Electronic Journal of Statistics
Vol. 8 (2014) 1973–1995
ISSN: 1935-7524
DOI: 10.1214/14-EJS941

Simulation extrapolation estimation in parametric models with Laplace measurement error*

Hira L. Koul and Weixing Song

Michigan State University and Kansas State University
e-mail: kou@stt.msu.edu; weixing@ksu.edu

Abstract: The normal-error-based simulation extrapolation (N-SIMEX) procedure provides a simulation-based method to remove or reduce the bias in estimators of parameters in measurement error models. This paper shows that the N-SIMEX procedure only works for the normal measurement error and does not work for Laplace measurement error. A new L-SIMEX procedure is particularly designed to remove or reduce the Laplace measurement errors in parametric models. Unlike in the N-SIMEX procedure where the measurement error is removed or reduced by adding independent normal errors controlled by the scale parameter, the proposed procedure removes or reduces the Laplace measurement error by adding a noise variable which is a difference between two independent gamma random variables, and where the noise level is governed by the shape parameter. Heuristic and rigorous arguments are provided to justify the proposed method and a Jackknife-type estimation procedure is provided to estimate the variance of the L-SIMEX estimate. Simulation studies and a real data example are presented to demonstrate the proposed estimation procedure. A finite sample comparison with some revised moment estimators of Hong and Tamer is also included.

MSC 2010 subject classifications: Primary 62F10, 62F40; secondary 62G09.
Keywords and phrases: Measurement error, Laplace distribution, simulation extrapolation, method of moments, jackknife.

Received February 2014.

Contents

*The research is supported by the collaborative grants from NSF DMS 1205271, 1205276.

1. Introduction

Due to wide applicability of measurement error models, the interest of searching for efficient statistical inference methodologies in these models has never receded. It is well known that simply ignoring the measurement error from the inference will lead to biased estimation and inefficient statistical inference. How to effectively remove or reduce the influence of the measurement error on the inference procedures is one of the main objective in this very important area of statistical inference. Among several promising procedures, the deconvolution and normal-error-based SIMEX (N-SIMEX) procedures are commonly used.

To proceed a bit more precisely, in the measurement error model of interest here, we have a response variable Y, an observable predicting variable W, and an unobservable covariate X, called latent variable. In this article, all of these variables are assumed to be scalars. Instead of observing X, we observe $Z = X + U$, where U is the measurement error. To achieve the bias reduction in an estimator of an underlying parameter, the deconvolution procedure tries to estimate the density function of the latent variable X. Assuming that the density function of the measurement error U is known and its characteristic function does not vanish on the whole real line, the characteristic function of X can be written as the ratio of the characteristic functions of Z and U. The characteristic function of Z can be smoothly estimated based on the observations on Z. Then, the deconvolution kernel density estimate is defined as the inverse Fourier transform of the ratio of the estimated characteristic function with that of U. Theoretical properties of the deconvolution density estimate are thoroughly discussed in [1, 5, 6, 7, 12, 19], among others.

Due to its mathematical complexity and burdensome technical details, applied statisticians might find the deconvolution procedures hard to understand intuitively, and when dealing with relatively simpler models, using deconvolution kernel seems like employing a steam engine to crack a nut. As a comparison, the N-SIMEX procedure proposed by [3] successfully avoids the technical complexity, and provides an easy-to-implement simulation-based method to remove or reduce the bias in parametric measurement error models. In N-SIMEX, the estimates are obtained by first adding an extra normal measurement error to the contaminated data in a resampling stage, where the variability of the normal measurement error is controlled by the scale parameter; then some naive-like estimates are constructed based on the noise-enhanced data set, and a trend of measurement error-induced bias versus the variance of the added normal measurement error is established; finally, the trend is extrapolated back to the case of no measurement error. As [3] described, the N-SIMEX procedure "...is completely general, it is also useful in applications when the particular model under consideration is novel and conventional approaches to estimation with the model have not been thoroughly studied and developed." In some simple cases, the N-SIMEX procedure does remove the bias completely. For example, in the multivariate linear regression setup, [18] showed that the difference between the N-SIMEX estimate and the bias-corrected moment estimate is asymptotically negligible even if the measurement error is not normal, since only the first two moments of the measurement error is needed in the proof. But in some other cases, in particular, if the true extrapolation function is not available, it only provides approximate solutions. Nevertheless, the N-SIMEX procedure certainly provides us a very useful exploratory tool for the model fitting in the measurement error context.

For the sake of completeness and ease of presentation, we briefly describe the N-SIMEX procedure as follows. With $Y, W, X, Z = X + U$ as described above, N-SIMEX assumes $U \sim N(0, \sigma_u^2)$, where σ_u^2 is known or can be reasonably estimated from other resources. Suppose θ, possibly multidimensional, is the model parameter to be estimated, and there is an existing estimation procedure

$$\hat{\theta}_{\text{TRUE}} = T(\{Y_i, W_i, X_i\}_{i=1}^n)$$

based on the data $\{Y_i, W_i, X_i\}_{i=1}^n$ which are i.i.d. copies from (Y, W, X). It is assumed that $\hat{\theta}_{\text{TRUE}} \to \theta_0$ in probability. But $\hat{\theta}_{\text{TRUE}}$ is not an estimate since $\{X_i\}_{i=1}^n$ is not available. Upon replacing X_i in this formula by the observed Z_i one obtains the naive estimate of θ given by

$$\hat{\theta}_{\text{NAIVE}} = T(\{Y_i, W_i, Z_i\}_{i=1}^n).$$

Fix a $\lambda \geq 0$, define

$$Z_{b,i} = Z_i + \sqrt{\lambda} U_{b,i} = X_i + U_i + \sqrt{\lambda} U_{b,i}, \quad b = 1, 2, \ldots,$$

where $U_{b,i}$ are i.i.d $N(0, \sigma_u^2)$ r.v.'s, independent of all the other r.v.'s in the model. Define

$$\hat{\theta}_b(\lambda) = T(\{Y_i, W_i, Z_{b,i}\}_{i=1}^n),$$

and calculate

$$\hat{\theta}(\lambda) = E\Big[\hat{\theta}_b(\lambda) \Big| \{Y_i, W_i, Z_i\}_{i=1}^n\Big].$$

Note that the expectation is with respect to the distribution of $\{U_{b,i}\}_{i=1}^n$ only. The above expectation might not have explicit form, but it can be approximated arbitrarily well by the average of $\hat{\theta}_b(\lambda)$, $b = 1, 2, \ldots, B$, based on B random samples $\{U_{b,i}\}_{i=1}^n$, $b = 1, 2, \ldots, B$, from $N(0, \sigma_u^2)$. Repeat the above computation for a sequence of λ values. According to [3], a rule of thumb is to choose grid points $0 = \lambda_1 < \lambda_2 < \cdots < \lambda_{k-1} < \lambda_k = 2$, where k is a known positive integer. This is the simulation step in the N-SIMEX procedure.

In the extrapolation step, a trend of $\hat{\theta}(\lambda)$ with respect to λ is identified. Except for some special cases, the dependence of $\hat{\theta}(\lambda)$ on λ is not clear. However, in general, a scatter plot of $\hat{\theta}(\lambda_j)$ versus λ_j, $j = 1, 2, \ldots, k$, is helpful in determining an approximate trend, and the least squares methodology might be called upon in helping to decide a reasonable analytic form of the extrapolation function. Once the trend is set, then extrapolating the extrapolant back to -1 gives the SIMEX estimate. As [3] suggested, the nonlinear form $a + b/(c + \lambda)$ usually produces a good fit to the true extrapolant, while the linear and quadratic extrapolants often give conservative answers.

Many simulation studies and real data examples show that the above N-SIMEX proedure works surprisingly well even in some complicated models. See [3] and other references. The popularity of N-SIMEX procedure is further enhanced by the theoretical justifications developed by some renowned statisticians. [20] discovered a strong relationship between N-SIMEX and jackknife estimation. They also provided a crucial lemma which delineates the reason why the N-SIMEX works for normal measurement errors. By considering the problems in which the regression parameter estimates are the solution to unbiased estimating equations and using asymptotic linearization results, an asymptotic distribution theory was developed in [2] under the assumption that the true extrapolant is known, including the verification of asymptotic normality and implementation of standard error estimates. To our best knowledge, there is no discussion in the literature on how N-SIMEX performs when the measurement error is not normal, at least from the theoretical perspective.

In this paper, we address three important questions about the SIMEX procedure: (1). Does the N-SIMEX work only for normal measurement error? (2). If the answer to (1) is positive, is there any similar procedure working for non-normal errors? (3). Can sufficient theoretical and empirical evidence be found to justify the proposed procedure. The answers to these questions form the core of this paper.

