 1992), Hesse (1999), Stefanski (1989), Stefanski (1990), Stefanski and Carroll (1990), and Wand (1998).

Suppose that the measurements $W_{j,r}$ and true values X_r are distributed as in the previous section, with the additional assumption that the X_r are independent and identically distributed with density f_X. For a suitable constant c the function $Q(\mu) = c\{\sin(\mu)/\mu\}^{2k}$ is a probability density function. It is also an entire function to which Theorem 2.3.1 applies and thus for fixed x and $\lambda > 0$, conditionally on the

true value X_r,

$$E\left\{\frac{1}{B}\sum_{b=1}^{B}\frac{1}{\lambda}\mathrm{Re}\left(Q\left[\frac{x-\left\{\overline{W}_r+i(1-1/m_r)^{1/2}\widehat{\sigma}_r T_{b,r}\right\}}{\lambda}\right]\right)\right\}=\frac{1}{\lambda}Q\left(\frac{x-X_r}{\lambda}\right)$$

It follows that

$$\widehat{f}_X(x)=\frac{1}{n}\sum_{r=1}^{n}\frac{1}{B}\sum_{b=1}^{B}\frac{1}{\lambda}\mathrm{Re}\left(Q\left[\frac{x-\left\{\overline{W}_r+i(1-1/m_r)^{1/2}\widehat{\sigma}_r T_{b,r}\right\}}{\lambda}\right]\right) \tag{25}$$

is, conditionally on $X_1,\ldots,X_n$, an unbiased estimator of $\frac{1}{n\lambda}\sum_{r=1}^{n}Q\left(\frac{x-X_r}{\lambda}\right)$, which is a kernel density estimator of f_X constructed with the kernel Q. So $\widehat{f}_X$ is an estimator of f_X that has the same bias as a kernel density estimator based on the true data. Its construction and conditional expectation properties are essentially identical to the deconvoluting-kernel density estimators studied in the references cited above. The key advantage is that the new estimator allows for heteroscedastic measurement error. A generalization and a numerical illustration is given in the next section.

4.3.1 Analysis of Processed Data

Consider a situation where for each of $r=1,\ldots,n$ subjects a full-rank linear model $\boldsymbol{Y}_r=\boldsymbol{X}_r\boldsymbol{\beta}_r+\boldsymbol{\epsilon}_r,\ \boldsymbol{\epsilon}_r\sim\mathrm{N}(\boldsymbol{0},\ \sigma_r^2\boldsymbol{I})$, is fit resulting in the n statistics $\widehat{\boldsymbol{\beta}}_r$, $\widehat{\sigma}_r^2$ and $(\boldsymbol{X}_r^T\boldsymbol{X}_r)^{-1}$, where $\widehat{\sigma}_r^2$ is the mean squared error on $d_r>0$ degrees of freedom. Then for a fixed $\boldsymbol{a}$, $\boldsymbol{a}^T\widehat{\boldsymbol{\beta}}_r\sim\mathrm{N}\left(\boldsymbol{a}^T\boldsymbol{\beta}_r,\ \sigma_r^2\boldsymbol{a}^T\left(\boldsymbol{X}_r^T\boldsymbol{X}_r\right)^{-1}\boldsymbol{a}\right)$. Note that there are two sources of heteroscedasticity in this case, subject-specific heteroscedasticity (σ_r^2), and design-induced heteroscedasticity ($\boldsymbol{X}_r$).

Suppose now that the regression coefficients $\boldsymbol{\beta}_r$ are modelled as independent and identically distributed random vectors and interest lies in estimating the density of $\boldsymbol{a}^T\boldsymbol{\beta}_r$. Define $\tau_r=\boldsymbol{a}^T\left(\boldsymbol{X}_r^T\boldsymbol{X}_r\right)^{-1}\boldsymbol{a}$. Then appealing to Theorem 2.3.1 to generalize (25) to the case under consideration shows that

