THE INSTITUTE OF STATISTICS

SIMULATION-EXTRAPOLATION:
THE MEASUREMENT ERROR JACKKNIFE

by

L. A. Stefanski and J. R. Cook

Institute of Statistics Mimeograph Series No. 2259

November 1993

NORTH CAROLINA STATE UNIVERSITY
Raleigh, North Carolina

SIMULATION-EXTRAPOLATION:
THE
MEASUREMENT ERROR JACKKNIFE

L. A. Stefanski
Department of Statistics
North Carolina State University
Raleigh, NC 27695

J. R. Cook
Merck Research Laboratories
Division of Merck & Co., Inc.
West Point, PA 19422

ABSTRACT

This paper provides theoretical support for the simulation-based estimation procedure, SIMEX, introduced by Cook and Stefanski (1992) for measurement error models. We do so by establishing a strong relationship between SIMEX estimation and jackknife estimation. A resultant of our investigations is the identification of a variance estimation method for SIMEX that parallels jackknife variance estimation. It is shown that the variance estimator is asymptotically valid in simple linear regression measurement error models. Data from the Framingham Heart Study are used to illustrate the variance estimation procedure in logistic regression measurement error models.

Keywords: Bias Reduction; Components of Variance; Extrapolation; Linear Regression; Logistic Regression; Simulation; UMVUE; Variance Estimation.

Note: This paper uses data supplied by the National Heart, Lung, and Blood Institute, NIH, and DHHS from the Framingham Heart Study. The views expressed in this paper are those of the authors and do not necessarily reflect the views of the National Heart, Lung, and Blood Institute, or of the Framingham Study.

1. INTRODUCTION

Cook and Stefanski (1992) introduced a simulation-based method of estimation for measurement error models, called SIMEX, wherein a *simulation* step is followed by a modelling and *extrapolation* step. Exact computations for linear models and simulation studies for nonlinear models were presented showing that SIMEX is competitive with other more conventional methods of analyzing data contaminated with measurement errors.

By establishing a close relationship between SIMEX and jackknife estimation, this paper provides further theoretical justification for SIMEX and identifies a complementary method of variance estimation. The relationship is such that SIMEX can be viewed as a specialized adaptation of the jackknife to measurement error models.

The problem with ignoring measurement errors in data is that parameter estimates so obtained are generally biased whenever the estimates are nonlinear functions of the variates measured with error. The jackknife (Quenouille, 1956) is a technique that was developed for the purpose of reducing bias in nonlinear estimators. Despite the apparent connection between these problems, there has been relatively little mention of the jackknife in the literature on measurement error methods.

In Sections 2 and 3 we review jackknife and SIMEX estimation, using for illustration a common example that is amenable to both methods. Section 4 digresses somewhat to address the extrapolation component of SIMEX estimation. Section 5 establishes the formal connection between SIMEX and the jackknife and studies the performance of SIMEX estimation in a class of problems that have typically been handled by the jackknife. Also in this section we describe a variance estimation method for SIMEX that parallels jackknife variance estimation.

Section 6 describes the implementation of the variance estimation method to measurement error models. We prove its validity for the simple, but representative, components-of-variance model, and also for the more complicated linear regression measurement error model. Section 6 ends with an application to a logistic regression measurement error model. Concluding remarks are given in Section 7.

Throughout the paper we assume the simple, additive measurement error model, often with known measurement error variance. The utility of this simple model for both applications and theoretical study is evident by the attention it has received in linear measurement error models, Fuller (1987).

2. JACKKNIFE EXTRAPOLATION

In order to present Quenouille's jackknife (Quenouille, 1956) as an extrapolation technique, we consider the problem of estimating $\theta = f(\mu)$, f nonlinear, with independent and identically distributed data, $\{X_j\}_1^n$, from a Gaussian population with mean μ and known variance σ^2. We take $f(\mu) = \exp(\mu)$ in this section, generalizing later to more general functions. This particular example was chosen primarily for its analytic tractibility. The results in this section do not depend on knowing σ^2, but results in later sections do.

It is well known that the maximum likelihood estimator, $\hat{\theta} = \exp(\bar{X})$, has expectation $E(\hat{\theta}) = \theta \exp\{\sigma^2/(2n)\}$, and thus is positively biased by the amount $[\exp\{\sigma^2/(2n)\} - 1]\theta$. The bias is of order n^{-1}, and, to terms of order n^{-3}, the bias is $\theta\left\{n^{-1}(\sigma^2/2) + n^{-2}(\sigma^4/8) + n^{-3}(\sigma^6/48)\right\}$.

The reduced-bias jackknife estimator is $n\hat{\theta} - (n-1)\bar{\hat{\theta}}_{(n-1),(\cdot)}$ where $\bar{\hat{\theta}}_{(n-1),(\cdot)}$ is the average of the n leave-1-out estimators $\hat{\theta}_{(n-1),(k)}$, $k = 1, \ldots, n$, based on samples of size $n - 1$.

Later we will make use of the notation $\bar{\hat{\theta}}_{(n-j),(\cdot)}$, to denote the average of the ${}_nC_j$ leave-j-out estimators $\hat{\theta}_{(n-j),(k)}$, $k = 1, \ldots, {}_nC_j$, based on samples of size $n - j$. This notation is further extended to denote both the maximum likelihood estimator, $\bar{\hat{\theta}}_{(n),(\cdot)}$, and the jackknife estimator, $\bar{\hat{\theta}}_{(\infty),(\cdot)}$. The maximum likelihood estimator is a leave-0-out estimator, and the jackknife estimator can be thought of, for reasons that will become evident in following sections, as a leave-$(-\infty)$-out estimator, hence the notation.

The jackknife estimator of $\theta = \exp(\mu)$ has a negative bias of order n^{-2}, and to terms of order n^{-3} is $-\theta\left\{n^{-2}(\sigma^4/8) + n^{-3}(\sigma^4/8 + \sigma^6/24)\right\}$.

Figure 1 illustrates the nature of the extrapolation involved. For this figure a pseudo-random, standard normal sample of size 4, $\{X_1 = -0.20544,\ X_2 = 0.33879,\ X_3 = 1.39088,\ X_4 = -1.02414\}$, was generated subject to certain constraints explained in Section 3.2. For these data $\mu = 0$ and thus $\theta = 1$. Plotted are the points $\left(\bar{\hat{\theta}}_{(k),(\cdot)},\ 1/k\right)$ for $k = 3$ (the average of the 4 leave-1-out estimates), $k = 4$ (the maximum likelihood estimate), and $k = \infty$ (the jackknife estimate). It is evident that the jackknife estimate is a simple linear extrapolation (the dashed line) of the line determined by the maximum likelihood estimate and the average of the leave-1-out estimates on an inverse sample size scale. Since $\theta = 1$, Figure 1 illustrates both the positive bias of the maximum likelihood estimator and the negative bias of the jackknife estimator. Our Figure 1 is similar to a figure in Efron (1982, Ch. 2, p. 7).

3. SIMULATION EXTRAPOLATION

3.1. Simulation Extrapolation Estimation

SIMEX estimation is a computational and graphical method-of-moments-like inference procedure for assessing and reducing measurement error-induced bias. The method starts with the observed data and the so-called naive estimator, *i.e.*, the estimator computed from the observed data without regard to the measurement errors. Estimators with yet greater bias are obtained by adding pseudo-random measurement errors to the data in a resampling-like stage and recalculating the naive estimator again from the contaminated data. These estimators are used to establish a trend of measurement error-induced bias versus the variance of the added measurement error. This trend is then extrapolated back to the case of no measurement error, producing estimators with less bias, and in some cases no bias, at least asymptotically. Details of the method and example applications are presented in Cook and Stefanski (1992).

We adopt notation consistent with Cook and Stefanski (1992). We use U to denote the true variate subject to measurement error (U is for unobserved), X to denote the measurement of U, and Y and V to denote response and covariable variates respectively, should the latter be present in the data. We assume initially that $X_j = U_j + \sigma Z_j, \quad j = 1, \ldots, n$, where $\{U_j\}_1^n$ is a sequence of unknown constants and $\{Z_j\}_1^n$ is a sequence of independent standard normal random variables, *i.e.*, a functional measurement error model with additive independent normal error.

The estimator that would be calculated in the absence of measurement error is denoted $\hat{\theta}_{\text{TRUE}} = T(\{Y_j, V_j, U_j\}_1^n)$, where T represents the function applied to the data, *e.g.*, for linear regression T would commonly represent least-squares estimation. The so-called naive estimator, *i.e.*, the estimator calculated if measurement error is ignored, is obtained by replacing U_j with X_j above, and is denoted $\hat{\theta}_{\text{NAIVE}} = T(\{Y_j, V_j, X_j\}_1^n)$.

The pseudo data constructed by adding random errors to $\{X_j\}_1^n$ are denoted

$$X_{b,j}(\lambda) = X_j + \lambda^{1/2}\sigma Z_{b,j}, \qquad j = 1, \ldots, n, \quad b = 1, \ldots, B,$$

where the added pseudo errors, $\{\{Z_{b,j}\}_{j=1}^n\}_{b=1}^B$ are mutually independent, independent of the data $\{Y_j, V_j, X_j\}_1^n$, and identically distributed, standard normal random variables.

The pseudo data are used to calculate

$$\hat{\theta}_b(\lambda) = T\left(\{Y_j, V_j, X_{b,j}(\lambda)\}_{j=1}^n\right), \qquad b = 1, \ldots, B, \tag{3.1}$$

which in turn are used to estimate, by averaging,

$$\hat{\theta}(\lambda) = E\left\{\hat{\theta}_b(\lambda) \mid \{Y_j, V_j, X_j\}_1^n\right\}. \tag{3.2}$$

The expectation in (3.2) is with respect to the distribution of $\{Z_{b,j}\}_{j=1}^n$ only.

Note that $\hat{\theta}(0) = \hat{\theta}_b(0) = \hat{\theta}_{\text{NAIVE}}$. Although analytic determination of $\hat{\theta}(\lambda)$ for $\lambda > 0$ is sometimes possible, it can always be estimated with arbitrarily small variance by generating a large number, B, of independent measurement error vectors, $\{\{Z_{b,j}\}_{j=1}^n\}_{b=1}^B$, computing the pseudo estimator $\hat{\theta}_b(\lambda)$ for each $b = 1, \ldots, B$, and estimating $\hat{\theta}(\lambda)$ by the sample mean of $\{\hat{\theta}_b(\lambda)\}_1^B$. This is the simulation component of SIMEX.

The extrapolation step of SIMEX entails modelling $\hat{\theta}(\lambda)$ as a function of λ for $\lambda \geq 0$ and extrapolating the model back to $\lambda = -1$. This results in the simulation-extrapolation estimator denoted $\hat{\theta}_{\text{SIMEX}}$. The procedure is illustrated in the next section.

3.2. Estimating $\theta = \exp(\mu)$ via Simulation Extrapolation

SIMEX estimation applies to the problem studied in Section 2 under the interpretation that X_j is a measurement of μ with measurement error variance σ^2. Adding additional measurement error with variance $\lambda\sigma^2$ to $X_1, \ldots, X_n$, results in the estimator $\hat{\theta}_b(\lambda) = \exp(\bar{X} + \sigma\sqrt{\lambda}\,\bar{Z}_b)$ where $\sigma\sqrt{\lambda}\bar{Z}_b$ is the mean of the added pseudo errors. Doing this for $b = 1, \ldots, B$, averaging the estimators, and letting $B \to \infty$ conceptually, results in

$$\hat{\theta}(\lambda) = \lim_{B\to\infty} \frac{1}{B} \sum_{b=1}^{B} \exp(\bar{X} + \sigma\sqrt{\lambda}\,\bar{Z}_b) = E\{\hat{\theta}_b(\lambda) \mid \bar{X}\} = \exp(\bar{X} + \lambda\sigma^2/(2n)),$$

where the limit is justified by an appeal to the Strong Law of Large Numbers, and the expectation can be found using moment generating function-like identities.

The functional form of $\hat{\theta}(\lambda)$ is known in this example only because the required expectations can be obtained analytically. The measurement error analyst will always have available $\{(\hat{\theta}(\lambda),\ \lambda) : \lambda > 0\}$, but will seldom know its functional form. Thus for this example we proceed under the contrived assumption of ignorance of the functional form of $\hat{\theta}(\lambda)$.

SIMEX next calls for plotting $\hat{\theta}(\lambda)$ versus λ for $\lambda > 0$, fitting a model to this curve and extrapolating the curve back to $\lambda = -1$, thereby obtaining $\hat{\theta}_{\text{SIMEX}}$. Upon examining the plot of $\hat{\theta}(\lambda)$ versus λ or by looking at residuals from the fit of a linear model, most data analysts would arrive at the realization that $\log(\hat{\theta}(\lambda))$ is *exactly* linear in λ and hence determine the exact functional form. Thus for this example SIMEX results in the estimator $\hat{\theta}_{\text{SIMEX}} = \hat{\theta}(-1) = \exp(\bar{X} - \sigma^2/(2n))$.