2. A motivating example

For any generic random variable T, let σ_t^2 denote the variance of T, and $\mu_{4,t}$ the fourth central moment $E(T-E(T))^4$ of T. We say that a r.v. $U \sim \text{Laplace}(0, \sigma_u^2)$ if its density f_u is Laplace density with mean zero and variance σ_u^2 given by $f_u(v) = (\sqrt{2}/\sigma_u)\exp\{-(\sqrt{2}/\sigma_u)|v|\}$, $|v| < \infty$. Moreover, in the sequel, all limits are taken as $n \to \infty$, unless mentioned otherwise, and $\to_p$ denotes the convergence in probability.

To answer the first question mentioned above, we start with the components-of-variance model $Z = X + U$. Suppose the variance σ_x^2 and the fourth central moment $\mu_{4,x}$ of X are the parameters to be estimated. Let $\{Z_i\}_{i=1}^n$ be the observed data satisfying $Z_i = X_i + U_i$, where U_i are i.i.d. measurement errors with known density function with mean 0 and the known variance $\sigma_u^2 > 0$.

Consider the N-SIMEX procedure first and focus on the estimation of σ_x^2. Assume that $U \sim N(0, \sigma_u^2)$. Following the steps stated in Section 1, we generate a random sample $\{U_{b,i}\}_{i=1}^n$ of size n from $N(0, \sigma_u^2)$ distribution, independent of all other r.v.'s in the model, and for a $\lambda \geq 0$, define $Z_{b,i}(\lambda) = Z_i + \sqrt{\lambda} U_{b,i}$. Following the N-SIMEX procedure, an estimate of σ_x^2 is

$$\hat{\mu}_{2,b}(\lambda) = \frac{1}{n-1} \sum_{i=1}^n (Z_{b,i} - \bar{Z}_b)^2.$$

Direct computation shows that $\hat{\mu}_2(\lambda) = E\left\{\hat{\mu}_{2,b}(\lambda) | \{Z_i\}_{i=1}^n\right\} = S_Z^2 + \lambda \sigma_u^2$, where S_Z^2 is the sample variance of $\{Z_i\}_{i=1}^n$. This fact continues to hold when $\{U_{b,i}\}_{i=1}^n$ are i.i.d. Laplace$(0, \sigma_u^2)$ r.v.'s. This implies that to estimate the variance σ_x^2, the

N-SIMEX procedure still produces a reasonable estimator, even when the true distribution of the measurement error is non-normal.

Next, consider the estimation of the fourth moment $\mu_{4,x}$. From [17], an unbiased estimate of $\mu_{4,x}$ based on method of moments when $\{X_i\}_{i=1}^n$ are observable is

$$\hat{\mu}_{4,x} = a_n \sum_{i=1}^n (X_i - \bar{X})^4 - b_n \left[\sum_{i=1}^n (X_i - \bar{X})^2\right]^2,$$

where

$$a_n = \frac{n^2 - 2n + 3}{(n^2 - 5n + 6)(n-1)}, \quad b_n = \frac{3(2n-3)}{n(n^2 - 5n + 6)(n-1)}.$$

Following the N-SIMEX procedure, we define

$$\hat{\mu}_{4,b}(\lambda) = a_n \sum_{i=1}^n (Z_{b,i} - \bar{Z}_b)^4 - b_n \left[\sum_{i=1}^n (Z_{b,i} - \bar{Z}_b)^2\right]^2.$$

A tedious but routine computation shows that

$$\begin{aligned}
\hat{\mu}_4(\lambda) \quad = \quad & E\left[\hat{\mu}_{4,b}(\lambda) \Big| \{Z_i\}_{i=1}^n\right] = a_n \sum_{i=1}^n (Z_i - \bar{Z})^4 - b_n \left[\sum_{i=1}^n (Z_i - \bar{Z})^2\right]^2 \\
& + \left[\frac{6\lambda a_n (n-1)\sigma_u^2}{n} - 2\lambda b_n (n+1)\sigma_u^2\right] \sum_{i=1}^n (Z_i - \bar{Z})^2 \\
& + a_n \lambda^2 \left[\frac{(n-1)(n^2 - 3n + 3)}{n^2} \mu_{4,u} + \frac{3(2n-3)(n-1)}{n^2} \sigma_u^4\right] \\
& - b_n \lambda^2 \left[\frac{(n-1)^2}{n} \mu_{4,u} + \frac{(n-1)(n^2 - 2n + 3)}{n} \sigma_u^4\right].
\end{aligned}$$

Because $\mu_{4,u} = 3\sigma_u^4$ for the normal error,

$$\hat{\mu}_4(\lambda) \to_p \mu_4(\lambda) = \mu_{4,x} + 6\sigma_x^2\sigma_u^2 + 3\sigma_u^4 + 6\lambda\sigma_u^2(\sigma_x^2 + \sigma_u^2) + 3\lambda^2\sigma_u^4.$$

Clearly, as expected, $\mu_4(\lambda) \to \mu_{4,x}$, as $\lambda \to -1$.

Now, assume that the true distribution of the measurement error U_i is Laplace with mean 0 and variance σ_u^2, but we still generate $U_{b,i}$ from $N(0, \sigma_u^2)$. Then $EU_i^2 = \sigma_u^2$, $EU_i^4 = 6\sigma_u^4$ and an argument similar to the above leads to

$$\hat{\mu}_4(\lambda) \to \mu_4(\lambda) = \mu_{4,x} + 6\sigma_x^2\sigma_u^2 + 6\sigma_u^4 + 6\lambda\sigma_u^2(\sigma_x^2 + \sigma_u^2) + 3\lambda^2\sigma_u^4.$$

Now, extrapolating back to $\lambda = -1$ yields $\mu_4(\lambda) \to \mu_{4,x} + 3\sigma_u^4$. This implies that the N-SIMEX does not work for Laplace measurement errors, in the sense that it does not yield a consistent estimator of the parameter of interest.

These findings from this simple example are significant. First, simply applying blindly the N-SIMEX to measurement error models can be very misleading if the true distribution of the measurement errors is not normal; secondly, the N-SIMEX might be a very general procedure with respect to a variety of parametric models, but the assumption of having normal measurement error is crucial, indicating that the N-SIMEX procedure is not robust with respect to the distribution of the measurement errors; finally, it is unfortunate that the N-SIMEX procedure cannot handle the Laplace measurement error, a typical example for ordinary smooth cases. Therefore, it is of interest, both theoretically and empirically, to consider the possibility of constructing similar SIMEX-type procedures for non-normal cases. Due to differences in error distributions, we can reasonably infer that a one-for-all SIMEX-type procedure does not exist, it must be an error distribution specific procedure. In the following, we will focus on the Laplace measurement error case.

As a matter of fact, statistical models with Laplace measurement errors are not rare at all in applications. For example, the global positioning system (GPS) collars are frequently used by ecologists to collect location data for animals moving across a landscape. However, the GPS data are subject to measurement error. In their empirical study, [13] observed that the type I error can be greatly reduced for testing the classification of observed locations into habitat types when the GPS measurement error is fitted by a Laplace distribution. [16] investigated an application of Laplace measurement error in the analysis of data from microarray experiments and reported that the Laplace distribution fits the microarray expression data much better than a normal distribution. In a study of the differential privacy in cryptography, [4] present an example in which the queried number of incidents in a database are masked by random numbers generated from a Laplace distribution. In addition to providing some insights in SIMEX, we expect that the proposed methodology will find its application in the real word problems such as the ones discussed in these papers.

3. SIMEX for Laplace measurement error

The answer to the question that why N-SIMEX works for the normal measurement errors is straightforward. We can look at the heuristic and a relatively rigorous argument supplied in [3], but the most convincing theoretical proof can be found in [20]. Lemmas 1 and 2 in [20] show that if U_1 and U_2 are i.i.d. standardized normal r.v.'s, then $E(U_1 + \mathbf{i}U_2)^n = 0$, where $\mathbf{i} = \sqrt{-1}$, for all $n = 1, 2, \ldots$, and $Ef(\mu + \sigma_u U_1 + \sqrt{\lambda}\sigma_u U_2) \to Ef(\mu + \sigma_u U_1 + \mathbf{i}\sigma_u U_2) = f(\mu)$, as $\lambda \to -1$, for any sufficiently smooth function f, respectively. From these two lemmas, we can easily see that a sufficient and necessary condition for the N-SIMEX procedure to work is that $E(U_i + \mathbf{i}U_{b,i})^n = 0$ must hold for all $n = 1, 2, \ldots$. Unfortunately, the following theorem shows that, under a very general condition on the distribution of the measurement errors, this is true only for the normal case.

Theorem 3.1. *Suppose U_1 and U_2 are i.i.d. symmetric absolute continuous r.v.'s having finite moment generating function. Then $E(U_1 + \mathbf{i}U_2)^n = 0$, for all $n = 1, 2, \ldots$, if and only if U_1 and U_2 have a normal distribution with mean 0.*

The sufficiency part is proved in Lemma 1 in [20]. So we only need to show the necessity. Under the assumption of the theorem, $E(U_1 + \mathbf{i}U_2)^n = 0$, for all $n = 1, 2, \ldots$, implies that the moment generating function, denoted by $m(t)$, and the characteristic function, denoted by $\psi(t)$ of U_1 or U_2 satisfy $m(t)\psi(t) = 1$. Then from Theorem 2 of [21], U_1 and U_2 must have a normal distribution with mean 0.