$$\widehat{f}_B(x)=\frac{1}{n}\sum_{r=1}^{n}\frac{1}{B}\sum_{b=1}^{B}\frac{1}{\lambda}\mathrm{Re}\left(Q\left[\frac{x-\left\{\boldsymbol{a}^T\widehat{\boldsymbol{\beta}}_r+i(d_r\tau_r)^{1/2}\widehat{\sigma}_r T_{b,r}\right\}}{\lambda}\right]\right) \tag{26}$$

is an estimator with the property that conditioned on $\boldsymbol{a}^T\boldsymbol{\beta}_1, \ldots, \boldsymbol{a}^T\boldsymbol{\beta}_n$

$$E\left\{\widehat{f}_B(x)\right\} = \frac{1}{n\lambda}\sum_{r=1}^{n} Q\left(\frac{x - \boldsymbol{a}^T\boldsymbol{\beta}_r}{\lambda}\right). \tag{27}$$

The expectation estimated by the r^{th} Monte Carlo mean in (26) is

$$Q^*\left(\frac{x - \boldsymbol{a}^T\widehat{\boldsymbol{\beta}}_r}{\lambda}, \lambda, d_r, \tau_r, \widehat{\sigma}_r\right)$$

where

$$Q^*\left(z, \lambda, d_r, \tau_r, \widehat{\sigma}_r\right) = \frac{\Gamma\left(\frac{d_r}{2}\right)}{2\pi}\int_{-\infty}^{\infty} e^{-itz}\left\{\frac{\theta(t)}{2}\right\}^{1-d_r/2} J_{d_r/2-1}\left(\theta(t)\right) C(t)\; dt,$$

with $\theta(t) = -it(d_r\tau_r)^{1/2}\widehat{\sigma}_r/\lambda$, and $C(t)$ is the characteristic function of $Q(t)$. Thus as $B \to \infty$ the estimator $\widehat{f}_B(x)$ converges to

$$\widehat{f}_\infty(x) = \frac{1}{n\lambda}\sum_{r=1}^{n} Q^*\left(\frac{x - \boldsymbol{a}^T\widehat{\boldsymbol{\beta}}_r}{\lambda}, \lambda, d_r, \tau_r, \widehat{\sigma}_r\right).$$

The function $Q^*\left(z, \lambda, d_r, \tau_r, \widehat{\sigma}_r\right)$ plays the same role in the current case (heteroscedastic measurement error with independent variance estimates) as the deconvoluting kernel defined in Stefanski and Carroll (1990) does in the homoscedastic measurement error, know-variance case. In fact, if $\widehat{\sigma}_r$ is set equal to σ_r, then in the limit as $d_r \to \infty$, $Q^*\left(z, \lambda, d_r, \tau_r, \widehat{\sigma}_r\right)$ converges to the deconvoluting kernel defined in Stefanski and Carroll (1990).

The estimators defined above assume heteroscedastic residual variation in the subject-specific linear models. For the case in which the assumption $\sigma_1^2 = \cdots = \sigma_n^2$ is warranted, the estimators defined above apply upon replacing $\widehat{\sigma}_r^2$ and d_r with the pooled estimator of variance, and pooled degrees of freedom for each r.

We illustrate the estimator (26) using a simulated data set of size $n = 300$. Random simple linear regression models were generated by: 1) generating a random sample size m_r uniformly distributed on $\{3,\ 4,\ 5,\ 6,\ 7\}$; 2) generating random design

points by adding N$(0, .04)$ errors to $\{1,\ 2,\ \ldots,\ m_r\}$; 3) generating a random slope from a shifted (to mean 1) and scaled (to variance 1) chi-squared(4) distribution; 4) generating Y_r according to the linear model with zero slope and random intercept, and common $\sigma_r = 2.0$ (i.e., homoscedastic residual error variation). The deconvoluting estimator of the random slopes was calculated as in (26) with $B = 50$ under the (correct) assumption of homoscedastic error variation (using the pooled estimate of variance) and under the (also correct but less precise) assumption of heteroscedastic residual error variances (using separate mean squared errors and degrees of freedom).