Figure 2 contains a plot of $\ln(\hat{\theta}(\lambda))$ versus λ for $0 \leq \lambda \leq 2$ (solid line), and the extrapolation to $\lambda = -1$ (dashed line), for the four-point data set set used in Section 2. The plotted points are for an equally-spaced grid of eleven λ values spanning [0, 2], and $\ln(\hat{\theta}_{\text{SIMEX}})$ at $\lambda = -1$. There is nothing special about the number or range of λ values employed in this example, except that both are large enough to provide convincing evidence of linearity. It transpires that for these data $\hat{\theta}_{\text{SIMEX}} = 1$, *i.e.*, exactly equal to θ. Thus Figure 2 illustrates both the positive bias of the maximum likelihood estimate, $\hat{\theta}(0)$, and the apparent unbiasedness of the SIMEX estimator.

It is time to confess. The four-point data set in Section 2 was generated subject to the restrictions that both $\bar{\hat{\theta}}_{(n),(\cdot)}$ and $\bar{\hat{\theta}}_{(n-1),(\cdot)}$, equal their expectations. Thus Figures 1 and 2 depict exact *expected* behaviour of the maximum likelihood, jackknife and SIMEX estimators.

Indeed, a simple calculation shows that $\hat{\theta}_{\mathrm{SIMEX}}$ is an unbiased estimator of θ in this problem. In fact, for what it's worth, $\hat{\theta}_{\mathrm{SIMEX}}$ is the uniform minimum variance unbiased estimator of θ. This is not a coincidence. If the naive estimator is a function of a sufficient statistic, then so too will be the SIMEX estimator. If the latter is unbiased, then it is necessarily uniform minimum variance unbiased.

Lest we be accused of painting too rosey a picture of SIMEX with this example, we emphasize that the unbiasedness of $\hat{\theta}_{\mathrm{SIMEX}}$ depends crucially on knowing σ^2, being able to determine the functional form of $\hat{\theta}(\lambda)$ exactly, and normality of the data. For simple problems $\hat{\theta}(\lambda)$ often has a simple functional form, and a little exploratory data analysis will often reveal it. However, the strength of SIMEX does not depend on knowing the functional form of $\hat{\theta}(\lambda)$. Our experience indicates that even for complicated problems $\hat{\theta}(\lambda)$ is well approximated by one or more of a handful of simple functional forms (see Cook and Stefanski (1992) and Section 4 of this paper), and that the approximations are generally more than adequate for applied work.

However, in most applications σ^2 is not known, and an estimate is used in its place. The effect of this can be analyzed in the simple example described above. When σ^2 is replaced by s_X^2, $\hat{\theta}_{\mathrm{SIMEX}} = \exp(\bar{X} - s_X^2/(2n))$. Routine calculations involving normal and chi-squared moment generating functions show that this estimator has expectation $a_n \exp(\mu)$ where

$$a_n = \exp(\sigma^2/(2n)) \left[1 + \sigma^2/\{n(n-1)\}\right]^{(1-n)/2}.$$

A little analysis shows that $a_n \geq 1$ for all n and that $n^3(a_n - 1) \to \sigma^4/4$ as $n \to \infty$. In other words the SIMEX bias, $E(\hat{\theta}_{\mathrm{SIMEX}}) - \theta$, is positive and of order n^{-3} for the case when σ^2 is estimated by s_X^2.

With regards to normality, SIMEX has no particular advantages or disadvantages relative to other nonlinear estimators. Many nonquadratic estimators that are unbiased under an assumption of normality, fail to remain unbiased under nonnormality in general. In the particular example of this section, it should be noted that normality of the data is important only to the extent that it guarantees the independence of $\bar{X}$ and s_X^2, and that expectations of certain functions of these statistics are the same as those obtained from the appropriate normal and chi-squared distributions.

4. EXTRAPOLANT CHARACTERISTICS AND EDA

4.1 Exploratory Data Analysis and Quadratic Extrapolants

An integral part of SIMEX is the fitting of extrapolant functions to the generated pseudo estimates, $\hat{\theta}(\lambda)$, $\lambda > 0$. The curves are necessarily smooth if B is large enough and in some cases have simple functional forms that are easily identified using exploratory data analysis techniques. However, in most problems the curves will not have simple functional forms although there will usually be some simple form that fits well. The problem is finding it. Since not all exploratory data analysts are created equal, it is difficult to present an objective description of a procedure that relies on an exploratory modelling step.

In the simulations and examples that follow we fit simple quadratic extrapolants in each case. Thus we avoid the problem of biasing the reported results by having prior knowledge of the appropriate, or best extrapolant function. In doing so, we provide a lower bound on the performance of SIMEX in the sense that statisticians are more competent curve fitters than our straw analyst who fits only quadratic functions. However, it will be evident that our straw analyst does quite well. The reason is that for many problems extrapolant curvature is slight and is well modeled by a quadratic.

4.2 The Range of λ

The simulation step can generate $\hat{\theta}(\lambda)$ for all $\lambda > 0$. There remains the problem of choosing a range of λ, say $[0, \Lambda]$, over which to fit the extrapolant model.

Having developed SIMEX as an exploratory method for analyzing and describing the effects of measurement error, we were led to consider initially $\Lambda = 2$. The plot of $\hat{\theta}(\lambda)$ for $0 \le \lambda \le 2$, yields immediate responses to queries about the effect of measurement error at levels two ($\lambda = 1$) and three ($\lambda = 2$) times as large as the nominal level σ^2 ($\lambda = 0$). Note that if the naive estimate is derived from measurements that are the means of replicate or triplicate measurements, then such queries translate directly into questions about the adequacy of a single measurement.

Subsequent simulation studies designed to investigate the choice of Λ have been generally inconclusive. This is good, since it suggests that SIMEX is relatively robust to choice of Λ. In addition, some useful conclusions are possible for the problem of estimating regression coefficients in large samples, under the assumption that B is large enough to ignore simulation variation.

Our observations are based on experience estimating regression parameters, suggesting that extrapolant functions are often monotone, concave or convex, and have decreasing curvature as λ increases. In such cases if a quadratic extrapolant is employed, then there is a tendency to undercorrect for the effects of measurement error. The reason is that the constant-curvature quadratic will diverge from the increasing-curvature true extrapolant as $\lambda \to -1$. The larger the value of Λ, the greater the undercorrection. However, a little undercorrection is desirable if minimum mean squared error is important. Thus there is, to a small extent, a bias/variance tradeoff associated with the choice of Λ when quadratic extrapolants are fit.

5. SIMEX COMPLEMENTS TO JACKKNIFE THEORY

5.1 The Jackknife in SIMEX Clothing

The connection between SIMEX and the jackknife suggests an alternative formulation of the latter. Let $\bar{\hat{\theta}}_{(n-j),(\cdot)}$ denote the average of the ${}_nC_j$ leave-j-out estimators $\hat{\theta}_{(n-j),(k)}$, $k = 1, \ldots, {}_nC_j$, based on samples of size $n - j$. For some fixed positive integer $j_{\max} \le n-1$ and for $\lambda_j = j/(n-j)$, $j = 0, 1, \ldots, j_{\max}$, $(0 \le \lambda_j \le \Lambda = j_{\max}/(n - j_{\max}))$, calculate $\hat{\theta}(\lambda_j) = \bar{\hat{\theta}}_{(n/(\lambda_j+1)),(\cdot)}$. Next plot $\hat{\theta}(\lambda_j)$ versus λ_j, examine this plot, and fit a model extrapolating the curve to $\lambda = -1$. The extrapolant at $\lambda = -1$ is the modified, reduced-bias jackknife estimate.

The mapping between λ_j and j (the number of points left out) is best understood in the context of estimating $\theta = f(\mu)$ via $f(\bar{X})$ with normally distributed, $N(\mu, \sigma^2)$, data. In this case $\bar{X}_{(n-j),(k)}$ and $\bar{X} + \sigma\sqrt{\lambda_j}\bar{Z}_b$ are identically distributed when $\lambda_j = j/(n-j)$. It follows that $f(\bar{X}_{(n-j),(k)})$ and $f(\bar{X} + \sigma\sqrt{\lambda_j}\bar{Z}_b)$ are identically distributed and thus possess the same moments, inparticular the same bias and variance. This is the key property relating SIMEX to the jackknife. Note that $j = -\infty$ corresponds to $\lambda_j = -1$; this correspondence is part of the explanation for our definition of jackknife estimator as a leave-$(-\infty)$-out estimator.

Others have studied the use of higher-order jackknife modifications (Gray *et al.*, 1972, 1975; Schucany *et al.*, 1971). Thus there is little new here except the proposal to examine the plot of $\hat{\theta}(\lambda)$ versus λ and use exploratory methods to suggest an appropriate extrapolant function.

The idea is intriguing. The usual two-point jackknife is a function of only two summary statistics, $\bar{\hat{\theta}}_{(n),(\cdot)}$, and $\bar{\hat{\theta}}_{(n-1),(\cdot)}$. Thus there is an inherent loss of information relative to the proposed modification which is a function of n summary statistics when $\Lambda = n - 1$.

Despite this advantage of the proposal, we doubt that it will supplant the usual jackknife. The reasons are twofold. First there is the inescapable bias-variance compromise. Reducing bias increases variability and thus there is always a point beyond which further bias reduction is counterproductive. For many real problems the usual jackknife seems to be near this optimal point. Second is the matter of knowing how to use the additional information in the modified procedure. The discussion in Section 4 makes clear that knowing $\hat{\theta}(\lambda)$ for large λ is useful only if the exact extrapolant function is known. When an approximate extrapolant is employed attention should be focused on λ in a neighborhood of 0. Applying the rule-of-thumb region, $0 \le \lambda \le 2$, sugested by our experience with SIMEX to the proposed modified jackknife bounds j via the inequality $2 \ge \lambda_j = j/(n-j)$, *i.e.*, $j \le 2n/3$.

Table 1 displays results of a Monte Carlo simulation designed to illustrate these ideas. The simulation employed the model of Section 3.2 with $n = 5$. Seven estimators were studied: the maximum likelihood estimate (MLE); the two-point linear jackknife estimator (LJ); three-, four-, and five-point quadratic modified jackknife estimators, *i.e.*, the modification suggested above using a quadratic extrapolant (QJ3–QJ5); the SIMEX estimator with s_X^2 in place of σ^2 (SIMEX1); and the SIMEX estimator with known σ^2 (SIMEX2). One-hundred thousand data sets were generated and analyzed.

The bias-corrected estimators do a good job of eliminating bias. Reductions ranged from 94% for LJ, to 100% for SIMEX2. All of the pairwise t-tests comparing MSEs are statistically significant, except the test comparing QJ3 and SIMEX1. However, except for SIMEX2, there is little practical difference between the estimators.

Of the three quadratic-extrapolant estimators the worst performing was QJ5 corresponding to $\lambda = 4$ which is much greater than the rule-of-thumb cutoff of 2. In contrast QJ3 and QJ4, with $\lambda = 2/3$ and $3/2$ respectively, performed better.

5.2 Jackknifing a Sample of Size One

In this section we use the term jackknife to denote both the specific techniques associated with the name, and the broader notion of a generally useful tool for reducing bias.

The connection between SIMEX and the jackknife makes clear that the essential role played by multiple observations in the latter is to provide a measure of variability. Jackknifing should be possible, in principle, with a sample of size one provided an external estimate of variability is available. One such manifestation of the jackknife under these conditions is SIMEX estimation.

The example in Section 3.2 makes this clear. Reexamination shows that, given either the raw data $\{X_j\}_1^n$, or $\bar{X}$ and either knowledge of σ^2 or an independent estimate of it, say $\hat{\sigma}^2$, it is possible to construct the SIMEX estimator. In the case that only $\bar{X}$ is given, the simulation step of SIMEX differs only in that the pseudo errors are generated from a N$(0,\ \hat{\sigma}^2/n)$ distribution and added to $\bar{X}$, where $\hat{\sigma}^2$ is either the known or estimated value of σ^2.

Thus we now consider bias reduction in the estimator $\hat{\theta}_{\mathrm{NAIVE}} = f(X)$ of $f(\mu)$, assuming that X is normally distributed with mean μ and known variance σ^2. We know how SIMEX is to be applied. Simulation is used to estimate $\hat{\theta}(\lambda) = E\{f(X + \sigma\sqrt{\lambda}\, Z_b) \mid X\}$ for $0 \le \lambda \le 2$, the plot of $\hat{\theta}(\lambda)$ vs λ is examined, and a model is fit to the curve. Extrapolation of the model to $\lambda = -1$ results in the reduced-bias SIMEX estimator.

Note that for the simple models studied in this section, $\hat{\theta}(\lambda)$ can be calculated by numerical integration more easily and efficiently than by simulation. However, the simple models are studied primarily to gain insight into the performance of SIMEX in more complicated settings where numerical integration is not feasible and thus we maintain the simulation component.

We now show that if the true extrapolant is employed, then for a general class of functions f, the SIMEX estimator is unbiased. The results that follow hold for smooth functions f, although the smoothness requirements are much greater than the common regularity conditions (one or two continuous derivatives) usually encountered in mathematical statistics. A minimum requirement is that f be analytic, *i.e.*, has a convergent power series expansion, and furthermore that expectation and summation can be interchanged in the power series expansion of f. The necessary smoothness conditions have been studied elsewhere (Stefanski, 1989) and will not be examined here. In the following $i = \sqrt{-1}$.

LEMMA 1. If Z_1 and Z_2 are independent and identically distributed standard normal variates, then $E\{(Z_1 + iZ_2)^n\} = 0,\ n = 1, 2, \ldots$.