Because of its computational simplicity, we really do not want to give up the "addition" component in the new SIMEX procedure to be developed, but the above theorem clearly indicates that just adding a multiple of the same random error to the contaminated data does not work. Of course, since we know the distribution of the measurement error, so it is natural to generate the extra random data from the same distribution or at least from the same distribution family. As mentioned in the previous section, the SIMEX-type procedure for non-normal measurement error must be error distribution specific. We will focus on the Laplace measurement error case. Accordingly, let U be a Laplace $(0, \sigma_u^2)$ r.v. We shall try to find the anti-measurement error variable (a term from [20]), say V characterized by some distributional parameters under addition, of U, in the sense that $E(U + V)^n$ will converge to 0 for all n, as the distributional parameters approach some particular values.

As seen from the brief justification of Theorem 3.1, $E(U_1 + \mathbf{i}U_2)^n = 0$, for all $n = 1, 2, \ldots$, if and only if the characteristic function and the moment generating function of U_1 has a reciprocal relationship. Similarly, in order to have $E(U + V)^n \to 0$, for all $n = 1, 2, \ldots$, it is sufficient to find V such that the characteristic function of $(U + V)$ converging to 1 as some distributional parameter of V tends to a certain particular value. Note that U, as a Laplace r.v., is the difference of two exponential or gamma r.v.'s, with the shape parameter 1 and scale parameter $\sigma_u/\sqrt{2}$. This motivates us to choose V, independent of U, to be the difference of two i.i.d. gamma r.v.'s with the shape parameter p and the scale parameter $\sigma_u/\sqrt{2}$. With such a choice, it is easy to see that the characteristic function $\phi_{U+V}(t)$ of $U + V$ is

$$\phi_{U+V}(t) = \left(\frac{1}{1 + \sigma_u^2 t^2/2}\right) \cdot \left(\frac{1}{1 + \sigma_u^2 t^2/2}\right)^p,$$

which will tend to 1, if we extrapolate the shape parameter p to -1. In fact, we have the following

Theorem 3.2. *Suppose $U \sim$ Laplace$(0, \sigma_u^2)$, and $V = V_1 - V_2$ with V_1, V_2 i.i.d. $\sim$ Gamma$(p, \sigma_u/\sqrt{2})$. Then $\lim_{p \to -1} E(U + V)^n = 0$, for all $n = 1, 2, \ldots$.*

Applying this theorem and an argument similar to the one used in the proof of Lemma 2 of [20], we obtain

Theorem 3.3. *Suppose f is sufficiently smooth, $U \sim$ Laplace$(0, \sigma_u^2)$, and $V = V_1 - V_2$ with V_1, V_2 i.i.d. $\sim$ Gamma$(p, \sigma_u/\sqrt{2})$. Then $\lim_{p \to -1} Ef(\mu + U + V) = f(\mu)$.*

The requirement of f being sufficiently smooth is indeed identical to the one used in Lemma 2 of [20]. It basically requires that f has a convergent Taylor expansion and the order of expectation and the summation can be interchanged in the Taylor expansion. Similar to [20], we can also describe the result in Theorem 3.3 in an amusing, yet descriptive way as "Laplace measurement error-induced bias is asymptotically annihilated by the addition of gamma anti-measurement error".

Based on the above results, now we propose the Laplace-error-based SIMEX procedure (L-SIMEX). Let Y, W, X still denote the response variable, the predictor without measurement error, and the latent variable, respectively. Suppose we observe $Z = X + U$, where $U \sim \text{Laplace}(0, \sigma_u^2)$, and as usual σ_u^2 is assumed to be known or can be reasonably estimated. Suppose θ is the model parameter to be estimated, and we start with an existing consistent estimation procedure $\hat{\theta}_{\text{TRUE}} = T(\{Y_i, W_i, X_i\}_{i=1}^n)$.

Fix a $p \geq 0$. In the simulation step, we first generate two independent sequences $\{V_{1,b,i}\}_{i=1}^n$ and $\{V_{2,b,i}\}_{i=1}^n$ of i.i.d. $\text{Gamma}(p, \sigma_u/\sqrt{2})$ r.v.'s, which are chosen to be independent of all other r.v.'s in the model. Then define

$$Z_{b,i}(p) = Z_i + V_{1,b,i} - V_{2,b,i} = X_i + U_i + V_{1,b,i} - V_{2,b,i}.$$

As before, b is a positive integer-valued index number. Let

$$\hat{\theta}_b(p) = T(\{Y_i, W_i, Z_{b,i}(p)\}_{i=1}^n),$$

and calculate

$$\hat{\theta}(p) = E[\hat{\theta}_b(p) \Big| \{Y_i, W_i, Z_i\}_{i=1}^n].$$

Again the expectation is with respect to the distribution of $\{V_{1,b,i}, V_{2,b,i}\}_{i=1}^n$ only. Similar to the N-SIMEX procedure, the above expectation might not have an explicit form, but it can be approximated arbitrarily well by the average of $\hat{\theta}_b(p)$, $b = 1, 2, \ldots, B$, based on B random samples $\{V_{1,b,i}, V_{2,b,i}\}_{i=1}^n$, $b = 1, 2, \ldots, B$. Repeat the above computations for a sequence of p values $0 = p_1 < p_2 < \cdots < p_{k-1} < p_k$ for some positive integer k. The convention we make here is that $p = 0$ corresponding to all observations being 0.

In the extrapolation step, unless the true relationship between p and $\hat{\theta}(p)$ can be identified, a trend of $\hat{\theta}(p)$ with respect to p should be approximately determined by checking the scatter plot of $(p_j, \hat{\theta}(p_j))$, $j = 1, 2, \ldots, k$. Again, least squares procedure might be needed to decide a reasonable analytic form of the extrapolation function. Once the trend is set, then extrapolating the final trend back to $p = -1$ gives the L-SIMEX estimate.

[3] provide a rigorous justification for their N-SIMEX procedure. We believe a similar argument will be also very beneficial to the understanding of the proposed L-SIMEX procedure for Laplace measurement errors. For this purpose, let $F_{Y,W,X}$ denote the distribution function (d.f.) of (Y, W, X) and $F_{Y,W,Z_b(p)}$ that of $(Y, W, Z_b(p))$. Suppose the true parameter $\theta_0 = S(F_{Y,W,X})$, where S is a continuous functional on the class of d.f.'s.

Let G_1, G_2 be two independent $\text{Gamma}(p+1, \sigma_u/\sqrt{2})$ r.v.'s and H denote the d.f. of the difference $G_1 - G_2$. From [9], for a r.v. $\Delta \sim H$,

$$E(\Delta^{2k}) = \frac{\sigma_u^{2k}(2k-1)!}{2^{k-1}(k-1)!}(p+k)(p+k-1)\cdots(p+1).$$

Direct calculations show that for some known constants $b_{j,k}$, $j = 1, \dots, k$, with $b_{k,k} = 1$,

$$\sigma_u^{2k}(p+k)(p+k-1)\cdots(p+1) = \sum_{j=1}^{k} b_{j,k}\big((p+1)\sigma_u^2\big)^j (\sigma_u^2)^{k-j}.$$

Note that H is a symmetric distribution around zero, hence all odd moments of H are zero. Moreover, H is completely determined by its moments. Since both k and σ_u^2 are known, the d.f. H thus depends only on the parameter $(p+1)\sigma_u^2$. To emphasize this dependence we now write $H_{(p+1)\sigma_u^2}$ for H, with the convention that $H_0 = \delta_0$, a distribution degenerate at zero. Note that $H_{(p+1)\sigma_u^2}$ converges weakly to δ_0, as $p \to -1$.