We used the kernel proportional to $\{\sin(x)/x\}^6$ standardized to have variance 1. Rates of convergence for deconvolution estimators with normal measurement error and known error variance are notoriously slow (Carroll and Hall, 1988), because variance increases exponentially as the bandwidth decreases. The intent of the example is to illustrate the unbiasedness property (27), and we did not attempt to estimate the bandwidth, but rather fixed it at the relatively large value $\lambda = 1.0$. A forthcoming paper will study convergence rates of (26) and examine the feasibility of deconvolution with normal errors and estimated variances in moderate sample sizes, as Wand (1998) has done for the known variance case.

Plots of estimated densities are displayed in Figure 1. The first panel contains the true density and the kernel estimator constructed using the error-free data, and makes evident the large amount of smoothing induced by the bandwidth. The second panel contains the estimator from the error-free data and the kernel estimator constructed from the error-contaminated data. The third and fourth panels display the two estimates from the second panel and the deconvolution estimates calculated under the assumptions of heteroscedastic and homoscedastic errors respectively.

We also calculated estimated moments, mean, variance, third and fourth central moments, for these simulated data using the true data, the observed data, and the

	True Data	Observed Data	Heteroscedastic	Homoscedastic
Mean	1.05	1.09	1.09	1.09
Variance	1.22	1.66	1.27	1.24
Third	2.30	2.20	2.22	2.18
Fourth	10.06	13.08	9.89	9.41

Table 3: Estimated moments of the slope from the simulated random coefficient model data.

m-estimation method described in Section 4.2 (in this case there is no response Y_r or error-free covariate $\boldsymbol{Z}_r$, and $\boldsymbol{\psi}$ is chosen to yield the indicated moments). The estimates are given in Table 3.

4.3.2 Comments on the Examples

A notable feature of the examples presented above which is not apparent in the results is the ease with which the Monte Carlo estimates were obtained. We exploited the complex number capabilities available in GAUSS to use the same code that would be used for standard analyses, evaluating functions at the complex arguments as indicated, and invoking GAUSS' "real-part" function as necessary. For the regression model example, the estimating equations were passed to an equation solver that provided solutions and sandwich-formula, asymptotic variance matrix estimates.

4.4 Robustness

In applications robustness is often an issue. The estimators we propose depend on normality and the smoothness assumptions on g. Novick and Stefanski (2002) provided evidence of robustness for a similar class of estimators. Whether our estimators possess a similar level of robustness is an open question that we plan to address in forthcoming papers that develop the applications in greater detail.