PROOF. The result can be verified by binomial expansion and direct calculation. Alternatively, consider that $Z_1 + \sqrt{\lambda}\, Z_2$ has a N(0, $1+\lambda$) distribution and thus $E\{(Z_1 + \sqrt{\lambda}\, Z_2)^n\}$ is proportional to $(1+\lambda)^{n/2}$. Let $\lambda \to -1$ to establish the result. This proof actually uses the result of Lemma 2, although in an easily justified case. ••

LEMMA 2. If f is sufficiently smooth and Z_1 and Z_2 are independent and identically distributed standard normal variates, then

$$\lim_{\lambda \to -1} E\{f(\mu + \sigma Z_1 + \sqrt{\lambda}\, \sigma Z_2)\} = E\{f(\mu + \sigma Z_1 + i\sigma Z_2)\} = f(\mu). \tag{5.1}$$

PROOF. Under the assumptions on f, and upon appeal to Lemma 1,

$$\begin{aligned}
\lim_{\lambda \to -1} E\{f(\mu + \sigma Z_1 + \sqrt{\lambda}\, \sigma Z_2)\} &= \lim_{\lambda \to -1} E\{f(\mu) + \sum_{n=1}^{\infty} \frac{f^{(n)}(\mu)\sigma^n}{n!}(Z_1 + \sqrt{\lambda}\, Z_2)^n\} \\
&= E\{f(\mu) + \sum_{n=1}^{\infty} \frac{f^{(n)}(\mu)\sigma^n}{n!}(Z_1 + iZ_2)^n\} \\
&= f(\mu) + \sum_{n=1}^{\infty} \frac{f^{(n)}(\mu)\sigma^n}{n!} E\{(Z_1 + iZ_2)^n\} = f(\mu), \qquad (5.2)
\end{aligned}$$

completing the proof. ••

Henceforth, we define a sufficiently smooth function f to be one for which the result of Lemma 2 holds.

The exact SIMEX extrapolant is just $\hat{\theta}(\lambda) = E\{f(X + \sqrt{\lambda}\, \sigma Z_b) \mid X\}$ and the corresponding exact SIMEX estimator is $\hat{\theta}_{\text{SIMEX}} = E\{f(X + i\sigma Z_b) \mid X\}$. Now with $Z = (X - \mu)/\sigma$,

$$E\{\hat{\theta}_{\text{SIMEX}}\} = E\{f(X + i\sigma Z_b)\} = E\{f(\mu + \sigma Z + i\sigma Z_b)\} = f(\mu),$$

the latter equality following from Lemma 2. That is, the exact SIMEX estimator is unbiased.

Lemma 2 and the unbiasedness of the exact SIMEX estimator have an interesting (and amusing) intuitive explanation. Estimator bias generally results from the fact that for a nonlinear function, g, and random variate X, $E\{g(X)\} \neq g(E\{X\})$ in general. In fact the latter is true in general only if X is degenerate, *i.e.*, Var(X)=0, in which case $E\{X^n\} = (E\{X\})^n$ for all n.

In the following we define $\text{Var}\{W\} = E\{W^2\} - (E\{W\})^2$ for complex-valued random variables W. Also if W_1 and W_2 are two complex-valued random variables we define their covariance as $\text{Cov}\{W_1, W_2\} = E\{W_1 W_2\} - E\{W_1\}E\{W_2\}$. It will become evident that this is a useful departure from convention. The reader is warned that these definitions give rise to some unusual looking results.

It follows from Lemma 2 that the complex variate $W = \mu + \sigma Z_1 + i\sigma Z_2$, where Z_1 and Z_2 are independent standard normal variates, has mean $E\{W\} = \mu$, and variance $\text{Var}\{W\} = 0$. In fact Lemma 2 shows that $E\{W^n\} = \mu^n = (E\{W\})^n$ for all $n \geq 1$. Thus W behaves very much like a degenerate N(μ, 0) random variate with respect to expectation, in the sense that for any smooth g, $E\{g(W)\} = g(E\{W\})$. Since adding $i\sigma Z_2$ to σZ_1 annihilates σZ_1 in an expected value sense, we refer to $i\sigma Z_2$ as the anti-measurement error of σZ_1. We now have the amusing, yet fairly descriptive paraphrasing of Lemma 2, "measurement error-induced bias is annihilated by the addition of anti-measurement error."

A simple example illustrates these ideas. Consider the problem of estimating $f(\mu) = \mu^2$. The estimator $f(X) = X^2$ is biased because of the measurement error $X - \mu = \sigma Z$. Let $X_* = X + i\sigma Z_b$. A candidate estimator is $X_*^2 = X^2 - \sigma^2 Z_b^2 + 2iX\sigma Z_b$. This has expectation $E\{X_*^2\} = \mu^2$ and is thus unbiased. We have annihilated the measurement error in an expected value sense. However, the new estimator has certain deficiencies, namely it is a complex variate and it depends on the arbitrary, generated pseudo error, Z_b. These deficiencies are remedied by taking the real part of the estimator, and then taking expectations conditional on X; or simply by taking expectations conditional on X, since the imaginary part vanishes upon conditional expectation. We are left with the unbiased estimator $\hat{\theta}_{\text{SIMEX}} = X^2 - \sigma^2$.

Although the previous paragraph dealt with $f(\mu) = \mu^2$ the main points apply to any smooth function f: a) $E\{f(X + i\sigma Z_b)\} = f(\mu)$; b) the imaginary part of $f(X + i\sigma Z_b)$ has zero conditional (on X) mean; and c) $E\{f(X + i\sigma Z_b) \mid X\}$ is an unbiased estimator of $f(\mu)$.

This section makes clear why SIMEX calls for plotting $\hat{\theta}(\lambda)$ as a function of λ, and not on some other scale, and the significance of extrapolating to $\lambda = -1$. It also provides the final piece of evidence justifying our notation for the jackknife estimator as a leave-$(-\infty)$-out estimator.

5.3 Variance Estimation

The jackkknife is used at least as often for variance estimation as it is for bias reduction (Tukey, 1958; Efron, 1982). We now examine SIMEX in regards to variance estimation.

The basic building blocks of a jackknife variance estimate are the n differences

$$\Delta_k = \hat{\theta}_{(n-1),(k)} - \bar{\hat{\theta}}_{(n-1),(\cdot)}, \qquad k = 1, \ldots, n.$$

The jackknife estimate of variance is $n^{-1}(n-1)\sum_{k=1}^{n} \Delta_k^2$.

In the SIMEX version of the jackknife with sample size = 1, corresponding quantities are

$$\Delta(\lambda) = \hat{\theta}_b(\lambda) - \hat{\theta}(\lambda), \qquad \lambda > 0.$$

We use the problem of estimating $\exp(\mu)$ to illustrate the manner in which these differences are employed in variance estimation. The SIMEX estimator is $\hat{\theta}_{\text{SIMEX}} = \exp(X - \sigma^2/2)$. It has expectation $\exp(\mu)$ and its variance is easily calculated and shown to be $\{\exp(\sigma^2) - 1\}\exp(2\mu)$.

For this problem,

$$\Delta(\lambda) = \hat{\theta}_b(\lambda) - \hat{\theta}(\lambda) = \exp(X)\{\exp(\sigma\sqrt{\lambda}\, Z_b) - \exp(\lambda\sigma^2/2)\}.$$

The variance of this difference is easily calculated and found to be,

$$\text{Var}\{\Delta(\lambda)\} = \exp(2\mu + 2\sigma^2)\{\exp(2\lambda\sigma^2) - \exp(\lambda\sigma^2)\}.$$

Note that since $\hat{\theta}(\lambda) = E\{\hat{\theta}_b(\lambda) \mid X\}$,

$$\text{Var}\{\Delta(\lambda)\} = E\{\text{Var}\{\Delta(\lambda) \mid X\}\} = E\{\text{Var}\{\Delta(\lambda) \mid X,\ \hat{\theta}(\lambda)\}\}. \tag{5.3}$$

The variance of $\hat{\theta}_{\text{SIMEX}}$, which we have already calculated directly, can also be obtained by taking the limit as $\lambda \to -1$ of $\text{Var}\{\Delta(\lambda)\}$, *viz.*,

$$\text{Var}\{\hat{\theta}_{\text{SIMEX}}\} = -\lim_{\lambda \to -1} \text{Var}\{\Delta(\lambda)\}. \tag{5.4}$$

It is easily verified that taking the indicated limit results in the correct variance. The formula is also easily checked for the case $f(\mu) = \mu^2$. Note that $\Delta(0) = 0$ and thus $\text{Var}\{\Delta(0)\} = 0$. Since for $\lambda > 0$, $\text{Var}\{\Delta(\lambda)\} > 0$, it is to be expected that as this curve is extrapolated to $\lambda < 0$, $\text{Var}\{\Delta(\lambda)\}$ takes on negative values. This seemingly preposterous result is due to our definition of variance for complex-valued random variates, *e.g.*, $\text{Var}\{i\sigma Z_b\} = -\sigma^2$. Since $\hat{\theta}_{\text{SIMEX}}$ is itself a limit as $\lambda \to -1$, the previous formula is naturally expressed as

$$\text{Var}\left\{\lim_{\lambda \to -1} \hat{\theta}(\lambda)\right\} = -\lim_{\lambda \to -1} \text{Var}\{\Delta(\lambda)\}. \tag{5.5}$$

Before demonstrating the correctness of the formula in general we describe its relevance in practice.

In problems that defy analytic treatment, the variance of the SIMEX estimator can be estimated by: 1) calculating $s^2_\Delta(\lambda)$, the sample variance of the $\hat{\theta}_b(\lambda)$, $b = 1, \ldots, B$, for each λ, thereby obtaining an unbiased estimator of $\text{Var}\{\Delta(\lambda) \mid X\}$, which by (5.3) is also unbiased for $\text{Var}\{\Delta(\lambda)\}$; 2) plotting $s^2_\Delta(\lambda)$ versus λ; and 3) modelling the curve and extrapolating back to $\lambda = -1$. The result is an estimator, albeit an approximate one when an approximate extrapolant is used, of $-\text{Var}\{\hat{\theta}_{\text{SIMEX}}\}$.

LEMMA 3. If f is sufficiently smooth, then (5.5) holds.

PROOF. The derivation of the variance formula depends on a simple but subtle point that is best attended to initially. Lemma 2 shows that when f is smooth, $\hat{\theta}_b(\lambda) = f(X + \sqrt{\lambda}\,\sigma Z_b)$ has expectation $E\{\hat{\theta}_b(\lambda)\} \to f(\mu)$ as $\lambda \to -1$. When f is smooth f^2 is also smooth, and thus Lemma 2 also shows that $E\{\hat{\theta}^2_b(\lambda)\} \to f^2(\mu)$ as $\lambda \to -1$. It follows that

$$\lim_{\lambda \to -1} \text{Var}\{\hat{\theta}_b(\lambda)\} = 0. \tag{5.6}$$

Now since $\hat{\theta}(\lambda) = E\{\hat{\theta}_b(\lambda) \mid X\}$, it follows that $\text{Cov}\{\hat{\theta}_b(\lambda),\ \hat{\theta}(\lambda)\} = \text{Var}\{\hat{\theta}(\lambda)\}$. Thus, via the usual variance decomposition,

$$\begin{aligned}\text{Var}\{\hat{\theta}_b(\lambda) - \hat{\theta}(\lambda)\} &= \text{Var}\{\hat{\theta}_b(\lambda)\} + \text{Var}\{\hat{\theta}(\lambda)\} - 2\text{Cov}\{\hat{\theta}_b(\lambda),\ \hat{\theta}(\lambda)\} \\ &= \text{Var}\{\hat{\theta}_b(\lambda)\} - \text{Var}\{\hat{\theta}(\lambda)\}.\end{aligned}$$

In light of (5.6) it is seen that

$$\lim_{\lambda \to -1} \text{Var}\{\hat{\theta}_b(\lambda) - \hat{\theta}(\lambda)\} = -\lim_{\lambda \to -1} \text{Var}\{\hat{\theta}(\lambda)\} = -\text{Var}\{\hat{\theta}(-1)\} = -\text{Var}\{\hat{\theta}_{\text{SIMEX}}\},$$

completing the proof. ●●

SIMEX is now a complete package for the problem of estimating $f(\mu)$ from $X \sim \text{N}(\mu,\ \sigma^2)$ with σ^2 known or independently estimated, say be $\hat{\sigma}^2$. Pseudo errors are generated to obtain $\hat{\theta}_b(\lambda) = f(X + \hat{\sigma}\sqrt{\lambda}\, Z_b)$, $b = 1, \ldots, B$. The sample mean, $\hat{\theta}(\lambda)$, and variance, $s^2_\Delta(\lambda)$, of $\{\hat{\theta}_b(\lambda)\}_1^B$ are then calculated for several λ. These are plotted as functions of λ, modelled and extrapolated back to $\lambda = -1$. The extrapolated mean function yields the estimator $\hat{\theta}_{\text{SIMEX}}$; the extrapolated variance function, multiplied by -1, yields an estimator of $\text{Var}\{\hat{\theta}_{\text{SIMEX}}\}$. The method is exact, in the sense of producing unbiased estimators, when σ^2 is known, *i.e.*, $\hat{\sigma}^2 = \sigma^2$, and the exact extrapolant functions are employed. Otherwise it is approximate.