If we further assume that $F_{Y,W,Z_b(p)}$ is totally determined by its moments, then from the above observations, and by a direct calculation, we obtain

$$F_{Y,W,Z_b(p)} = F_{Y,W,X} * \begin{pmatrix} \delta_0 \\ \delta_0 \\ H_{(p+1)\sigma_u^2} \end{pmatrix},$$

$$S(F_{Y,W,Z_b(p)}) = S\left(F_{Y,W,X} * \begin{pmatrix} \delta_0 \\ \delta_0 \\ H_{(1+p)\sigma_u^2} \end{pmatrix}\right).$$

A similar phenomenon happens to hold in the normal case, where $S(F_{Y,W,Z_b(\lambda)})$ depends on λ only through $(1+\lambda)\sigma_u^2$. If we further assume that $\hat{\theta}_{\text{TRUE}}$, $\hat{\theta}_b(p)$ and $\hat{\theta}(p)$ converge in probability to their expectations, which are assumed to be finite, and $\hat{\theta}_{\text{TRUE}} \to_p \theta_0 = S(F_{Y,W,X})$, then we also have $\hat{\theta}(p) \to_p \theta(p) = S(F_{Y,W,Z_b(p)})$. Finally, the continuity of S implies

$$\theta(p) \to_p S\left(F_{Y,W,X} * \begin{pmatrix} \delta_0 \\ \delta_0 \\ \delta_0 \end{pmatrix}\right) = S(F_{Y,W,X}) = \theta_0, \quad \text{as } p \to -1.$$

Similar to the N-SIMEX procedure, the vexing part in the proposed L-SIMEX procedure is the extrapolation step. As we mentioned earlier, the true extrapolant usually does not have an explicit form which makes the extrapolation step not easy to follow. Although some exploratory data analysis such as scatter plots and least squares estimation can help us to determine an approximate extrapolant, the accuracy of the approximation is heavily influenced by the Monte Carlo error in the simulation step. Increasing the sample size and the replication time B of course can reduce the Monte Carlo error, but at the cost of prolonged computation time. Empirical and theoretical arguments show that we can actually choose the extrapolant from the linear, quadratic, and nonlinear forms as in the N-SIMEX case. In fact, exactly the same argument as in [3] indicates that the quadratic and the nonlinear extrapolants produce L-SIMEX estimates with asymptotic bias of the order $O(\sigma_u^6)$. It is also true that using NON-IID pseudo errors improves the convergence of $B^{-1}\sum_{b=1}^{B}\hat{\theta}_b(p)$ to $\hat{\theta}(p)$ as $B \to \infty$.

In the N-SIMEX procedure, one uses the same set of random numbers $\{U_{b,i}\}$ generated from $N(0, \sigma_u^2)$ for all the grid points λ_j's. As a result, the scatter plot of $\hat{\theta}(\lambda_j)$'s versus λ_j's will have a smooth pattern and it might be very helpful for deciding the form of the extrapolant. However, this is not the case for the L-SIMEX procedure. Because for each grid point p, the random numbers $\{U_{b,i}\}$ are generated from the gamma distribution with shape parameter p. Therefore, the trend shown on the scatter plot of $\hat{\theta}(p_j)$'s versus p_j's will not be very smooth. Consequently, the L-SIMEX estimates would not be very stable. One possible way to avoid this is to use more grid points to decide the extrapolant, for example, use three or four grid points to determine the linear form instead of just two points as in the N-SIMEX case.

4. Variance estimation for L-SIMEX

By establishing a strong relationship between the N-SIMEX estimation and the jackknife estimation, [20] lay a solid theoretical justification for their N-SIMEX procedure for normal measurement error. More importantly, inspired by the jackknife variance estimation, a variance estimation method for N-SIMEX is constructed. In this section, we extend their methodology to the L-SIMEX procedure.

We will start with single observation case, which is parallel to the discussion in Section 4.3 of [20]. Suppose we want to estimate $\theta = \exp(\mu)$ from the location model $Z = \mu + U$, where $U \sim$ Laplace $(0, \sigma_u^2)$ and σ_u^2 is known. Since only one observation is available, so $\exp(Z)$ is the MLE of θ. Now let $Z_b = Z + V_b$, where $V_b = V_{1b} - V_{2b}$ and V_{1b}, V_{2b} are independent and both have a Gamma distribution with shape parameter p and scale parameter $\sigma_u/\sqrt{2}$. Then in the L-SIMEX version of the jackknife with sample size 1, the building block for variance estimation is $\Delta(p) = \hat{\theta}_b(p) - \hat{\theta}(p)$, where $\hat{\theta}_b(p) = \exp(Z_b)$, and $\hat{\theta}(p) = E(\hat{\theta}_b(p)|Z)$. Direct calculations yield

$$\Delta(p) = \exp(Z)\left[\exp(V_b) - \left(1 - \frac{\sigma_u^2}{2}\right)^{-p}\right],$$

with $E\Delta(p) = 0$, and

$$\mathrm{Var}(\Delta(p)) = \frac{\exp(2\mu)}{1 - 2\sigma_u^2}\left[\left(\frac{1}{1 - 2\sigma_u^2}\right)^{p} - \left(1 - \frac{\sigma_u^2}{2}\right)^{-2p}\right].$$

Note that $\hat{\theta}_{\text{SIMEX}} = \hat{\theta}(-1) = \exp(Z)[1 - \sigma_u^2/2]$, and that

$$\mathrm{Var}(\hat{\theta}_{\text{SIMEX}}) = \exp(2\mu)\left[\left(1 - \frac{\sigma_u^2}{2}\right)^{2}\left(\frac{1}{1 - 2\sigma_u^2}\right) - 1\right].$$

Therefore, for this particular example, we have

$$\mathrm{Var}(\hat{\theta}_{\text{SIMEX}}) = -\lim_{p \to -1} \mathrm{Var}(\Delta(p)).$$

This relationship also holds when $\theta = f(\mu) = \mu^2$. In general, for the location model and for a function $\theta = f(\mu)$ of μ, we have the following result, which is an analogue of Lemma 3 in [20]. In this case, $\hat{\theta}_b(p) = f(Z_b)$, $\hat{\theta}(p) = E(\hat{\theta}_b(p)|Z)$, and $\Delta(p) = \hat{\theta}_b(p) - \hat{\theta}(p)$.

Theorem 4.1. *Suppose f is sufficiently smooth, then*

$$Var\left[\lim_{p\to -1} \hat{\theta}(p)\right] = -\lim_{p\to -1} Var(\Delta(p)).$$

After replacing λ with p, the proof of Theorem 4.1 is almost a repetition of the proof of Lemma 3 in [20] with the help of Theorem 3.2 in Section 3 above, hence omitted for the sake of brevity.

Clearly, the above result has limited practical significance other than its theoretical importance. In the following we shall propose a variance estimation procedure for the general case. Since the theory of the multivariate Laplace distribution is still developing, we will only focus on the one dimensional case, in other word, the latent variable X and the measurement error variable U are assumed to be one dimensional.

To proceed further, we assume that in the case $X_i, i = 1, \dots, n$, are observable, there exist unbiased estimators $\hat{\theta}_{\text{TRUE}} = T(\{Y_i, W_i, X_i\})$ of θ, and $T_{\text{Var}}(\{Y_i, W_i, X_i\})$ of $\text{Var}(\hat{\theta}_{\text{TRUE}})$, respectively. Then an estimate of $\text{Var}(\hat{\theta}_{\text{SIMEX}})$ can be constructed via the following steps:

(1). Calculate $T_{\text{Var}}(\{Y_i, W_i, Z_{b,i}(p)\})$ for $b = 1, 2, \dots, B$, and denote the average as $\hat{\tau}^2_B(p)$. If possible, calculate the limit of the average as $B \to \infty$, and denote the limit as $\hat{\tau}^2(p)$;
(2). Calculate the sample variance of $\hat{\theta}_b(p)$, $b = 1, 2, \dots, B$. Denote it as $s^2_\Delta(p)$;
(3). Extrapolating $\hat{\tau}^2_B(p) - s^2_\Delta(p)$ or $\hat{\tau}^2(p) - s^2_\Delta(p)$ to $p = -1$ to get an estimator of the variance of $\hat{\theta}_{\text{SIMEX}}$.

If the exact extrapolant is used, then the resulting estimator is unbiased.

We conclude this section by the similar component-of-variance problem discussed as in [20] to provide a tangibly informative demonstration of the above ideas. To be specific, the component-of-variance model has the form of $Z = X + U$, where the latent variable $X \sim N(0, \sigma_x^2)$, and the measurement error $U \sim \text{Laplace}(0, \sigma_u^2)$, with σ_u^2 a known positive constant. The parameter of interest is $\theta = \sigma_x^2$. We start with the true estimate $\hat{\theta}_{\text{TRUE}} = S_X^2$, the sample variance of $\{X_i\}_{i=1}^n$. The naive estimate is simply $\hat{\theta}_{\text{NAIVE}} = S_Z^2$. We have shown in Section 2 that $\hat{\theta}_{\text{SIMEX}} = S_Z^2 - \sigma_u^2$, and after some tedious computation, we obtain

$$\hat{\tau}^2(-1) = \frac{2}{n+1}\left[(S_Z^2 - \sigma_u^2)^2 + \frac{(3-n)\sigma_u^4}{n(n-1)} - \frac{4S_Z^2\sigma_u^2}{n-1}\right],$$

and

$$s^2_\Delta(-1) = \frac{(3-n)\sigma_u^4}{n(n-1)} - \frac{4S_Z^2\sigma_u^2}{n-1}.$$

Therefore,

$$\hat{\tau}^2(-1) - s_\Delta^2(-1) = \frac{2S_Z^4}{n+1} + \frac{3(n-1)\sigma_u^4}{n(n+1)}.$$

This, together with the result

$$ES_Z^4 = \frac{3\sigma_x^4 + 6\sigma_x^2\sigma_u^2 + 6\sigma_u^4}{n} + \frac{(n^2-2n+3)(\sigma_x^2+\sigma_u^2)^2}{n(n+1)},$$

implies

$$E\left[\hat{\tau}^2(-1) - s_\Delta^2(-1)\right] = \frac{2\sigma_x^4}{n-1} + \frac{4\sigma_x^2\sigma_u^2}{n-1} + \frac{(5n-3)\sigma_u^4}{n(n-1)},$$

which is exactly the variance of $\hat{\theta}_{\text{SIMEX}}$.