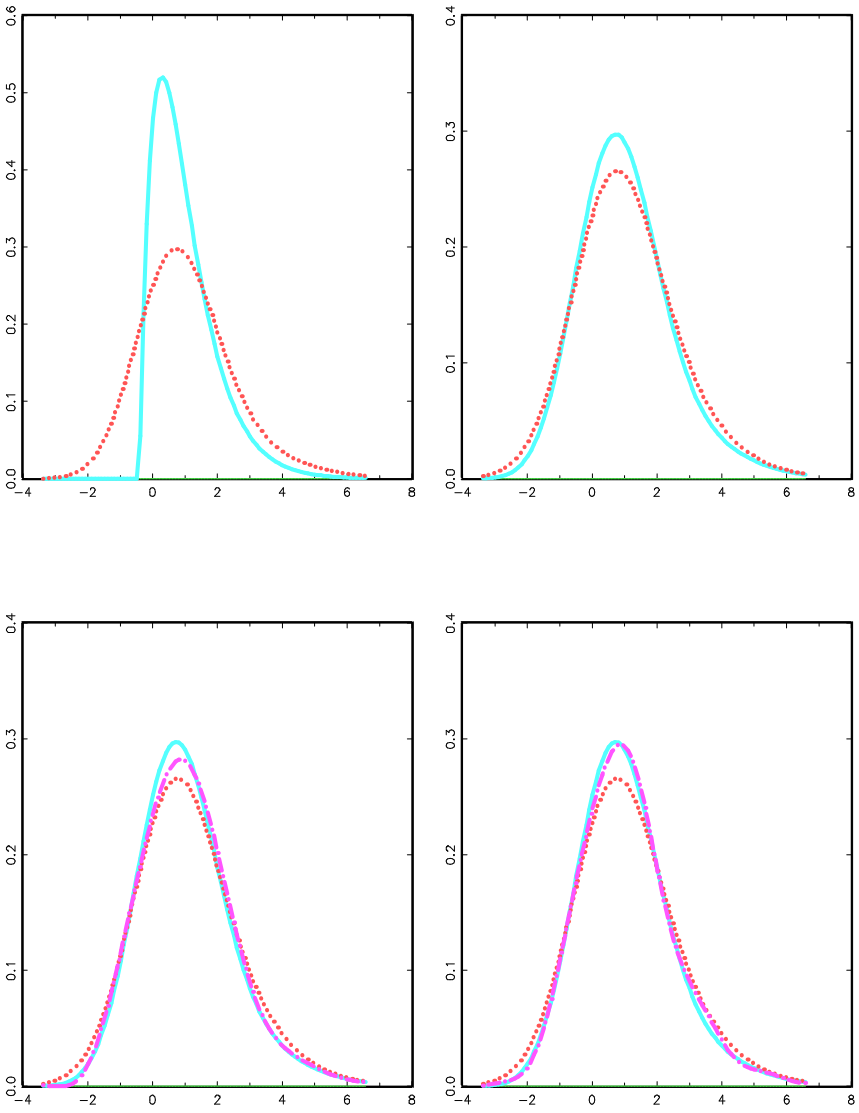

Figure 1: Density Estimates. Panel 1: solid, true density; dotted, true-data estimate. Panel 2: solid, true-data estimate; dotted, observed-data estimate. Panel 3: solid, true-data estimate; dotted, observed-data estimate; dot-dash, deconvolution estimate assuming heteroscedasticity. Panel 4: solid, true-data estimate; dotted, observed-data estimate; dot-dash, deconvolution estimate assuming homoscedasticity. Note that Panels 2-4 are scaled differently than Panel 1.

5 Appendix: Technical Details

Lemma 5.0.1 *If ε_1 and ε_2 are iid $N(0,\ \tau^2)$, then $E\left\{(\varepsilon_1 + i\varepsilon_2)^k\right\} = 0$, for $k = 1, 2, 3, \ldots$.*

Proof. The moment generating function of $\varepsilon_1 + i\varepsilon_2$ is equal to 1 and thus all of its derivatives are 0. ♠

Lemma 5.0.2 *Let $\boldsymbol{Z}$, $\boldsymbol{T}$ and $\boldsymbol{P_X}$ be defined as in (11). Then: (i) $\boldsymbol{Z}^T(\boldsymbol{I}_m - \boldsymbol{P_X})\boldsymbol{Z}$ is positive definite almost surely if and only if $m \geq 2p$; (ii) for fixed non-random vectors $\boldsymbol{u}$ and $\boldsymbol{s}$ the distribution of $\boldsymbol{u}^T\boldsymbol{T}^T\boldsymbol{s}$ is symmetric, possesses moments of all orders, and depends on $\boldsymbol{s}$ only through $\boldsymbol{s}^T(\boldsymbol{I}_m - \boldsymbol{P_X})\boldsymbol{s}$.*

Proof. Because $\text{rank}\,(\boldsymbol{Z}^T(\boldsymbol{I}_m - \boldsymbol{P_X})\boldsymbol{Z}) \leq \text{rank}\,(\boldsymbol{I}_m - \boldsymbol{P_X}) = m - p,\ m - p \geq p$ is necessary. Now assume that $m \geq 2p$. Write $(\boldsymbol{I}_m - \boldsymbol{P_X}) = \boldsymbol{Q}\boldsymbol{D}\boldsymbol{Q}^T$, where $\boldsymbol{Q}$ is orthogonal and $\boldsymbol{D}$ is diagonal with first $m - p$ entries equal to 1, and last p entries equal to 0. Because $\boldsymbol{Q}^T\boldsymbol{Z}$ has the same distribution as $\boldsymbol{Z}$, $(\boldsymbol{Z}^T(\boldsymbol{I}_m - \boldsymbol{P_X})\boldsymbol{Z})$ and $\boldsymbol{Z}^T\boldsymbol{D}\boldsymbol{Z}$ are identically distributed. The latter is equal to $\sum_{i=1}^{m-p}\boldsymbol{r}_i\boldsymbol{r}_i^T$ where $\boldsymbol{r}_i^T$ is the i^{th} row of $\boldsymbol{Z}$. The matrix $\sum_{i=1}^{m-p}\boldsymbol{r}_i\boldsymbol{r}_i^T$ has a Wishart distribution and is positive definite almost surely provided $m - p \geq p$ (Srivastava and Khatri, 1979, p 72).