For the problem of estimating $f(\mu) = \exp(\mu)$, it is readily verified that as $B \to \infty$, $s^2_\Delta(\lambda) \to v(X, \lambda)$, where $v(X, \lambda) = \mathrm{Var}\{\hat{\theta}_b(\lambda) \mid X\} = \{\exp(2\sigma^2\lambda) - \exp(\sigma^2\lambda)\}\exp(2X)$ is the exact extrapolant function for variance estimation. Thus if the exact extrapolant is employed, the SIMEX estimator of $\mathrm{Var}\{\hat{\theta}_{\mathrm{SIMEX}}\}$ is $-v(X, -1)$. Under the assumed normality, $-v(X, -1)$ is the uniform minimum variance unbiased estimator of $\mathrm{Var}\{\hat{\theta}_{\mathrm{SIMEX}}\}$. It is tempting to debunk this optimality in light of the small likelihood of identifying the exact variance extrapolant function in this case without prior knowledge its the functional form. However, the important point is that the approximate SIMEX variance estimator, approximates an estimator that makes efficient use of the data.

Cook and Stefanski (1992) suggested the use of pseudo errors contstrained to have certain population moments, as a means of reducing variability in the Monte Carlo estimate of $\hat{\theta}(\lambda)$ and decreasing computation time. However, they did not consider SIMEX variance estimation. In order for the variance estimator suggested herein to be valid in general, the pseudo errors must be independent and identically distributed.

5.3.1 Variance Estimation Simulation

We now illustrate this procedure with an example and study its performance via simulation. We take $f(\mu) = \mu/(1 + \exp(-\mu))$. This function is not analytic and thus Lemma 2 does not apply. We chose this function for two reasons. First, to demonstrate the utility of the method even when the mathematical theory does not apply and an approximate extrapolant, a quadratic, is necessarily employed.

Second, is the relevance of this function to understanding measurement error in logistic regression. The normal equations for simple logistic regression are $\sum_{j=1}^{n}\{Y_j - G(\alpha + \beta U_j)\}(1,\ U_j)^T = (0,\ 0)^T$ where G is the logistic distribution function. If U_j is replaced by a measurement X_j, then the normal equations lose the essential feature of Fisher consistency since

$$E\left\{\{Y_j - G(\alpha + \beta X_j)\}(1,\ X_j)^T\right\} = G(\alpha + \beta U_j)(1,\ U_j)^T - E\left\{G(\alpha + \beta X_j)(1,\ X_j)^T\right\} \neq (0,\ 0)^T.$$

Nonlinearity of the functions $G(t)$ and $tG(t)$ is responsible for the lack of Fisher consistency. The latter is just the function $f(t) = t/(1 + \exp(-t))$ chosen for our simulation.

For the simulation study $\mu = 1$ was fixed. The measurement error variance σ^2 had levels 0.25, 0.50, 0.75 and 1.00. Two independent $\mathrm{N}(\mu,\ 2\sigma^2)$ measurements of μ were generated as well as a second independent sample of size $d + 1$ of $\mathrm{N}(\mu,\ \sigma^2)$ measurements. The first sample of size 2 was averaged to form the single measurement $X \sim \mathrm{N}(\mu,\ \sigma^2)$. The second sample was used to construct an estimate of σ^2 based on d degrees of freedom. This was done for $d = 1,\ 16,\ 256$ and ∞, *i.e.*, σ^2 known.

This design allowed us to control both the measurement error variance and the precision of the independent estimate, $\hat{\sigma}^2$, of σ^2 required by SIMEX, while also allowing for the calculation of the standard, linear-extrapolated jackknife estimator from the initial sample of two measurements.

The study compared three estimator/variance estimator pairs: the maximum likelihood estimator $\hat{\theta}_{\text{NAIVE}} = f(X)$, with Δ-method variance estimator $\{f'(X)\}^2\hat{\sigma}^2$; the SIMEX estimator and corresponding variance estimator employing quadratic extrapolants for each; and the jackknife estimator and corresponding variance estimator based on the sample of size 2. The SIMEX estimator was calculated with $B = 1000$ and 4000. We report on only the results for $B = 4000$. The smaller value of B was studied as a check on the adequacy of $B = 4000$. The results for $B = 1000$ and $B = 4000$ were similar. For each combination of σ^2 and d, 25,000 replications were performed.

The intent of the study is to show that even with the use of simple quadratic extrapolants, and for a case where the exact theory does not hold, and when σ^2 is replaced by $\hat{\sigma}^2$, that SIMEX reduces bias as well as the jackknife, and that the proposed variance estimation method is comparable to the jackknife variance estimator *as an estimator of the variance of the reduced-bias jackknife estimator.* Relevant results of the study are displayed in in Figures 3a–3d and 4a–4d.

With 25,000 replications, visually detectably differences and statistically significant differences are roughly the same. In order to preserve clarity in the figures we elected not to display standard errors. However, some feeling for the simulation variability can be inferred from the plots upon realizing that the jackknife procedure made no use of the data used to generate the independent estimate of σ^2. Thus the four dashed lines in Figures 3a–3d and in Figures 4a–4d are independent replicates.

Figures 3a–3d contain plots of relative bias as functions of σ^2 for each d. The SIMEX estimator displayed uniformly smallest bias across levels of σ^2 and d, with the jackknife a close second.

Figures 4a–4d display plots of the (mean of the variance estimator)/(true estimator variance) as functions of σ^2 for each d. In all cases the SIMEX procedure came closest to the ideal value of unity, *i.e.*, unbiased variance estimation. The true variances of the SIMEX and jackknife estimators were virtually identical over all factor level combinations in the experiment. Thus the differences in Figures 4a–4d are due almost exclusively to differences in the means of the estimated variances. Both procedures underestimate variability, the jackknife more than SIMEX.

These results are not in conflict with the well-known conservativeness of the jackknife variance estimator (Efron and Stein, 1981), for that result pertains to the use of the jackknife variance estimator as an estimator of the variance of the initial estimator, the maximum likelihood estimator in this case, and not as an estimator of the variance of the reduced-bias jackknife estimator.

Figures 5a and 5b display an example of the quadratic mean and variance extrapolations used in the SIMEX procedure. Figures 5a and 5b were constructed from a data set generated as in the simulation with $\sigma^2 = 0.5$ and $d = 256$, and employed B=1000 replications in the simulation step. The unusually poor fit of the extrapolant in Figure 5a suggested that B was not large enough for this particular sample, *i.e.*, an unusually variable sequence of pseudo errors had been generated. We reanalyzed the data set taking $B = 16,000$ resulting in the quadratic mean and variance extrapolants plotted in Figures 5c and 5d. Comparison between Figures 5a and 5c suggests that even though our simulations failed to reveal noteworthy differences between $B = 1000$ and $B = 4000$, that differences might be detected at even larger values of B.

5.3.2. Jackknife Variance Estimation Revisited

The SIMEX description of the jackknife presented in Section 5.1 and Lemma 3 suggest an alternative method of deriving a jackknife variance estimator. Suppose that $\Lambda = 1/(n-1)$ so that the usual linear jackknife is obtained. Lemma 3 then suggests fitting a line to the two points (λ_j, V_j), $j = 1,2$ where $\lambda_1 = 0$ and $V_1 = 0$, and $\lambda_2 = 1/(n-1)$ and $V_2 = n^{-1}\sum_1^n \Delta_k^2$, the population variance of the n differences $\Delta_k = \hat{\theta}_{(n-1),(k)} - \bar{\hat{\theta}}_{(n-1),(\cdot)}$; then extrapolating the line back to $\lambda = -1$, resulting in $-V_2/\lambda_2$. Upon multiplication by -1 we arrive at the new variance estimator, $n^{-1}(n-1)\sum_{k=1}^n \Delta_k^2$, that is, the usual jackknife variance estimator. In other words, the ususal jackknife variance estimator is an extrapolation of the form (5.5), but with an approximate (linear) extrapolant.

It may appear that this result was fudged by use of the divisor n in the definition of V_2. However, for SIMEX variance estimation, $s_\Delta^2(\lambda)$ is an unbiased estimator of the variance of $\hat{\theta}_b(\lambda)$ about its conditional mean $\hat{\theta}(\lambda)$. Likewise V_2 should be an unbiased estimator of the variance of $\hat{\theta}_{(n-1),(k)}$ about $\bar{\hat{\theta}}_{(n-1),(\cdot)}$. In the special case that $\hat{\theta}$ is the sample mean, then either exactly for finite samples and normally distributed data, or approximately for large samples upon appeal to the Central Limit Theorem, the conditional distribution of $\hat{\theta}_{(n-1),(k)}$ given $\bar{\hat{\theta}}_{(n-1),(\cdot)}$ is normal with mean $\bar{\hat{\theta}}_{(n-1),(\cdot)}$. It follows that the divisor n in V_2 yields an unbiased estimator of the appropriate conditional variance. The same argument applys asymptotically for non-sample mean statistics after linearization.

6. JACKKNIFE COMPLEMENTS TO SIMEX THEORY

In this section we extend the jackknife-like variance estimation procedure for SIMEX estimators studied in Section 5.3, to SIMEX estimators for the class of measurement error models studied by

Cook and Stefanski (1992).

In Section 3.1 we introduced the function T to denote the estimator under study. We now introduce a second function, $T_{\rm Var}$ to denote an associated variance estimator. That is, with $\hat{\theta}_{\rm TRUE} = T(\{Y_j, V_j, U_j\}_1^n)$, $T_{\rm Var}(\{Y_j, V_j, U_j\}_1^n)$ is an estimator of $\text{Var}\{\hat{\theta}_{\rm TRUE}\}$. We allow T to be p-dimensional, in which case $T_{\rm Var}$ is $p \times p$-matrix valued, and variance refers to the variance-covariance matrix. We use τ^2 to denote the parameter $\text{Var}\{\hat{\theta}_{\rm TRUE}\}$, $\hat{\tau}^2_{\rm TRUE}$ to denote the estimator $T_{\rm Var}(\{Y_j, V_j, U_j\}_1^n)$, $\hat{\tau}^2_{\rm NAIVE}$ to denote the naive estimator $T_{\rm Var}(\{Y_j, V_j, X_j\}_1^n)$, and so on.

6.1 Multivariate Anti-Measurement Error

We start this section with a generalization of Lemma 2 to multivariate normal random variates.

LEMMA 4. If f is a sufficiently smooth function from $\mathbf{R}^p$ to $\mathbf{R}^q$, $\mu \in \mathbf{R}^p$, and Z_1 and Z_2 are independent and identically distributed p-dimensional N(0, Ω) random vectors, then

$$\lim_{\lambda \to -1} E\{f(\mu + Z_1 + \sqrt{\lambda}\, Z_2)\} = E\{f(\mu + Z_1 + iZ_2)\} = f(\mu) \tag{6.1}$$

PROOF. The assumed smoothness allows us to represent the q components of f as multivariate power series and to exchange expectation and series summation. First assume that Ω is the identity matrix. In this case the key step in the proof entails showing that

$$E\left\{\prod_{j=1}^{p}(Z_{1,j} + iZ_{2,j})^{r_j}\right\} = 0$$

when at least one of the nonnegative interger powers $r_j > 0$, where $Z_{1,j}$ and $Z_{2,j}$, $j = 1, \ldots, p$ are the components of Z_1 and Z_2 respectively. The truth of this assertion follows from independence and appeal to Lemma 1. It then follows that upon taking expectations in the series expansion of $f(\mu + Z_1 + iZ_2)$ around μ, that all terms vanish except the first, that is $f(\mu)$.

For the case with general Ω define g via $g(t) = f(\mu + \Omega^{1/2}t)$ where $\Omega^{1/2}$ is the symmetric square root of Ω. By the independence case just established, $E\{g(\Omega^{-1/2}Z_1 + i\Omega^{-1/2}Z_2)\} = g(0)$. Invoking the relationship between f and g establishes the result. ●●

6.2 Exact SIMEX, Finite-Sample Theory

In this section we assume that $\hat{\theta}_{\rm TRUE} = T(\{Y_j, V_j, U_j\}_1^n)$ is an unbiased estimator of θ and that $T_{\rm Var}(\{Y_j, V_j, U_j\}_1^n)$ is an unbiased estimator of $\text{Var}\{\hat{\theta}_{\rm TRUE}\}$. Furthermore we assume that both T and $T_{\rm Var}$ are smooth functions of the arguments $U_1, \ldots, U_n$.