5. Numerical studies

Parallel to the Section 4 in [3], this section will report two simulation studies conducted for some well known parametric models and a real data example. Through these numberical studies, not only can one further get more familiar with the methodology, but also appreciate the wide applicability of the L-SIMEX procedure when the measurement error is truly Laplace.

Simulation 1: Polynomial regression models

Consider the quadratic measurement error regression model $Y = \beta_0 + \beta_1 X + \beta_2 X^2 + \varepsilon$ and $Z = X + U$. We would like to point out that the best extrapolant function is not of the form $a + b/(c+p)$ anymore. In fact, this is also true for the N-SIMEX procedure. For the sake of simplicity, we assume that X has expectation 0.

If $U_i \sim N(0, \sigma_u^2), U_{b,i} \sim N(0, \sigma_u^2)$, all mutually independent, and $Z_{b,i} = X_i + U_i + \sqrt{\lambda}U_{b,i}$, then the estimate of β_1 for extrapolating has the form

$$\begin{pmatrix}\hat{\beta}_{0,b}(\lambda)\\ \hat{\beta}_{1,b}(\lambda)\\ \hat{\beta}_{2,b}(\lambda)\end{pmatrix} = \begin{pmatrix} n & \sum_{i=1}^n Z_{b,i} & \sum_{i=1}^n Z_{b,i}^2 \\ \sum_{i=1}^n Z_{b,i} & \sum_{i=1}^n Z_{b,i}^2 & \sum_{i=1}^n Z_{b,i}^3 \\ \sum_{i=1}^n Z_{b,i}^2 & \sum_{i=1}^n Z_{b,i}^3 & \sum_{i=1}^n Z_{b,i}^4 \end{pmatrix}^{-1} \begin{pmatrix} \sum_{i=1}^n Y_i \\ \sum_{i=1}^n Z_{b,i}Y_i \\ \sum_{i=1}^n Z_{b,i}^2 Y_i \end{pmatrix}$$

It is not easy, if not impossible, to compute $\hat{\beta}_j(\lambda) = E(\hat{\beta}_{j,b}(\lambda)|\{Y_i, Z_i\}_{i=1}^n)$ for $j = 0, 1, 2$, hence the exact extrapolating function. But we can have a good guess by checking the limit of $\hat{\beta}_{j,b}(\lambda)$. Let $\mu_k = EX^k$ for $k = 2, 3, 4$, and $\tilde{\lambda} = (1+\lambda)\sigma_u^2$. Then, one can show that

$$\hat{\beta}_{1,b}(\lambda) \to P \frac{2\tilde{\lambda}\beta_2\mu_3(2\mu_2 + \tilde{\lambda}) + \beta_1[\mu_2(\mu_4 - 3\mu_2^2 + 2(\mu_2+\tilde{\lambda})^2) - \mu_3^2]}{(\mu_4 - 3\mu_2^2)(\mu_2 + \tilde{\lambda}) + 2(\mu_2+\tilde{\lambda})^3 - \mu_3^2},$$

$$\hat{\beta}_{2,b}(\lambda) \to P \frac{\tilde{\lambda}\beta_1\mu_3 + \beta_2[(\mu_2+\tilde{\lambda})(\mu_4 - \mu_2^2) - \mu_3^2]}{(\mu_4 - 3\mu_2^2)(\mu_2 + \tilde{\lambda}) + 2(\mu_2+\tilde{\lambda})^3 - \mu_3^2}.$$

Therefore, the reasonable extrapolant should be of the form

$$a_0 + \frac{a_1 + a_2\lambda + a_3\lambda^2}{1 + a_4\lambda + a_5\lambda^2 + a_6\lambda^3},$$

which is rather complicated. The same form of extrapolant function can be obtained for the case of Laplace measurement errors.

For the sake of simplicity, in the following simulation, we will adopt a simpler quadratic model in which $\beta_1 = 0$. Then we can show that in the normal measurement error case,

$$\hat{\beta}_{2,b}(\lambda) \to P \frac{\beta_2(\mu_4 - \mu_2^2)}{\mu_4 - \mu_2^2 + 4\mu_2\sigma_u^2(1+\lambda) + 2(1+\lambda)^2\sigma_u^4}.$$

If $U_i \sim$ Laplace $(0, \sigma_u^2)$, and $U_{b,i} \sim V_{1i} - V_{2i}$ with $V_{1i}, V_{2i} \sim$ Gamma $(p, \sigma_u/\sqrt{2})$, $Z_{b,i} = X_i + U_i + U_{b,i}$, then we can show that

$$\hat{\beta}_{2,b}(p) \to P \frac{\beta_2(\mu_4 - \mu_2^2)}{\mu_4 - \mu_2^2 + (4\mu_2\sigma_u^2 + 3\sigma_u^4)(1+p) + 2(1+p)^2\sigma_u^4}.$$

Therefore, a proper extrapolant for the quadratic regression model has a simpler form, $a + b/(c + d\lambda + \lambda^2)$ in the normal case and $a + b/(c + dp + p^2)$ in the Laplace case. Note that the square terms are related to σ_u^4, so if σ_u^2 is small, then using $a + b/(c + \lambda)$ or $a + b/(c + p)$ to extrapolate may still get satisfying answers, but if σ_u^2 is large, then this replacement might be very dangerous.

In the simulation, we chose $\beta_0 = \beta_2 = 1$, the sample size to be 300. To see the effect of σ_u^2 on the SIMEX procedures, we chose $\sigma_u^2 = 0.3^2$ and $\sigma_u^2 = 0.6^2$. The latent variable $X \sim N(0, 1)$. Thus the reliability ratio is 91.7% when $\sigma_u^2 = 0.3^2$, and 73.5% when $\sigma_u^2 = 0.6^2$. Naive, linear, quadratic, two nonlinear SIMEX estimates are compared. For convenience, we call the nonlinear extrapolant $a + b/(c+\lambda)$ or $a + b/(c+p)$ as Type I nonlinear extrapolant, and $a + b/(1 + c\lambda + d\lambda^2)$ or $a + b/(1 + cp + dp^2)$ as Type II nonlinear extrapolant. Similar to the linear case, we also simulate the mismatch cases, that is, N-SIMEX procedure is applied to the Laplace measurement error and L-SIMEX procedure is applied to the normal measurement error.

Figures 1 to 6 consist of the kernel density estimates (KDEs) of the distributions of β_2 estimates corresponding to the standard normal kernel, where the bandwidth used is 0.3 times the median of the standard deviations of all the estimators in the study. In all plots, the solid curve is for the naive estimate, the short dashed curve is for the linear L-SIMEX estimate, the dotted curve is for the quadratic L-SIMEX estimate, the long dashed curve is for the Type I nonlinear L-SIMEX estimate, and the dash-dotted curve is for the Type II nonlinear L-SIMEX estimate. Table 1 and 2 are means and MSEs for all the estimates. Just for exploratory purpose, we also calculate the average of the quadratic and the two nonlinear estimates, denoted by Combined in Table 1 and 2. In these simulations, the L-G SIMEX outperforms the other two mismatch cases.