Symmetry of the distribution of $\boldsymbol{u}^T\boldsymbol{T}^T\boldsymbol{s}$ follows from the fact that

$$\boldsymbol{u}^T\boldsymbol{T}^T\boldsymbol{s} = \boldsymbol{u}^T\left\{\boldsymbol{Z}^T(\boldsymbol{I}_m - \boldsymbol{P_X})\boldsymbol{Z}\right\}^{-1/2}\boldsymbol{Z}^T(\boldsymbol{I}_m - \boldsymbol{P_X})\boldsymbol{s} = h(\boldsymbol{Z}), \tag{28}$$

where $h(-\boldsymbol{z}) = -h(\boldsymbol{z})$ and $h(\boldsymbol{Z})$ and $h(-\boldsymbol{Z})$ have the same distribution. From the second property in (12), $|\boldsymbol{u}^T\boldsymbol{T}^T\boldsymbol{s}| \leq ||\boldsymbol{u}||||\boldsymbol{s}||$ from which it follows that all moments of $\boldsymbol{u}^T\boldsymbol{T}^T\boldsymbol{s}$ exist.

It remains to show that the distribution of $\boldsymbol{u}^T\boldsymbol{T}^T\boldsymbol{s}$ depends on $\boldsymbol{s}$ only through $\boldsymbol{s}^T(\boldsymbol{I}_m - \boldsymbol{P_X})\boldsymbol{s}$. Note that by (28) if $\boldsymbol{s}$ is in the column space of $\boldsymbol{X}$ then $\boldsymbol{u}^T\boldsymbol{T}^T\boldsymbol{s} =$

0 almost surely and the claim is trivially true. Now consider the case in which $(\boldsymbol{I}_m - \boldsymbol{P}_{\boldsymbol{X}})\,\boldsymbol{s} \neq \boldsymbol{0}$. Define

$$\boldsymbol{P}_{\dagger} = \frac{(\boldsymbol{I}_m - \boldsymbol{P}_{\boldsymbol{X}})\,\boldsymbol{s}\boldsymbol{s}^T\,(\boldsymbol{I}_m - \boldsymbol{P}_{\boldsymbol{X}})}{||\,(\boldsymbol{I}_m - \boldsymbol{P}_{\boldsymbol{X}})\,\boldsymbol{s}||^2}.$$

From the decomposition $(\boldsymbol{I}_m - \boldsymbol{P}_{\boldsymbol{X}})\,\boldsymbol{Z} = \boldsymbol{P}_{\dagger}\,(\boldsymbol{I}_m - \boldsymbol{P}_{\boldsymbol{X}})\,\boldsymbol{Z} + (\boldsymbol{I}_m - \boldsymbol{P}_{\dagger})\,(\boldsymbol{I}_m - \boldsymbol{P}_{\boldsymbol{X}})\,\boldsymbol{Z}$, it follows that $\boldsymbol{Z}^T\,(\boldsymbol{I}_m - \boldsymbol{P}_{\boldsymbol{X}})\,\boldsymbol{Z} = \boldsymbol{Z}_1^T\boldsymbol{Z}_1 + \boldsymbol{Z}_2^T\boldsymbol{Z}_2$ where $\boldsymbol{Z}_2 = (I_m - \boldsymbol{P}_{\dagger})\,(I_m - \boldsymbol{P}_{\boldsymbol{X}})\,\boldsymbol{Z}$, and $\boldsymbol{Z}_1 = \boldsymbol{s}^T\,(\boldsymbol{I}_m - \boldsymbol{P}_{\boldsymbol{X}})\,\boldsymbol{Z}/||\,(I_m - \boldsymbol{P}_{\boldsymbol{X}})\,\boldsymbol{s}||$. Thus

$$\boldsymbol{s}^T\boldsymbol{T} = ||\,(\boldsymbol{I}_m - \boldsymbol{P}_{\boldsymbol{X}})\,\boldsymbol{s}||\boldsymbol{Z}_1\,(\boldsymbol{Z}_1^T\boldsymbol{Z}_1 + \boldsymbol{Z}_2^T\boldsymbol{Z}_2)^{-1/2}\,. \tag{29}$$