The fundamental identity $\hat{\theta}(\lambda) = E\{\hat{\theta}_b(\lambda) \mid \{Y_j, V_j, X_j\}_1^n\}$ and Lemma 4 are used to establish that

$$\lim_{\lambda \to -1} E\{\hat{\theta}_b(\lambda)\} = E\{\hat{\theta}_b(-1)\} = \theta, \qquad \lim_{\lambda \to -1} E\{\hat{\theta}(\lambda)\} = E\{\hat{\theta}(-1)\} = \theta,$$

and

$$\text{Var}\{\hat{\theta}_b(\lambda) - \hat{\theta}(\lambda)\} = \text{Var}\{\hat{\theta}_b(\lambda)\} - \text{Var}\{\hat{\theta}(\lambda)\}.$$

Further appeal to smoothness, Lemma 4, and upon invoking our definition of variance in the case of complex-valued p-dimensional variates, show that

$$\begin{aligned}\text{Var}\{\hat{\theta}_{\text{SIMEX}}\} = \text{Var}\{\hat{\theta}(-1)\} &= \lim_{\lambda \to -1} \text{Var}\{\hat{\theta}(\lambda)\} \\ &= \lim_{\lambda \to -1} \text{Var}\{\hat{\theta}_b(\lambda)\} - \lim_{\lambda \to -1} \text{Var}\{\hat{\theta}_b(\lambda) - \hat{\theta}(\lambda)\} \\ &= \text{Var}\{\hat{\theta}_b(-1)\} - \lim_{\lambda \to -1} \text{Var}\{\hat{\theta}_b(\lambda) - \hat{\theta}(\lambda)\}. \qquad (6.2)\end{aligned}$$

But the fact that TT^T is smooth whenever T is smooth, and further appeal to Lemma 4 show that

$$\begin{aligned}\text{Var}\{\hat{\theta}_b(-1)\} &= E\{\hat{\theta}_b(-1)\hat{\theta}_b^T(-1)\} - E\{\hat{\theta}_b(-1)\}(E\{\hat{\theta}_b(-1)\})^T \\ &= E\{E\{\hat{\theta}_b(-1)\hat{\theta}_b^T(-1) \mid \{Y_j, V_j, X_j\}_1^n\}\} - \theta\theta^T \\ &= E\{\hat{\theta}_{\text{TRUE}}\hat{\theta}_{\text{TRUE}}^T\} - \theta\theta^T \\ &= \text{Var}\{\hat{\theta}_{\text{TRUE}}\} \\ &= \tau^2.\end{aligned}$$

SIMEX estimation can be used to estimate (the components of) τ^2. That is, $\hat{\tau}_b^2(\lambda) = T_{\text{Var}}(\{Y_j, V_j, X_{b,j}(\lambda)\}_1^n)$ is calculated for $b = 1, \ldots, B$, and upon averaging and letting $B \to \infty$, results in $\hat{\tau}^2(\lambda)$, and so on. Furthermore, the sample variance matrix of $\{\hat{\theta}_b(\lambda)\}_{b=1}^B$, call it $s_\Delta^2(\lambda)$, is an unbiased estimator of $\text{Var}\{\hat{\theta}_b(\lambda) - \hat{\theta}(\lambda) \mid \{Y_j, V_j, X_j\}_1^n\}$ for all $B > 1$ and converges in probability to its expectation as $B \to \infty$. It follows that $E\{s_\Delta^2(\lambda)\} = \text{Var}\{\hat{\theta}_b(\lambda) - \hat{\theta}(\lambda)\}$.

The plots of (the components of) $s_\Delta^2(\lambda)$ versus $\lambda > 0$, extrapolated back to $\lambda = -1$ provide an estimator of

$$\lim_{\lambda \to -1} \text{Var}\{\hat{\theta}_b(\lambda) - \hat{\theta}(\lambda)\}.$$

The estimator is unbiased if the exact extrapolant is used. In light of (6.2) the difference, $\hat{\tau}_{\text{SIMEX}}^2 - s_\Delta^2(-1)$, is an unbiased estimator of $\text{Var}\{\hat{\theta}_{\text{SIMEX}}\}$.

SIMEX estimation is now a complete package for the special case of this section. The simulation step results in $\hat{\theta}(\lambda)$, $\hat{\tau}^2(\lambda)$ and $s_\Delta^2(\lambda)$. The extrapolation of $\hat{\theta}(\lambda)$ to $\lambda = -1$, $\hat{\theta}_{\text{SIMEX}}$, provides an unbiased estimator of θ, and extrapolation of (the components of) the difference, $\hat{\tau}^2(\lambda) - s^2_\Delta(\lambda)$ to $\lambda = -1$ provides an unbiased estimator of $\mathrm{Var}\{\hat{\theta}_{\mathrm{SIMEX}}\}$.

6.2.1. Components-of-Variance Estimation

The components-of-variance problem provides a simple yet informative illustration of these ideas. Suppose that $U_1, \ldots, U_n$ are independent and identically distributed $\mathrm{N}(\mu_U, \theta)$. Thus $\hat{\theta}_{\mathrm{TRUE}} = s^2_U$ and $\hat{\theta}_{\mathrm{NAIVE}} = s^2_X$. Cook and Stefanski (1992) show that for this problem $\hat{\theta}(\lambda) = s^2_X + \lambda\sigma^2$ and thus $\hat{\theta}_{\mathrm{SIMEX}} = s^2_X - \sigma^2$. The normality of $U_1, \ldots, U_n$ is significant only to the extent that it allows us to easily identify an unbiased variance estimator, $\hat{\tau}^2_{\mathrm{TRUE}}$, of $\hat{\theta}_{\mathrm{TRUE}}$.

It is readily verified that $\tau^2 = 2\theta^2/(n-1)$, $\hat{\tau}^2_{\mathrm{TRUE}} = 2s^4_U/(n+1)$,

$$\hat{\tau}^2(\lambda) = \frac{2}{n+1}\left\{(s^2_X + \lambda\sigma^2)^2 + \frac{2\lambda^2\sigma^4 + 4\lambda\sigma^2 s^2_X}{n-1}\right\},$$

and

$$s^2_\Delta(\lambda) = \frac{2\lambda^2\sigma^4 + 4\sigma^2\lambda s^2_X}{n-1}.$$

The difference, $\hat{\tau}^2(\lambda) - s^2_\Delta(\lambda) = 2s^4_X/(n+1)$, is constant in λ and thus extrapolation to $\lambda = -1$ is trivial. It is equally trivial to show that the constant difference, $2s^4_X/(n+1)$, is an unbiased estimator of $\mathrm{Var}\{\hat{\theta}_{\mathrm{SIMEX}}\}$.

6.3 Exact SIMEX, Large-Sample Theory

In many problems, even in the absence of measurement error, the functions T and T_{Var} do not yield unbiased estimators, but do so only asymptotically, *i.e.*, they yield consistent estimators as the sample size, $n \to \infty$. We now argue heuristically that the SIMEX variance estimator works asymptotically in general for such problems.

Our intent is explain how SIMEX works in large samples and not to provide rigorous conditions under which the asymptotic results are valid. These are best handled on a case-by-case basis. We begin the heuristics by replacing the finite sample estimators with their large sample linearizations, that are assumed to exist.

Thus we let

$$\hat{\theta}_{\mathrm{TRUE}} = \theta + \frac{1}{n}\sum_{j=1}^{n} IC(Y_j, V_j, U_j, \theta),$$

where IC is the influence curve associated with T. Having so defined $\hat{\theta}_{\mathrm{TRUE}}$, it is natural to define

$$\tau^2 = \frac{1}{n^2}\sum_{j=1}^{n} E\{IC(Y_j, V_j, U_j, \theta) IC^T(Y_j, V_j, U_j, \theta)\},$$

$$\hat{\tau}^2_{\mathrm{TRUE}} = \frac{1}{n^2}\sum_{j=1}^{n} IC(Y_j, V_j, U_j, \theta) IC^T(Y_j, V_j, U_j, \theta),$$

Related quantities required in SIMEX estimation are

$$\hat{\theta}_b(\lambda) = \theta(\lambda) + \frac{1}{n}\sum_{j=1}^{n} IC(Y_j, V_j, X_{b,j}(\lambda), \theta(\lambda)),$$

$$\hat{\theta}(\lambda) = \theta(\lambda) + \frac{1}{n}\sum_{j=1}^{n} E\{IC(Y_j, V_j, X_{b,j}(\lambda), \theta(\lambda)) \mid Y_j, V_j, X_j\},$$

$$\hat{\tau}_b^2(\lambda) = \frac{1}{n^2}\sum_{j=1}^{n} IC(Y_j, V_j, X_{b,j}(\lambda), \theta(\lambda)) IC^T(Y_j, V_j, X_{b,j}(\lambda), \theta(\lambda)),$$

and

$$\hat{\tau}^2(\lambda) = \frac{1}{n^2}\sum_{j=1}^{n} E\{IC(Y_j, V_j, X_{b,j}(\lambda), \theta(\lambda)) IC^T(Y_j, V_j, X_{b,j}(\lambda), \theta(\lambda)) \mid Y_j, V_j, X_j\}.$$

It is now a simple exercise to verify that all of the results from the previous section hold under the assumption that IC is a smooth function of U_j. Thus for SIMEX and SIMEX variation estimation to work in large samples, smoothness of the functions T and T_{Var} is not important *per se*, but rather it is smoothness of the influence curve that is relevant.

These large-sample heuristics are useful for they indicate the key components involved analyzing SIMEX in large samples, namely linearization and smoothness of the linearized form, but they hide the fact that a rigorous demonstration is quite involved. We illustrate this with a detailed analysis of simple linear regression in the presence of measurement error.

6.3.1. SIMEX Variance Estimation in Simple Linear Regression

The regression model under investigation is $Y_j = \alpha + \beta U_j + \epsilon_j$, $j = 1, \ldots, n$, where it is assumed that the equation errors $\{\epsilon_j\}_1^n$ are identically distributed with $E\{\epsilon_1\} = 0$ and $\text{Var}\{\epsilon_1\} = \sigma_\epsilon^2$, mutually independent, and are independent of the measurement errors. Furthermore we assume the functional version of this model, *i.e.*, that $U_1, \ldots, U_n$ are nonrandom constants. We add the regularity condition that the sample variance of $\{U_j\}_1^n$ converges to $\sigma_U^2 > 0$, as $n \to \infty$. This is stronger than is necessary, but simplifies the presentation.

We will not address asymptotic normality *per se*, since this follows from the easily established asymptotic equivalence of the SIMEX estimator and the much-studied method-of-moments estimator (Fuller, 1987, pp 15-17); see also Lombard *et al.* (1993). Our focus is on variance estimation. Define $S_{YU} = n^{-1}\sum(Y_j - \bar{Y})(U_j - \bar{U})$ and $S_{UU} = n^{-1}\sum(U_j - \bar{U})^2$, *etc.* We proceed under the minimal assumptions that: 1) $S_{YY} = \beta^2\sigma_U^2 + \sigma_\epsilon^2 + o_p(1)$; 2) $\sqrt{n}(S_{YX} - S_{UU}\beta) = O_p(1)$; and 3) $\sqrt{n}(S_{XX} - S_{UU} - (n-1)\sigma^2/n) = O_p(1)$. Note that 3) implies that $S_{XX} = \sigma_U^2 + \sigma^2 + o_p(1)$.

We consider estimating $\theta = \beta$. Thus

$$\hat{\theta}_{\text{TRUE}} = \frac{S_{YU}}{S_{UU}}.$$

The estimator $\hat{\theta}_{\text{TRUE}}$ is unbiased in finite samples. However, the results of Section 6.2 do not apply since the estimator is not a sufficiently smooth function of $U_1, \ldots, U_n$. Indeed, if we view $\hat{\theta}_{\text{TRUE}}$ as a function of U_1 alone, then it has the form $g(U_1)$ where $g(u_1) = (du_1+e)/(au_1^2+bu_1+c)$ for suitably defined real constants $a, \ldots, e$. The quadratic in the denominator has either one real root or two complex roots, since $S_{UU} \geq 0$. In either case the function $g(w)$ of the complex variable w has at least one singularity in the complex plane and this violates the assumed smoothness (Stefanski, 1989).

Define $\tau^2 = \sigma_\epsilon^2/(nS_{UU})$ and $\hat{\tau}^2_{\text{TRUE}} = (nS_{UU})^{-1}\hat{\sigma}^2_{\epsilon,\text{TRUE}}$, where $\hat{\sigma}^2_{\epsilon,\text{TRUE}}$ is the true-regression mean squared error. For this problem

$$\hat{\theta}_b(\lambda) = \frac{\hat{N}_b(\lambda)}{\hat{D}_b(\lambda)}, \tag{6.3}$$

where

$$\hat{N}_b(\lambda) = \frac{1}{n}\sum_{j=1}^{n}(Y_j - \bar{Y})(X_j + \sigma\sqrt{\lambda}\, Z_{b,j})$$

$$\hat{D}_b(\lambda) = \frac{1}{n}\sum_{j=1}^{n}\{(X_j - \bar{X}) + \sigma\sqrt{\lambda}\,(Z_{b,j} - \bar{Z}_b)\}^2.$$

Define

$$\hat{N}(\lambda) = E\{\hat{N}_b(\lambda) \mid \{Y_j, X_j\}\} = S_{YX},$$

$$N(\lambda) = E\{S_{YX}\} = S_{UU}\beta,$$

$$\hat{D}(\lambda) = E\{\hat{D}_b(\lambda) \mid \{Y_j, X_j\}\} = S_{XX} + \left(\frac{n-1}{n}\right)\lambda\sigma^2,$$

$$D(\lambda) = E\{S_{XX} + \left(\frac{n-1}{n}\right)\lambda\sigma^2\} = S_{UU} + \left(\frac{n-1}{n}\right)(\lambda+1)\sigma^2.$$

The Cauchy-Schwartz inequality is used to show that

$$\hat{\theta}_b^2(\lambda) \leq \frac{nS_{YY}}{\sum_{j=1}^{n}\{X_j - \bar{X} + \sqrt{\lambda}\,\sigma(Z_{b,j} - \bar{Z}_b)\}^2}. \tag{6.4}$$

Conditioned on $\{Y_j, X_j\}_1^n$, the right hand side of (6.4) is proportional to the reciprocal of a noncentral Chi-squared random variable with $n-1$ degrees of freedom and thus has its first k moments finite provided $n > 1 + 2k$. So for $k > 0$, $\hat{\theta}_b(\lambda)$ possesses a finite $2k$ conditional moment, provided $n > 1 + 2k$. In other words, the random quantities $\hat{\theta}(\lambda)$ and $s^2_\Delta(\lambda)$ are well defined for $n > 3$.