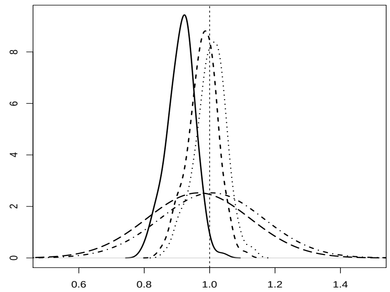

FIG 1. *KDEs of the distributions of β_2 estimates in the quadratic regression (L-G), $\sigma = 0.3$.*

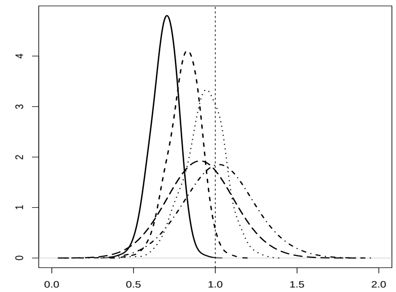

FIG 2. *KDEs of the distributions of β_2 estimates in the quadratic regression (L-G), $\sigma = 0.6$.*

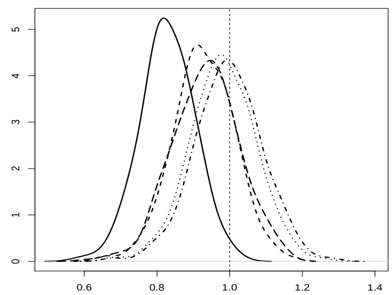

FIG 3. *KDEs of the distributions of β_2 estimates in the quadratic regression (L-N), $\sigma = 0.3$.*

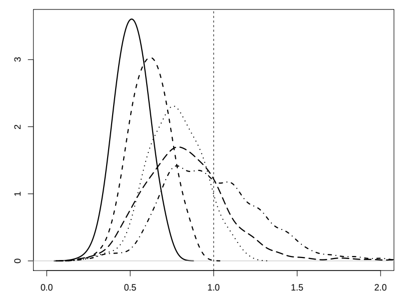

FIG 4. *KDEs of the distributions of β_2 estimates in the quadratic regression (L-N), $\sigma = 0.6$.*

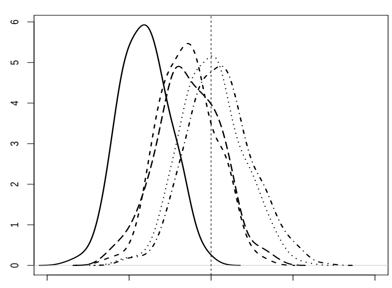

FIG 5. *KDEs of the distributions of β_2 estimates in the quadratic regression (N-L), $\sigma = 0.3$.*

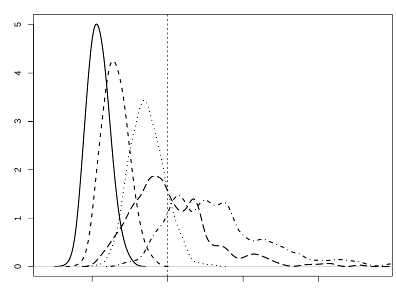

FIG 6. *KDEs of the distributions of β_2 estimates in the quadratic regression (N-L), $\sigma = 0.6$.*

It might be worth mentioning that if the measurement error is normal, but the L-SIMEX procedure is used, then as $n \to \infty$ and $p \to -1$, $\hat{\beta}_{2,b}(p) \to P\beta_2(\mu_4 - \mu_2^2)/(\mu_4 - \mu_2^2 - 3\sigma_u^4)$, which over estimates β_2, and if the measurement error is Laplace, but the N-SIMEX procedure is used, then, as $n \to \infty$ and $p \to -1$, $\hat{\beta}_{2,b}(p) \to P\beta_2(\mu_4 - \mu_2^2)/(\mu_4 - \mu_2^2 + 3\sigma_u^4)$, which under estimates β_2.

TABLE 1
Mean and MSE of β_2 estimates in quadratic regression, $\sigma = 0.3$

	$\hat{\beta}_{1,\text{NAIVE}}$	Linear	Quadratic	Nonlinear (I)	Nonlinear (II)	Combined
L-G Mean	0.9119	0.9783	1.0022	1.0068	0.9657	0.9915
MSE	0.0018	0.0021	0.0024	0.0025	0.0024	0.0023
L-N Mean	0.8289	0.9285	0.9806	0.9965	0.9338	0.9703
MSE	0.0054	0.0070	0.0081	0.0086	0.0084	0.0081
N-L Mean	0.8374	0.9420	1.0013	1.0212	0.9456	0.9893
MSE	0.0037	0.0047	0.0057	0.0065	0.0060	0.0057

TABLE 2
Mean and MSE of β_2 Estimates in Quadratic Regression, $\sigma = 0.6$

	$\hat{\beta}_{1,\text{NAIVE}}$	Linear	Quadratic	Nonlinear (I)	Nonlinear (II)	Combined
L-G Mean	0.6907	0.8141	0.9352	1.0241	0.9104	0.9566
MSE	0.0059	0.0087	0.0147	0.0265	0.0221	0.0195
L-N Mean	0.5095	0.6182	0.7746	1.0038	0.8331	0.8705
MSE	0.0091	0.0136	0.0268	0.0993	0.0667	0.0560
N-L Mean	0.5407	0.6548	0.8654	1.3888	1.0244	1.0929
MSE	0.0049	0.0072	0.0132	0.4616	0.0797	0.0936

Simulation 2: Logistic regression

In this simulation, we adopt the similar logistic regression model as in [3] except the measurement error now has a Laplace distribution. To be specific, the response variable Y and covariates X, W are related through

$$P(Y = 1|X, W) = F(\beta_1 + \beta_x X + \beta_w W + \beta_{xw} XW),$$

where F is the logistic d.f. Here, W is observable, X is measured with Laplace error $Z = X + U$, $U \sim$ Laplace $(0, 0.5^2)$, and X, W are bivariate normal with standard normal as their marginal distributions, and their correlation coefficient is set to be $1/\sqrt{5}$. Thus the reliability ratio is 80%. The regression parameters are set to $\beta_1 = -2$, $\beta_x = 1$, $\beta_w = 0.25$, and $\beta_{xw} = 0.25$. Simulation is conducted for 400 data sets of sample size $n = 1500$ for each data set. To visually illustrate the L-SIMEX procedure, we take $p \in \{0, 1/8, 2/8, \ldots, 16/8\}$ as the grid points and $B = 200$. Figure 7 shows L-SIMEX steps from three different extrapolants, linear, quadratic and nonlinear based on the data from all grid points. For this particular data set and p values, the performance of the L-SIMEX estimators for $\beta_x, \beta_w, \beta_{xw}$ are well ordered in the sense that the nonlinear extrapolation estimator generally has smaller estimation bias than the quadratic extrapolating estimator, the quadratic extrapolating estimator has smaller estimation bias than the linear extrapolating estimator, and the naive estimator has the largest estimation error. The exception occurs for the estimate of β_1 with quadratic L-SIMEX performing the best, followed by linear, nonlinear and naive. We also tried some other data sets, and same phenomenon holds for most cases.

We also conducted the simulation study using the above set up, but where the three $\sigma_u = 0.3, 0.6, 1$ values were used in the simulation in order to see the effect

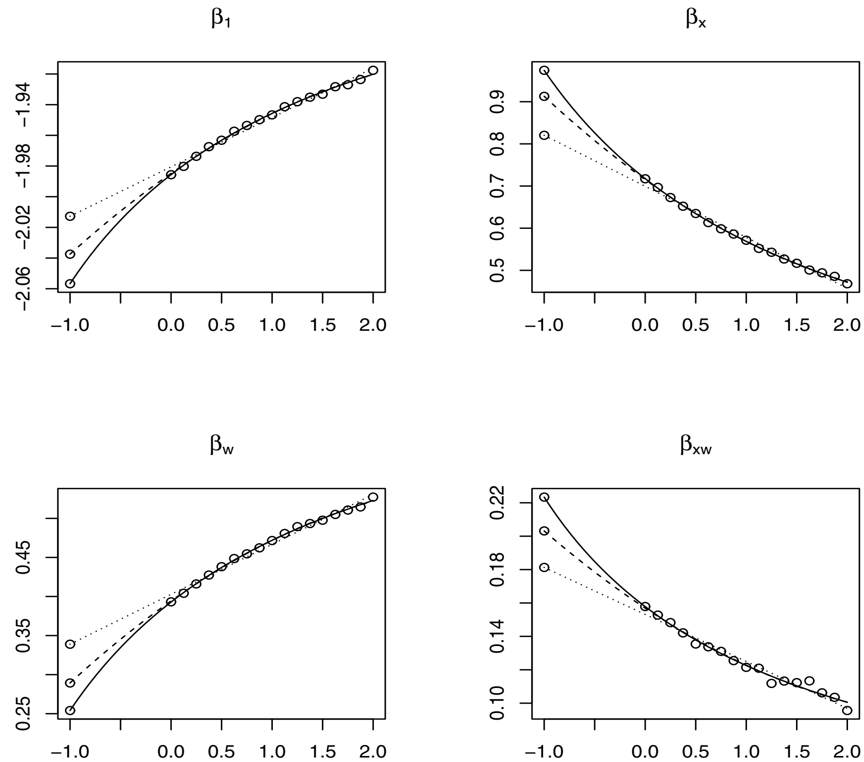

FIG 7. *Extrapolation Plot for logistic regression.*

TABLE 3
Mean (MSE) of the estimates in logistic regression, $\sigma = 1$

		Naive	Linear	Quadratic	Nonlinear
β_1	L-L	-1.8169 (0.0066)	-1.8388 (0.0069)	-1.8852 (0.0079)	-2.0287 (0.0191)
	N-L	-1.8213 (0.0064)	-1.8432 (0.0068)	-1.8882 (0.0079)	-2.0155 (0.0169)
	L-N	-1.8236 (0.0069)	-1.8454 (0.0072)	-1.8922 (0.0082)	-2.0444 (0.0213)
β_x	L-L	0.4072 (0.0035)	0.4804 (0.0049)	0.6293 (0.0089)	1.0209 (0.0383)
	N-L	0.4044 (0.0041)	0.4773 (0.0056)	0.6235 (0.0099)	0.9960 (0.0357)
	L-N	0.4064 (0.0037)	0.4792 (0.0052)	0.6282 (0.0092)	1.0296 (0.0410)
β_w	L-L	0.5352 (0.0073)	0.4973 (0.0079)	0.4198 (0.0096)	0.2074 (0.0228)
	N-L	0.5471 (0.0081)	0.5097 (0.0086)	0.4363 (0.0098)	0.2565 (0.0180)
	L-N	0.5465 (0.0079)	0.5094 (0.0084)	0.4318 (0.0100)	0.2060 (0.0251)
β_{xw}	L-L	0.1092 (0.0033)	0.1293 (0.0046)	0.1741 (0.0079)	0.2992 (0.8687)
	N-L	0.1035 (0.0035)	0.1232 (0.0049)	0.1604 (0.0081)	0.2926 (2.3963)
	L-N	0.1079 (0.0036)	0.1270 (0.0051)	0.1733 (0.0089)	0.2779 (2.0053)