We now show that: $\boldsymbol{Z}_1^T$ has a p-variate standard normal distribution; $\boldsymbol{Z}_1^T$ and $\boldsymbol{Z}_2^T$ are independent; and the distribution of $\boldsymbol{Z}_2^T\boldsymbol{Z}_2$ does not depend on $\boldsymbol{s}$. It follows that the distribution of $\boldsymbol{s}^T\boldsymbol{T}$ depends on $\boldsymbol{s}$ only through the scalar multiplier $||\,(\boldsymbol{I}_m - \boldsymbol{P}_{\boldsymbol{X}})\,\boldsymbol{s}||$ in (29).

Note that $\boldsymbol{Z}_1^T = \boldsymbol{Z}^T\boldsymbol{v}/||\boldsymbol{v}||$ where $\boldsymbol{v} = (\boldsymbol{I}_m - \boldsymbol{P}_{\boldsymbol{X}})\,\boldsymbol{s}$. Because the rows of $\boldsymbol{Z}^T$ are iid it follows that the rows of $\boldsymbol{Z}_1^T$ are also iid. Let $\boldsymbol{c}_1^T$ denote the first row of $\boldsymbol{Z}^T$. Then the first component of $\boldsymbol{Z}_1^T$ is $\boldsymbol{c}_1^T\boldsymbol{v}/||\boldsymbol{v}|| \sim \mathrm{N}(0,\ 1)$.

Independence of the rows of $\boldsymbol{Z}^T$ implies that the i^{th} row of $\boldsymbol{Z}_1^T$ is independent of the j^{th} row of $\boldsymbol{Z}_2^T$ for $i \neq j$. Also the rows are identically distributed so it is only necessary to establish that the first rows of $\boldsymbol{Z}_1^T$ and $\boldsymbol{Z}_2^T$ are independent. These are $\boldsymbol{c}_1^T\,(\boldsymbol{I}_m - \boldsymbol{P}_{\boldsymbol{X}})\,\boldsymbol{s}/||\boldsymbol{v}||$ and $\boldsymbol{c}_1^T\,(\boldsymbol{I}_m - \boldsymbol{P}_{\boldsymbol{X}})\,(\boldsymbol{I}_m - \boldsymbol{P}_{\dagger})$ respectively, which are uncorrelated and hence independent.

The matrix $\boldsymbol{Z}_2^T\boldsymbol{Z}_2 = \boldsymbol{Z}^T\boldsymbol{M}\boldsymbol{Z}$ where $\boldsymbol{M} = (\boldsymbol{I}_m - \boldsymbol{P}_{\boldsymbol{X}})\,(\boldsymbol{I}_m - \boldsymbol{P}_{\dagger})\,(\boldsymbol{I}_m - \boldsymbol{P}_{\boldsymbol{X}})$. By the spectral composition we can find a diagonal matrix $\boldsymbol{D}_*$ with $0-1$ entries such that $\boldsymbol{M} = \boldsymbol{Q}_*\boldsymbol{D}_*\boldsymbol{Q}_*^T$ where $\boldsymbol{Q}_*^T$ is orthonormal and thus $\boldsymbol{Q}_*^T\boldsymbol{Z}$ and $\boldsymbol{Z}$ are identically distributed. It follows that $\boldsymbol{Z}_2^T\boldsymbol{Z}_2$ and $\boldsymbol{Z}^T\boldsymbol{D}_*\boldsymbol{Z}$ are identically distributed and the distribution of the latter obviously does not depend on $\boldsymbol{s}$. ♠