Define

$$\tilde{\theta}_b(\lambda) = \frac{N(\lambda)}{D(\lambda)} + \frac{\hat{N}_b(\lambda) - N(\lambda)\hat{D}_b(\lambda)/D(\lambda)}{D(\lambda)}.$$

Then $\hat{\theta}_b(\lambda) = \tilde{\theta}_b(\lambda) + \tilde{R}_b(\lambda)$ where

$$\tilde{R}_b(\lambda) = \left(\frac{\hat{D}_b(\lambda) - D(\lambda)}{\hat{D}_b(\lambda)}\right)\left(\frac{\hat{N}_b(\lambda) - N(\lambda)\hat{D}_b(\lambda)/D(\lambda)}{D(\lambda)}\right).$$

Letting $\tilde{\theta}(\lambda) = E\{\tilde{\theta}_b(\lambda) \mid \{Y_j, X_j\}_1^n\}$ and $\tilde{R}(\lambda) = E\{\tilde{R}_b(\lambda) \mid \{Y_j, X_j\}_1^n\}$ we have that

$$\hat{\theta}(\lambda) = E\{\hat{\theta}_b(\lambda) \mid \{Y_j, X_j\}_1^n\} = \tilde{\theta}(\lambda) + \tilde{R}(\lambda)$$

where

$$\tilde{\theta}(\lambda) = \frac{N(\lambda)}{D(\lambda)} + \frac{\hat{N}(\lambda) - N(\lambda)\hat{D}(\lambda)/D(\lambda)}{D(\lambda)}.$$

In the Appendix it is shown that the remainder terms, $\tilde{R}_b(\lambda)$ and $\tilde{R}(\lambda)$, can be ignored asymptotically. Thus upon employing the linearizations $\tilde{\theta}_b(\lambda)$ and $\tilde{\theta}(\lambda)$ it follows that

$$\begin{aligned} s_\Delta^2(\lambda) &= \mathrm{Var}\{\hat{\theta}_b(\lambda) - \hat{\theta}(\lambda) \mid \{Y_j, X_j\}_1^n\} \\ &\approx \mathrm{Var}\{\tilde{\theta}_b(\lambda) - \tilde{\theta}(\lambda) \mid \{Y_j, X_j\}_1^n\} \\ &= \frac{\lambda\sigma^2}{nD^2(\lambda)}\left\{S_{YY} + \frac{N^2(\lambda)}{D^2(\lambda)}(4S_{XX} + 2\lambda\sigma^2) - 4\frac{N(\lambda)S_{YX}}{D(\lambda)}\right\}. \end{aligned}$$

For the linear regression model

$$\hat{\tau}_b^2(\lambda) = \frac{S_{YY}}{n\hat{D}_b(\lambda)} - \frac{\hat{\theta}_b^2(\lambda)}{n},$$

and it is readily seen that

$$\lim_{n\to\infty} n\hat{\tau}_b^2(\lambda) = \frac{\sigma_\epsilon^2\sigma_U^2 + (\lambda+1)\sigma^2(\beta^2\sigma_U^2 + \sigma_\epsilon^2)}{[\sigma_U^2 + (\lambda+1)\sigma^2]^2}.$$

Of course $n\hat{\tau}^2(\lambda)$ also has the same limit asymptotically.

Assuming that the exact extrapolant function is employed, the validity of SIMEX variance estimator depends on

$$\lim_{n\to\infty} n\{\hat{\tau}^2(-1) - s_\Delta^2(-1)\} = \frac{\sigma_\epsilon^2(\sigma_U^2 + \sigma^2) + \beta^2\sigma^2(\sigma_U^2 + 2\sigma^2)}{\sigma_U^4}.$$

The indicated limit coincides with the asymptotic variance reported in Fuller(1987, pp 15–17).

The function, $n\{\hat{\tau}_b^2(\lambda) - s_\Delta^2(\lambda)\}$, that must be extrapolated to obtain the SIMEX variance estimate is clearly not quadratic asymptotically. However, it is well approximated by a quadratic over a useful range of conditions. We will not provide evidence of this claim electing instead to address this issue in a more complicated problem in the next section.

6.3.2. SIMEX Variance Estimation in Simple Logistic Regression

Logistic regression provides an example for which SIMEX theory is exact in neither finite samples or asymptotically, yet has been shown by Cook and Stefanski (1992) to be competitive with estimation methods that are optimal asymptotically. We now present an example demonstrating the SIMEX variance estimator based on a quadratic extrapolant in a nontrivial application.

We make use of data from the Framingham Heart Study. The model we study relates Y=CHD, an indicator of coronary heart disease over an eight-year period following enrollment, to systolic blood pressure (SBP).

The data contain blood pressure measurements at two-year intervals during the study. We make use of the first two such measurements and work under the assumption that on a logarithmic scale these are replicate measurements; see Carroll and Stefanski (1993) for a discussion of the replicate measurements assumption. Thus with X_1 and X_2 denoting the logarithms of the two blood pressure measurements, our working assumptions are that $X_1 = U + \sqrt{2}\sigma Z_1$ and $X_2 = U + \sqrt{2}\sigma Z_2$ where Z_1 and Z_2 are independent measurement errors. Thus the best measurement, $X = (X_1 + X_2)/2$, has mean U and variance σ^2. Implied by our assumptions is the definition of U as the long-term average ln(SBP).

We have subsetted the data, selecting data on males only, and eliminating cases with missing information. The resulting sample size is 1615. Since it is unlikely that data are missing at random, we view our analysis as illustrative only.

A components-of-variance analysis results in the estimate $\hat{\sigma}^2 = 3.15 \times 10^{-3}$ corresponding to an estimated linear model correction-for-attenuation (the inverse of the reliability ratio) of 1.21.

Table 2 displays the results of three estimator/variance estimator combinations along with bootstrap variance estimates. The NAIVE procedure refers to ordinary logistic regression of Y on X (= the average of the two measurements). Standard errors were obtained from the usual large-sample, inverse-information variance matrix estimate.

The SUFFICIENCY procedure refers to the method based on sufficient statistics as described in Stefanski and Carroll (1985, 1987). This is a conditional likelihood-based procedure and is known to be asymptotically optimal under certain conditions. It provides a benchmark with which to compare the SIMEX procedure. Standard errors were obtained as for the NAIVE procedure employing the inverse, conditional-information matrix.

The SIMEX procedure employed quadratic-extrapolant parameter and variance estimation as described in Section 6.2, with T and T_{VAR} the usual, *i.e.*, non-measurement error, maximum likelihood estimator and inverse information matrix respectively, and with $B = 1000$.

In both the SUFFICIENCY and SIMEX procedures, the estimate $\hat{\sigma}^2$ was employed as if it were the known. This is not unreasonable in light of the sample size.

Five hundred bootstrap samples were drawn and analyzed for the purpose of obtaining bootstrap standard errors for each of the three procedure.

Table 2 shows that the two measurement-error procedures result in nearly identical analyses that differ predictably from the NAIVE procedure. The almost-constant differences between the bootstrap standard errors and the procedure-based standard errors, and the relative magnitudes of these differences suggest two conclusions. First is that the SIMEX standard error estimates are comparably to the conditional likelihood-based estimates. Second is that all of the procedure-based standard error estimates do an adequate job of estimating variability.

7. SUMMARY

The close relationship between SIMEX and the jackknife established in this paper adds much credibility to the former. The SIMEX simulation step, corresponds to the jackknife leave-j-out step, and the extrapolation steps in both procedures are conceptually identical.

In addition to providing theoretical justification for SIMEX estimation, the investigation of SIMEX as a variant of the jackknife made obvious the variance estimation procedure described in Section 6. Except for the fact that the theory assumes σ^2 to be known, SIMEX provides a widely applicable method for statistical inference in the presence of measurement error.

In situations where σ^2 is estimated and the estimation variation is not negligible, it alsways possible to jackknife, or bootstrap, the combined estimation procedure, *i.e.*, estimation of σ^2 followed by SIMEX estimation. Alternatively, the large-sample distribution theory developed by Lombard *et al.* (1993) yields asymptotically valid standard errors for a large class of SIMEX estimators with estimated σ^2.

8. REFERENCES

Cook, J. R. and Stefanski, L. A. (1992). Simulation Extrapolation Estimation in Parametric Measurement Error Models, *Journal of the American Statistical Association*, in review.

Efron, B. and Stein, C. (1981). The Jackknife Estimate of Variance, *Annals of Statistics*, 9, 586–96.

Efron, B. (1982). *The Jackknife, the Bootstrap and Other Resampling Plans*, SIAM, Philadelphia.

Fuller, W. A. (1987). *Measurement Error Models.* Wiley, New York.

Gray, H., Watkins, T. and Adams, J. (1972). On the Generalized jackknife, Its Extensions, and Its

Relation to e_n-transformations, *Annals of Mathematical Statistics*, 43, 1–30.

Gray, H., Schucany, W. and Watkins, T. (1975). On the Generalized jackknife and Its Relation to Statistical Differentials, *Biometrika*, 62, 637–42.

Lombard, F., Carroll, R. J. and Stefanski, L. A. (1993). Asymptotics for the SIMEX Estimator in Structural Measurement Error Models, in review.

Quenouille, M. (1956). Notes on Bias and estimation, *Biometrika*, 43, 353–60.

Stefanski, L. A. and Carroll, R. J. (1985). Covariate Measurement Error in Logistic Regression, *The Annals of Statistics*, 13, 1335–1351.

Stefanski, L. A. and Carroll, R. J. (1987). Conditional Scores and Optimal Scores for Generalized Linear Measurement-Error Models. *Biometrika*, 74, 703–716.

Stefanski, L. A. (1989). Unbiased Estimation of a Nonlinear Function of a Normal Mean with Application to Measurement Error Models, *Communications in Statistics, Theory and Methods*, 18, 4335–4358.

Tukey, J. (1958). Bias and Confidence in Not Quite Large Samples, *Annals of Mathematical Statistics*, 29, 614.

9. APPENDIX

We now show that the remainder terms, $\tilde{R}_b(\lambda)$ and $\tilde{R}(\lambda)$, encountered in Section 6.3.1 can be ignored asymptotically. Note that it is not sufficient to show that these are $o_p(n^{-1/2})$. The variance estimation method depends on the convergence in probability of the scaled conditional variance, $ns^2_\Delta(\lambda) = n\text{Var}\{\hat{\theta}_b(\lambda) - \hat{\theta}(\lambda) \mid \{Y_j, X_j\}_1^n\}$, $\lambda > 0$. Thus it must be shown that $E\{n[\tilde{R}_b(\lambda) - \tilde{R}(\lambda)]^2 \mid \{Y_j, X_j\}_1^n\} = o_p(1)$. However, since $E\{n[\tilde{R}_b(\lambda) - \tilde{R}(\lambda)]^2 \mid \{Y_j, X_j\}_1^n\} = E\{n\tilde{R}_b^2(\lambda) \mid \{Y_j, X_j\}_1^n\} - n\tilde{R}^2(\lambda)$ and $n\tilde{R}^2(\lambda) \le E\{n\tilde{R}_b^2(\lambda) \mid \{Y_j, X_j\}_1^n\}$ almost surely, it is sufficient to show that $E\{n\tilde{R}_b^2(\lambda) \mid \{Y_j, X_j\}_1^n\} = o_p(1)$.

For the remainder of this section let $E\{\cdot\}$ denote conditional expectation $E\{\cdot \mid \{Y_j, X_j\}_1^n\}$, and let all equalities and inequalities be interpreted in an almost sure sense.

The Cauchy-Schwartz inequality and the inequality $(a+b)^4 \le 8(a^4 + b^4)$ are used to show that $E\{n\tilde{R}_b^2(\lambda)\}$ is bounded above by

$$8E\left\{\left(\frac{\hat{D}_b(\lambda) - D(\lambda)}{\hat{D}_b(\lambda)}\right)^4\right\} E\left\{\left[\frac{n^{1/2}[\hat{N}_b(\lambda) - N(\lambda)]}{D(\lambda)}\right]^4 + \left[\frac{N(\lambda)}{D(\lambda)}\frac{n^{1/2}[\hat{D}_b(\lambda) - D(\lambda)]}{D(\lambda)}\right]^4\right\}.$$

Thus it is sufficient to show that

$$E\{(\sqrt{n}[\hat{N}_b(\lambda) - N(\lambda)])^4\} = O_p(1);$$
$$E\{(\sqrt{n}[\hat{D}_b(\lambda) - D(\lambda)])^4\} = O_p(1);$$
$$E\left\{\left(\frac{\hat{D}_b(\lambda) - D(\lambda)}{\hat{D}_b(\lambda)}\right)^4\right\} = o_p(1).$$

Now conditioned on $\{Y_j, X_j\}_1^n$, $\sqrt{n}[\hat{N}_b(\lambda) - N(\lambda)]$ is normally distributed with mean $M_C = \sqrt{n}(S_{YX} - S_{UU}\beta)$ and variance $V_C = \lambda\sigma^2 S_{YY}$. Its fourth (conditional) moment is thus bounded by $8(M_C^4 + 3V_C^2)$ which is $O_p(1)$ provided $\sqrt{n}(S_{YX} - S_{UU}\beta)$ and S_{YY} are $O_p(1)$.