of measurement error on the L-SIMEX procedure. An interesting finding is that when σ_u is small (0.3 and 0.6), both match and mismatch SIMEX procedures work almost equally well, but if σ_u is large ($\sigma_u = 1$), the advantage of the match SIMEX procedure becomes obvious, in particular, when estimating the parameter β_{xw}. Also see Table 3 for detail.

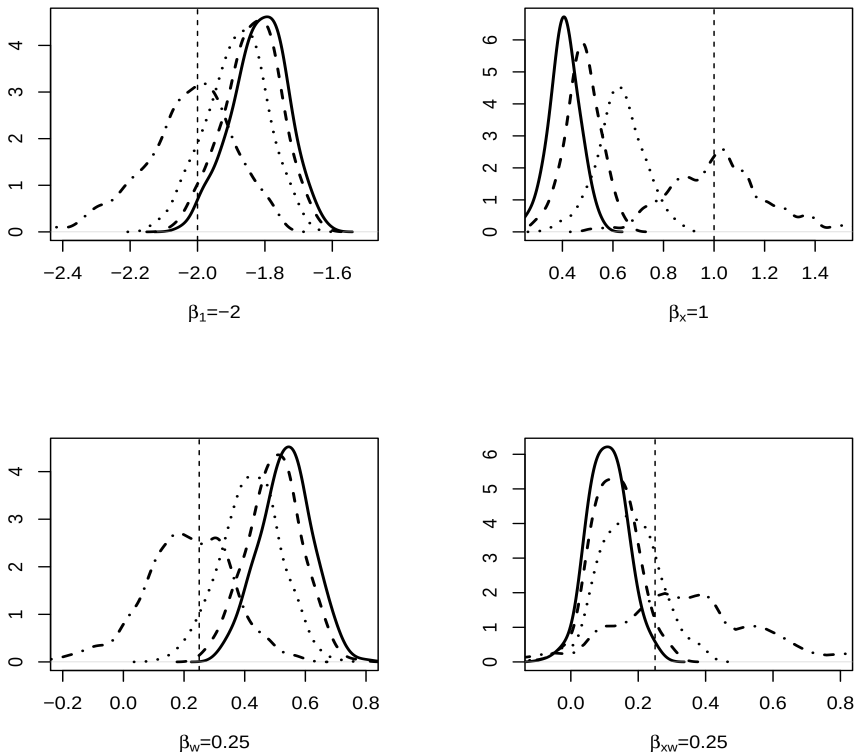

FIG 8. *KDE plots for logistic regression, L-G, $\sigma = 1$.*

For illustration, Figure 8 is the KDE plots for the parameter estimates using the proposed L-SIMEX procedure. Again, normal kernel function is used for all cases and the common bandwidth is chosen to be 0.3 times the median of the standard deviations of all the estimators in the study. The solid curve is for the naive estimate, the dashed curve is for the linear L-SIMEX, the dotted curve is for the quadratic L-SIMEX and the dash-dotted curve is for the nonlinear L-SIMEX. As expected, the naive estimate, linear, quadratic SIMEX become more conservative with larger σ, the nonlinear L-SIMEX estimate stays unbiased but with increasing variability. For the sake of brevity, the KDE plots for mismatch cases are omitted, since the overall patterns are similar.

Real data example: Framinham heart study

In the following, we shall apply the proposed L-SIMEX procedure to a data set from the Framinham Heart Study, which consists of several exams taken two years apart. The data set includes 1615 observations from men aged from 31 to 65 years. The response variable Y is the indicator of the first evidence of coronary heart disease (CHD) within an 8-year period following the second exam. There are 128 cases of CHD. The predictors of interests include the age

Table 4
SIMEX estimates and S.E.'s from the Framinham data, SBP measured with error

	Intercept	Age	Smoking	Log(SCL-100)	Log(SBP-50)
N-SIMEX, Linear	-20.3417	0.0561	0.6044	1.5927	1.5822
	2.3997	0.0118	0.2510	0.3389	0.3896
L-SIMEX, Linear	-20.3381	0.0561	0.6046	1.5927	1.5813
	2.3909	0.0118	0.2510	0.3389	0.3862
N-SIMEX, Quadratic	-21.5532	0.0538	0.6138	1.5784	1.8972
	2.6002	0.0120	0.2517	0.3396	0.4621
L-SIMEX, Quadratic	-21.5430	0.0539	0.6139	1.5786	1.8944
	2.6010	0.0120	0.2517	0.3396	0.4626
N-SIMEX, Nonlinear	-22.6926	0.0511	0.6246	1.5662	2.1963
	3.0152	0.0127	0.2548	0.3416	0.5986
L-SIMEX, Nonlinear	-22.6832	0.0511	0.6237	1.5666	2.1937
	3.1098	0.0130	0.2548	0.3415	0.6280

at Exam 2, systolic blood pressures (SBP) at Exam 2 and Exam 3, smoking status, and serum cholesterol levels (SCL) at Exam 2 and Exam 3. For each individual, SBP are measured twice by independent examiners at each exam. The covariates age and smoking status do not have measurement errors, but the measurement error is presents in SBP and SCL. A logistic regression model is used to investigate the relationship between CHD and covariates. In our analysis, the SBP measurement is transformed to $\log(SBP-50)$. But different from [2], the SCL measurement is transformed to $\log(SCL-100)$. A logistic regression model is used to investigate the relationship between CHD and covariates. To be specific, if we denote $X_1 = \log(SBP-50)$, and $X_2 = \log(SCL-100)$, then the logistic regression model used in our analysis can be written as

$$P(Y=1|\text{Age}, \text{Smoking}, X_1, X_2) = F(\beta_1 + \beta_2\text{Age} + \beta_3\text{Smoking} + \beta_4 X_1 + \beta_5 X_2),$$

where F is the logistic distribution function.

Similar to the set up in [2], we ignore the measurement error in SCL and assume that the SBP is the only covariate measured with error. In our analysis, the average Z of all four $\log(SBP-50)$ readings from each individual will be used as the surrogate. Therefore, implicitly, the true latent variable X is the long-term average of Z. The average of $\log(SCL-100)$ from Exam 2 and 3 is used as the true value of $\log(SCL-100)$.

A component of variance analysis produces an estimate of 0.01285 for the variance of measurement error σ_u^2, which is slightly different from 0.01259, the one reported in [2]. In the simulation, $B = 2000$ and eight equally spaced values of p from $[0, 8]$ are used in the simulation step. Using the same B and p values, the variance estimation procedure developed in Section 5 is used to obtain the standard errors for each estimate. Linear, quadratic and Type I nonlinear extrapolant are used for the extrapolation. Table 4 presents the analysis results from both N-SIMEX procedure and L-SIMEX procedure. In each cell, the number on the top is the SIMEX estimate for the regression parameters, and the number on the bottom is the SIMEX estimate for the standard error. It is interesting to notice that the estimates from both SIMEX procedures behave almost