When $p = 1$ and $\boldsymbol{P_x} = m^{-1}\mathbf{1}\mathbf{1}^T$, the vector $\boldsymbol{T} = (T_1, \ldots, T_m)^T$ is such that $\sum T_j = 0$ and $\sum T_j^2 = 1$. Thus $\boldsymbol{T}$ lies on the intersection of an $(m-1)$-dimensional hyperplane and the m-dimensional unit ball which is an $(m-1)$-dimensional ball. In this case $\boldsymbol{T}$ is equal in distribution to $\boldsymbol{A}\boldsymbol{T}_*$ where $\boldsymbol{A}_{m\times(m-1)}$ is a matrix such that $\boldsymbol{A}\boldsymbol{A}^T = \boldsymbol{I}_m - m^{-1}\mathbf{1}\mathbf{1}^T$ and $\boldsymbol{A}^T\boldsymbol{A} = \boldsymbol{I}_{m-1}$, and $\boldsymbol{T}_*$ is uniformly distributed on the $(m-1)$-dimensional unit ball. So $\boldsymbol{T}^T\boldsymbol{s}$ and $\boldsymbol{T}_*^T\boldsymbol{A}^T\boldsymbol{s}$ have the same distribution and the conclusions of Lemma 5.0.2 can be deduced from known properties of the distribution of $\boldsymbol{T}_*$. (Watson, 1983, p 45).

REFERENCES

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

Carroll, R. J., Ruppert, D. and Stefanski, L. A. (1995). *Measurement Error in Nonlinear Models*, London: Chapman Hall.

Devanarayan, V. & Stefanski, L. A. (2002), Empirical Simulation Extrapolation for Measurement error Models with Replicate Measurements, *Statistics and Probability Letters*, 59, 219-225.

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

Fan, J. (1991b). Asymptotic normality for deconvolving kernel density estimators, *Sankhyā, Series A*, 53, 97-110.

Fan, J. (1991c). Global behavior of deconvolution kernel estimates. *Statistica Sinica*, 1, 541-551.

Fan, J. (1992). Deconvolution with supersmooth distributions. *Canadian Journal of Statistics*, 20, 23-37.

Gray H. L., Watkins, T. A., and Schucany W. R., (1973). On the Jackknife Statistic and its Relation to UMVU Estimators in the Normal Case, *Communications in Statistics*, 2(4), 285-320.

Hesse, C. H. (1999). Data-driven deconvolution. *Journal of Nonparametric Statistics*, 10, 343-373.

Kitagawa, T. (1956). The operational calculus and the estimations of functions or parameters admitting sufficient statistics, *Bulletin of Mathematical Statistics*, 6, 95-108.

Kolmogorov, A. N. (1950). Unbiased estimates, *Izvestiya Akademii Nauk SSSR, Ser. Mat.*, 14, 303-326; see also *Am. Math. Soc. Translation* No. 98.

Nakamura, T. (1990). Corrected score functions for errors-in-variables models: methodology and application to generalized linear models. *Biometrika*, 77, 127-137.

Neyman, J., and Scott, E. L. (1960). Corrections for bias introduced by a transformation of variables, *Annals of Mathematical Statistics*, 31, 643-655.

Novick, S. J. and Stefanski, L. A. (2002). Corrected Score Estimation via Complex Simulation Extrapolation, *Journal of the American Statistical Association*, 97, 472-481.

Srivastava, M. S. and Khatri, C. G. (1979). *An Introduction to Multivariate Statistics*, North Holland, New York.

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

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

Stefanski, L. A. & Carroll, R. J. (1990). Deconvoluting kernel density estimators. *Statistics*, 21, 165-184.

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

Washio, Y., Morimoto, H., and Ikeda, N. (1956). Unbiased estimation based on sufficient statistics, *Bulletin of mathematical Statistics*, 6, 69-93.

Watkins T. A. (1979). Estimating A Function Of The Mean Of A Multivariate Normal-Distribution, *Communications in Statistics A, Theory*, 8, 245-256.

Watson, Geoffrey S. (1983). *Statistics on Spheres*, John Wiley & Sons, Inc., New

York.

Woodward, W. A. (1977). Subroutines For Computing Minimum Variance Unbiased Estimators Of Functions Of Parameters In Normal And Gamma Distributions, *Communications in Statistics B, Simulation*, 6, 63-73.