Define $\tau_n = nS_{XX}/(2\lambda\sigma^2)$. Then conditioned on $\{Y_j, X_j\}_1^n$, $\sqrt{n}[\hat{D}_b(\lambda) - D(\lambda)]$ is equal in distribution to

$$A_n = \sqrt{n}\left\{\frac{\lambda\sigma^2}{n}\left[\chi^2_{(n-1)}(\tau_n) - 2\tau_n - (n-1)\right] + \left(S_{XX} - S_{UU} - \left(\frac{n-1}{n}\right)\sigma^2\right)\right\},$$

where $\chi^2_{(n-1)}(\tau_n)$ denotes a noncentral Chi-squared random variable with noncentrality τ_n.

The inequality $(a+b)^4 \leq 8(a^4+b^4)$, and evaluation of the fourth central moment of a noncentral Chi-squared distribution are used to show that $E\{(\sqrt{n}[\hat{D}_b(\lambda) - D(\lambda)])^4\} \leq 8(B_n + C_n)$ where

$$B_n = \frac{\sigma^8\lambda^4}{n^2}\{48(8\tau_n + n - 1) + 3[8\tau_n + 2(n-1)]^2\},$$
$$C_n = (\sqrt{n}[S_{XX} - S_{UU} - (n-1)\sigma^2/n)])^4.$$

Thus $E\{(\sqrt{n}[\hat{D}_b(\lambda) - D(\lambda)])^4\} = O_p(1)$ provided C_n and τ_n/n are $O_p(1)$.

Note that

$$\left[\frac{\hat{D}_b(\lambda) - D(\lambda)}{\hat{D}_b(\lambda)}\right]^4 = \left[1 - \frac{D(\lambda)}{\hat{D}_b(\lambda)}\right]^4$$

and thus to show that

$$E\left\{\left(\frac{\hat{D}_b(\lambda) - D(\lambda)}{\hat{D}_b(\lambda)}\right)^4\right\} = o_p(1)$$

it is sufficient to show that $E\{[D(\lambda)/\hat{D}_b(\lambda)]^j\} = 1 + o_p(1)$ for $j = 1, \ldots, 4$. We present the proof for $j = 4$ only. Let τ_n and $\chi^2_{(n-1)}(\tau_n)$ be defined as above. We must show that

$$\left[S_{UU} + \left(\frac{n-1}{n}\right)(\lambda+1)\sigma^2\right]^4 E\left\{n^4\left[\lambda\sigma^2\chi^2_{(n-1)}(\tau_n)\right]^{-4}\right\} = 1 + o_p(1). \quad (9.1)$$

When $n > 9$ the indicated expectation exists. Furthermore,

$$E\left\{n^4\left[\chi^2_{(n-1)}(\tau_n)\right]^{-4}\right\} = n^4\sum_{j=1}^{\infty}\frac{\exp(-\tau_n)\tau_n^j}{j!}\frac{1}{(n+2j-3)\times\cdots\times(n+2j-9)}.$$

Let a_k be the coefficient of t^{4-k} in the expansion of $(1-t)^3/48$, $k = 1, \ldots, 4$. Then

$$\sum_{k=1}^{4} \frac{a_k}{n+2j-2k-1} = \frac{1}{(n+2j-3)\times\cdots\times(n+2j-9)}.$$

For $0 \le s \le 1$ define the generating function

$$g_n(s) = n^4 \sum_{j=1}^{\infty} \frac{\exp(-\tau_n)\tau_n^j}{j!} \sum_{k=1}^{4} \frac{a_k s^{n+2j-2k-1}}{n+2j-2k-1}.$$

Note that $g_n(1) = E\left\{n^4\left[\chi^2_{(n-1)}(\tau_n)\right]^{-4}\right\}$. The derivative of $g_n(s)$ with respect to s, $g_n'(s)$, exists (almost surely) and furthermore

$$\begin{aligned} g_n'(s) &= n^4 s^{n-10} \sum_{j=1}^{\infty} \frac{\exp(-\tau_n)(\tau_n s^2)^j}{j!} \sum_{k=1}^{4} a_k (s^2)^{4-k} \\ &= \frac{n^4 s^{n-10}(1-s^2)^3}{48} \exp(\tau_n(s^2-1)). \end{aligned}$$

Integrating, making the change-of-variables $y = -\tau_n(s^2-1)$, and appealling to the Lebesgue Dominated Convergence Theorem using the fact that $\tau_n/n \to (\sigma_U^2+\sigma^2)/(2\sigma^2\lambda)$ in probability, we find that

$$\begin{aligned} g_n(1) &= n^4 \int_0^1 \frac{n^4 s^{n-10}(1-s^2)^3}{48} \exp(\tau_n(s^2-1))\, ds \\ &= \frac{n^4}{96\tau_n^4} \int_0^{\tau_n} \left(1-\frac{y}{\tau_n}\right)^{(n-11)/2} y^3 \exp(-y)\, dy \\ &\to \frac{1}{96}\left(\frac{2\lambda\sigma^2}{\sigma_U^2+\sigma^2}\right)^4 \int_0^{\infty} y^3 \exp\left(-y - \frac{y2\sigma^2\lambda}{2(\sigma_U^2+\sigma^2)}\right) dy \\ &= \frac{(\lambda\sigma^2)^4}{[\sigma_U^2+\sigma^2(\lambda+1)]^4}. \end{aligned}$$

The '$\to$' in the preceeding set of equations denotes convergence in probability. In light of (9.1) it follows that $E\{[D(\lambda)/\hat{D}_b(\lambda)]^4\} = 1 + o_p(1)$.

TABLE 1. COMPARISON OF SIMEX AND JACKKNIFE ESTIMATORS

Results from the simulation described in Section 5.1. MLE, maximum likelihood estimator; LJ, linear jackknife; QJ3–QJ5, quadratic jackknife base on 3–5 points; SIMEX1, SIMEX with estimated variance; SIMEX2, SIMEX with known variance.

	MLE	LJ	QJ3	QJ4	QJ5	SIMEX1	SIMEX2
Mean	1.105	0.994	1.001	1.002	1.004	1.003	1.000
Variance	0.275	0.231	0.233	0.233	0.235	0.233	0.225
Mean Squared Error	0.286	0.231	0.233	0.233	0.235	0.233	0.225

TABLE 2. LOGISTIC REGRESSION VARIANCE ESTIMATION

Variance estimation example from Section 6.3.2. NAIVE, naive procedure; SUFFICIENCY, sufficiency procedure; SIMEX, SIMEX procedure; Information, inverse information-based standard errors; Bootstrap, bootstrap standard errors based on 500 resampled data sets.

	NAIVE		SUFFICIENCY		SIMEX	
	Intercept	Slope	Intercept	Slope	Intercept	Slope
Estimate	-18.89	3.37	-21.23	3.85	-21.04	3.81
Information	2.94	0.60	3.31	0.67	3.29	0.67
Bootstrap	2.77	0.56	3.08	0.63	3.07	0.63

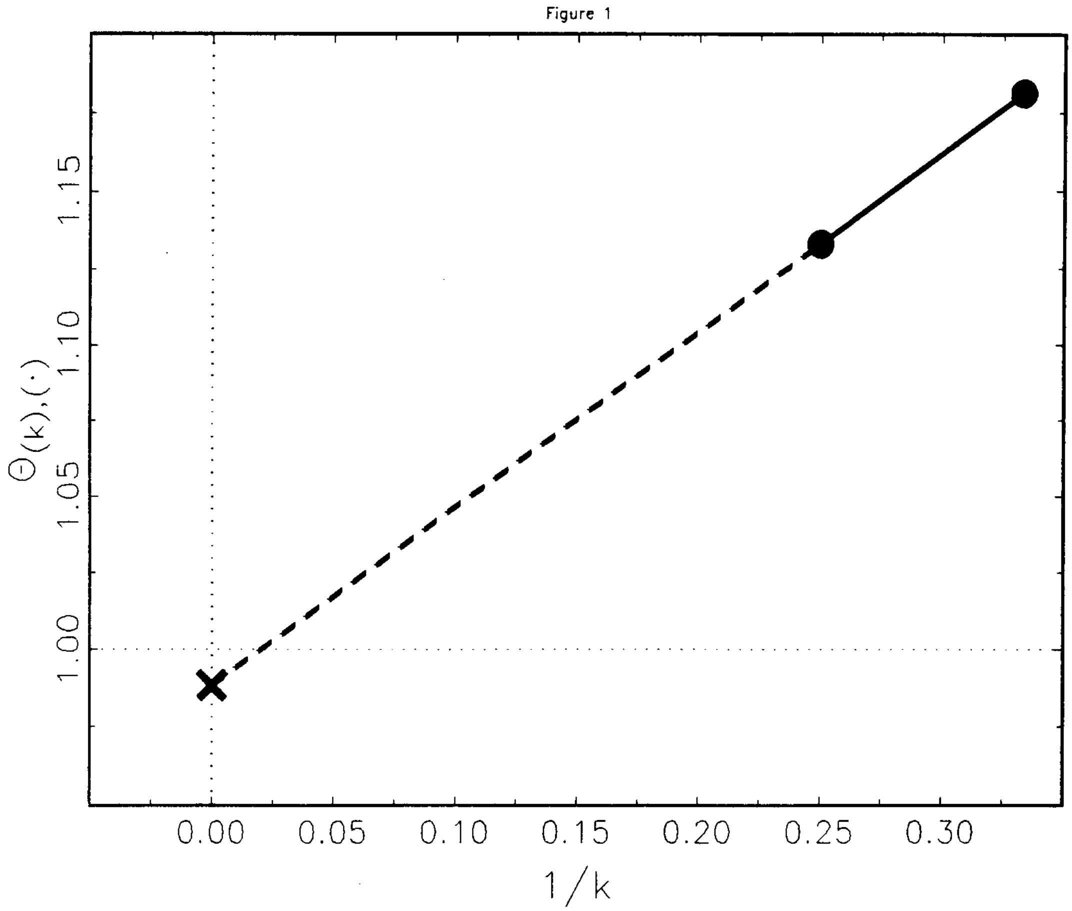

FIGURE 1. JACKKNIFE EXTRAPOLATION

The solid line determined by the plot of $\bar{\hat{\theta}}_{(k),(\cdot)}$ versus $1/k$ (circles) for $k = 4$ (the maximum likelihood estimate) and $k = 3$ (the average of the 4 leave-1-out estimates) is extrapolated (dashed line) to 0 ($= 1/\infty$) thereby obtaining the jackknife estimate (cross).

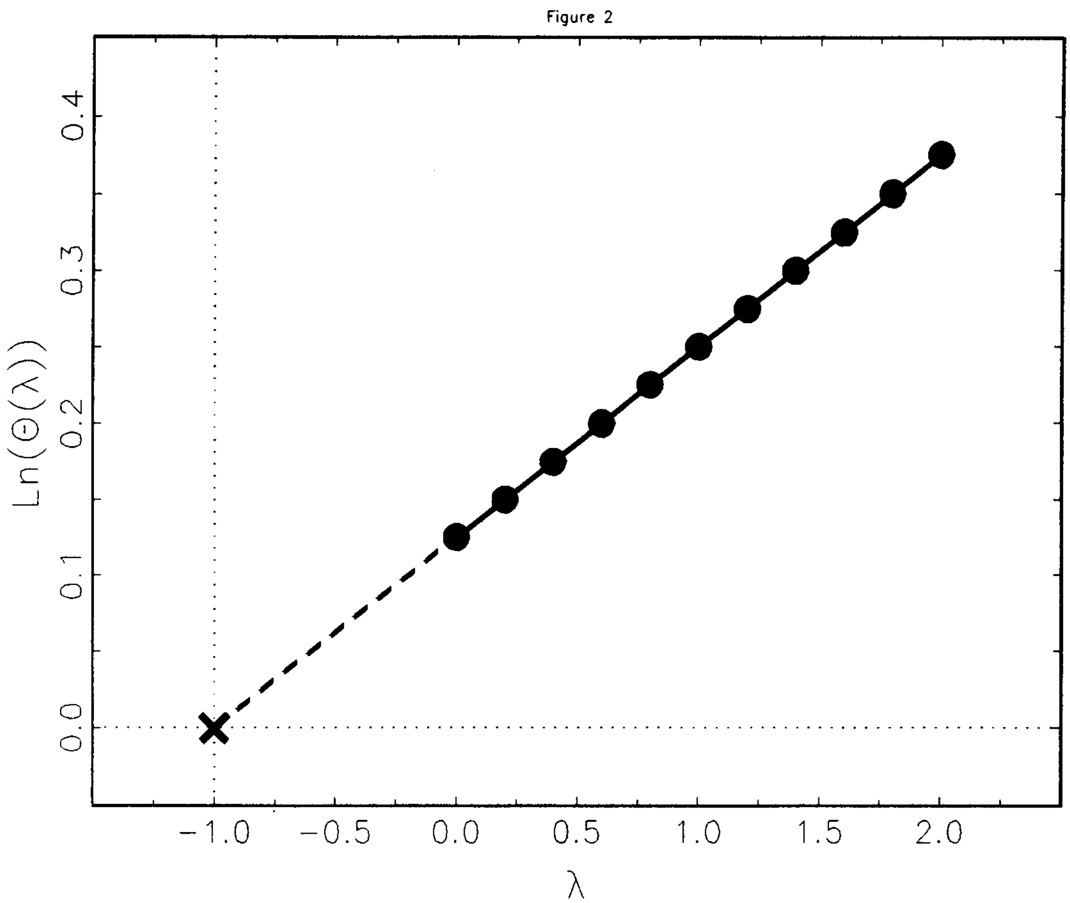

FIGURE 2. SIMEX EXTRAPOLATION

The solid line determined by the plot of $\ln(\hat{\theta}(\lambda))$ versus λ (circles) for eleven, equally-spaced values of λ spanning [0, 2], is extrapolated (dashed line) to $\lambda = -1$ thereby obtaining the natural logarithm of the SIMEX estimate (cross).

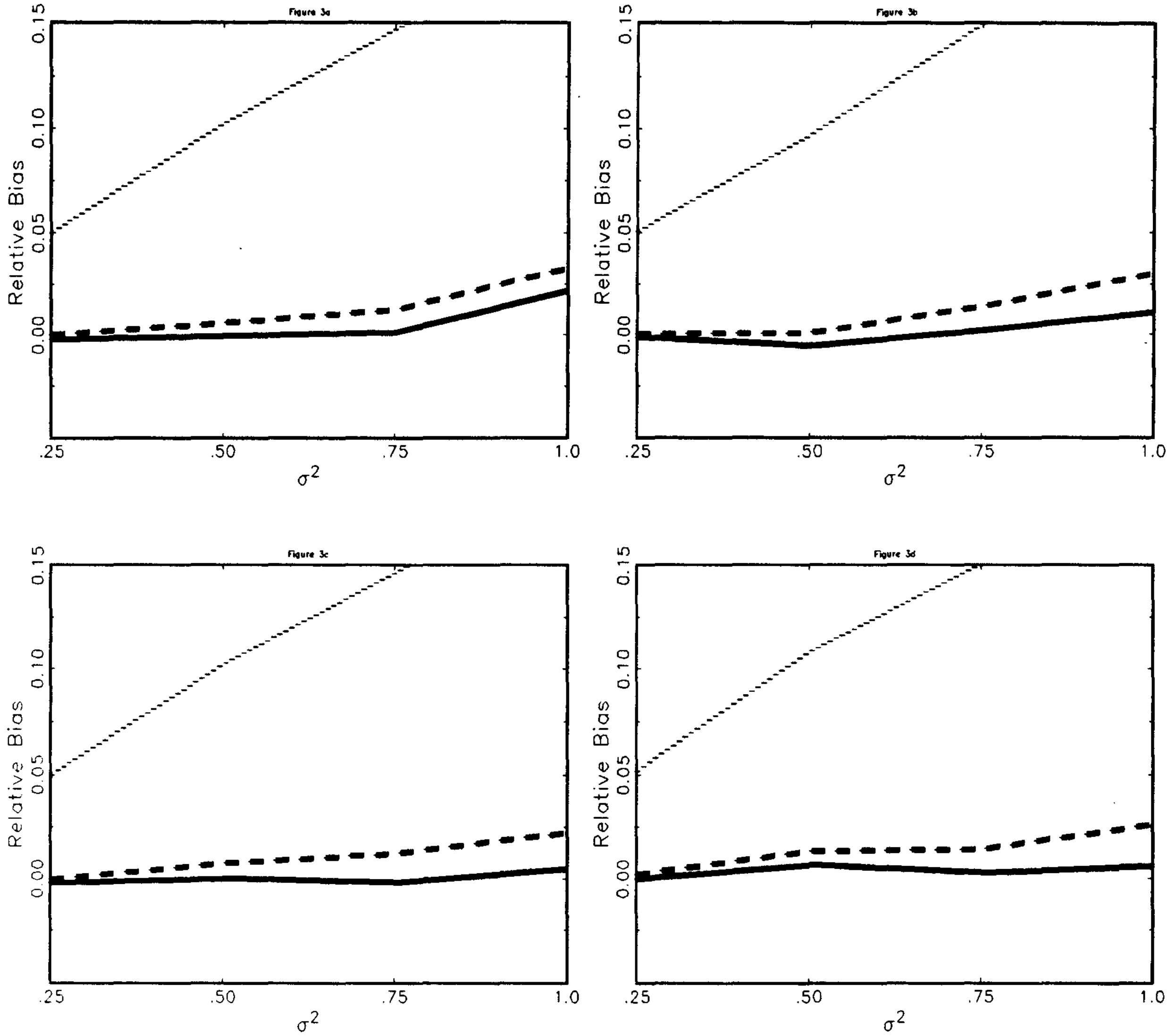

FIGURE 3. RELATIVE BIAS PLOTS

Results from the simulation described in Section 5.3.1. Clockwise from top left, $d = 1, 16, 256, \infty$. Dotted lines, maximum likelihood; Dashed lines, linear jackknife; solid lines, SIMEX. Unbiased estimation corresponds to a horizontal line through the origin.

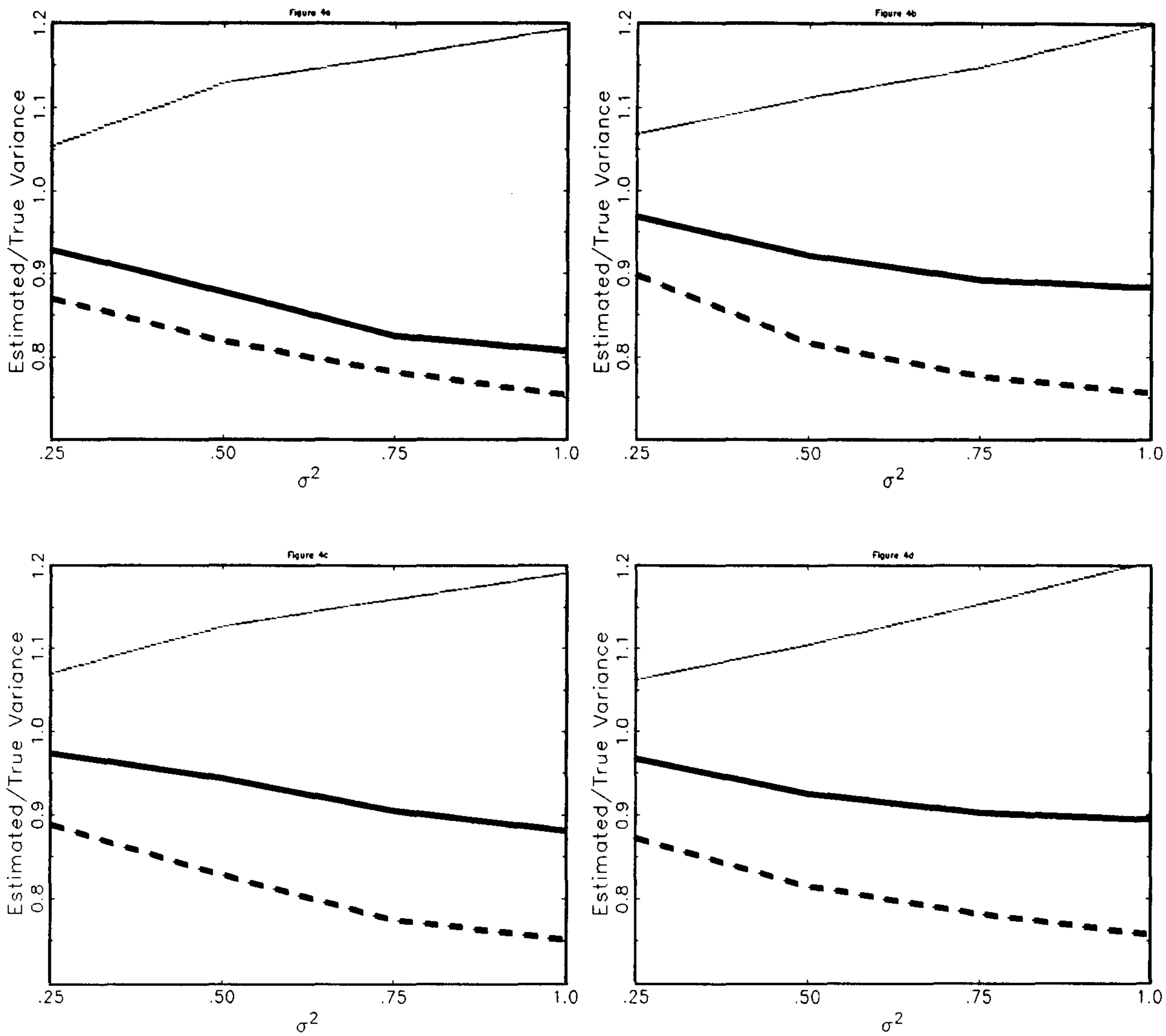

FIGURE 4. RELATIVE VARIANCE PLOTS

Results from the simulation described in Section 5.3.1. Clockwise from top left, $d = 1, 16, 256, \infty$. Dotted lines, maximum likelihood; Dashed lines, linear jackknife; solid lines, SIMEX. Unbiased estimation corresponds to a horizontal line with intercept at 1.

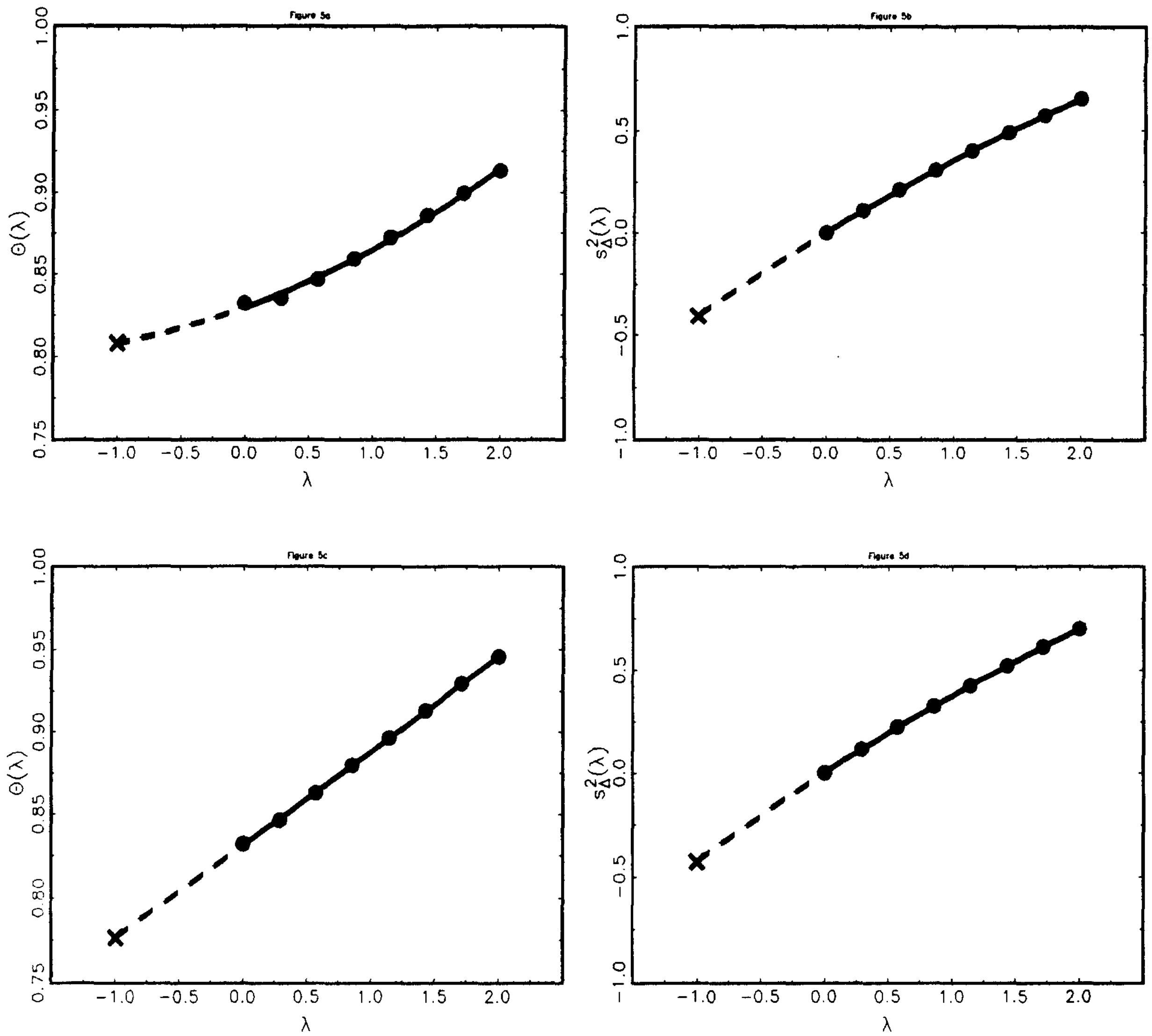

FIGURE 5. MEAN AND VARIANCE EXTRAPOLANTS

Sample mean and variance functions fit in the simulation described in Section 5.3.1 for $\sigma^2 = 0.5$, $d = 256$. Top left, $B = 1000$; quadratic extrapolant (solid line) fit to the plot of $\hat{\theta}(\lambda)$ versus λ (circles) for eight, equally-spaced values of λ spanning [0, 2], is extrapolated (dashed line) to $\lambda = -1$ thereby obtaining the SIMEX estimate (cross). Top right, $B = 1000$; quadratic extrapolant (solid line) fit to the plot of $s^2_\Delta(\lambda)$ versus λ (circles) for eight, equally-spaced values of λ spanning [0, 2], is extrapolated (dashed line) to $\lambda = -1$ thereby obtaining the SIMEX variance estimate (cross). Bottom left and right; same as top left and right with $B = 16,000$.