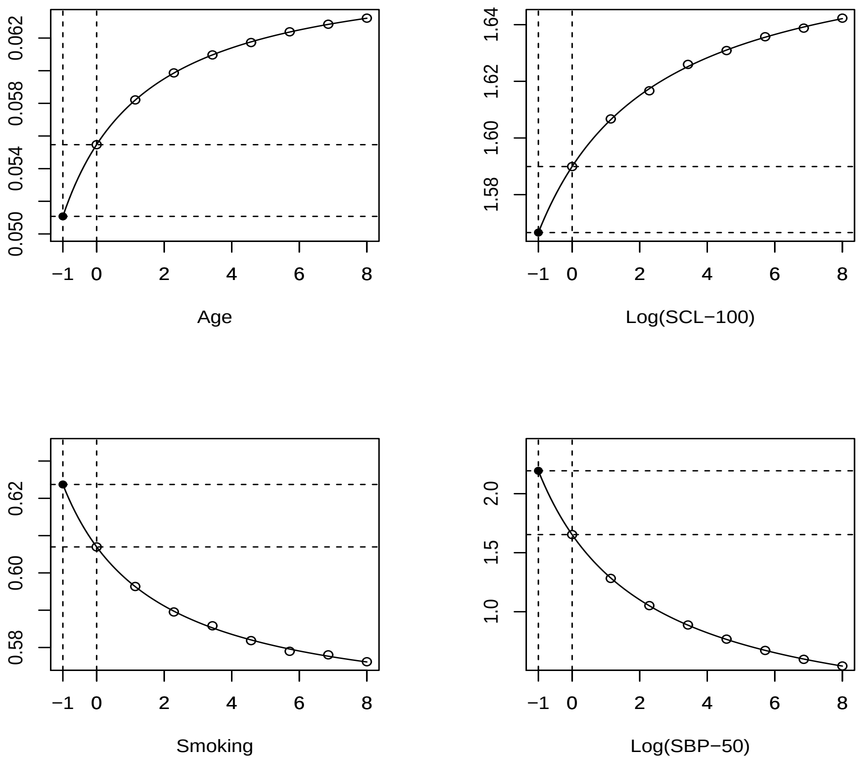

FIG 9. *Extrapolation plot for regression parameters estimates, SBP measured with error.*

the same. This is mainly due to the fact that the variance of the measurement error (0.01285) is very small and this phenomenon well matches the simulation results in logistic regression reported in Section 4.

Figures 9 and 10 are the L-SIMEX extrapolation plots for the regression parameters except for the intercept. The x-axis is the value of p, and the y-axis in Figure 9 is the estimated value of regression parameter, and the y-axis in Figure 10 is the estimated variance. From Figure 9 and 10, we can see that the effect of the measurement error is significant on the regression coefficient for $\log(SBP - 50)$. This finding is similar to that of [2]. Since the N-SIMEX extrapolation plots are very similar to the L-SIMEX extrapolation plots, they are omitted here for the sake of brevity.

6. Discussion

Here, we shall summarize the contribution of this paper and make a finite sample comparison of the L-SIMEX with the revised moment estimates of [10].

This paper has developed a novel L-SIMEX procedure in the measurement error models when the measurement error has a Laplace distribution. Theo-

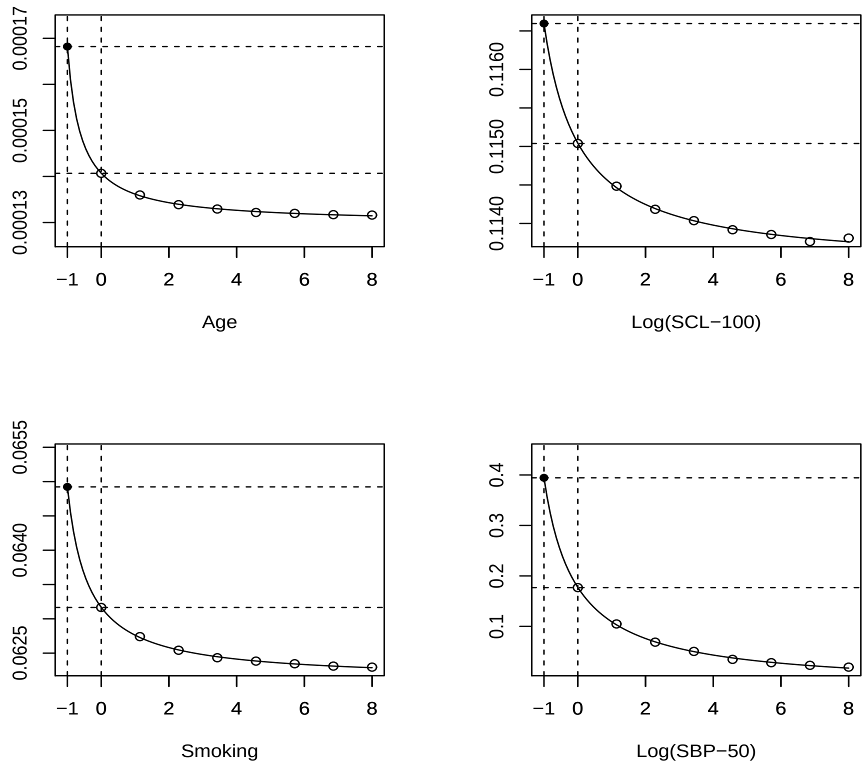

FIG 10. *Extrapolation plot for variance estimates, SBP measured with error.*

retical and empirical evidence clearly shows that the L-SIMEX procedure can successfully reduce or remove the bias induced by the Laplace error. Although the difference between N-SIMEX and L-SIMEX procedures is not very significant when the variance of measurement error is small, extra caution has to be taken if the parameter to be estimated involves higher than the fourth moment and the measurement error is not negligible. Same normal pseudo-errors can be repeatedly used in the simulation step for different λ values in N-SIMEX procedure, but in L-SIMEX procedure, the same gamma pseudo-variables can not be repeatedly used in the simulation step. Therefore, a bigger value of B is preferred in the application in order to reduce the Monte Carlo simulation bias.

Finding the exact extrapolant usually is not possible, in particular, when the model and the estimation procedure is complicated. The simulation study shows that if the measurement error is small, then the nonlinear form $a + b/(c + p)$ works well in the Laplace measurement error case. For example, in the quadratic regression model discussed in Section 4, although the true extrapolant has a form of $a + b/(c + dp + p^2)$, but the square term p^2 actually related to σ_u^4 which is negligible when σ_u^2 is small. Asymptotic theory for L-SIMEX procedure will be developed following the work in [2] at a later date.

For the cases of multiple predictors measured with Laplace error, if the components in the measurement error vector are independent with mean 0 and possibly different variances, we can still adapt and apply the proposed L-SIMEX, and the modification is obvious and straightforward. But if the measurement error vector follows a general Laplace distribution, such as the one defined in [11] or in [22], then extending the proposed L-SIMEX to multiple predictors cases is not easy. A natural choice would be searching for a version of multidimensional gamma distribution, such as the ones discussed in [8, 14, 15], to see if it still maintains the similar result as in Theorem 3.2. But up to now, we are still investigate this possibility.

Finally we would like to comment on [10]'s revised moment estimation procedure when data are contaminated with Laplace measurement error. To be specific, suppose the parameter of interest, say β, is defined by a set of population moment conditions $Em(X;\beta) = 0$, where m is twice differentiable. Instead of observing X, one observes its surrogate $Z = X + U$. Both the parameter β and the variable X could be multidimensional. If the components of U are independent and follow Laplace distributions with mean 0 and possibly different variances, then the population moment conditions can be replaced by the revised moment conditions proposed in [10] based on the surrogate Z. Thus, estimating equations based on the empirical versions of the revised moment conditions can be constructed. In the regression setup, if $E(Y|X = x) = g(x;\beta)$, then they showed that for a twice differentiable weight function $h(x)$,

$$E\left[h(Z)R(Y,Z;\beta) - \frac{\sigma_u^2}{2}(h''(Z)R(Y,Z;\beta) - 2h'(Z)g'(Z;\beta) - h(Z)g''(Z;\beta))\right] = 0,$$

where $R(Y,Z;\beta) = Y - g(Z;\beta)$, $h'(z)$, $h''(z)$ are the first and second order derivatives of h w.r.t. z, and $g'(z;\beta)$,$g''(z;\beta)$ are the first and second order derivatives of g w.r.t. z. The asymptotic properties of the estimator based on the revised moment estimating equation are also discussed in [10]. Although the revised moment estimation procedure has a clean mathematical structure and a nice asymptotic theory, the relatively complicated revised moment estimating equation might create some computational difficulties when implementing the method. This is extremely undesirable for applied statisticians, in particular at the beginning stage of a study, they might only want to do some exploratory analysis. If so, the proposed L-SIMEX procedure surely can satisfy their needs. For the illustration purpose, we conduct a comparison study through simulation to show that the finite sample performance of the L-SIMEX procedure versus the revised moment estimation procedure.

Logistic regression: L-SIMEX vs. revised moment estimate

Consider the regression model $Y = I[\beta_0 + \beta_1 X + \varepsilon \geq 0]$, where ε follows a standard logistic distribution. The same model is also discussed in [10], with the exception of σ_u^2 being known here. For any measurable function $h(x)$, we have $E[h(X)(Y - F(\beta_0 + \beta_1 X))] = 0$, where F is the logistic CDF. In this case, the revised moment equations have complicated forms. The following two different

Table 5
MSE of regression parameter estimators

	$\hat{\beta}_0$			$\hat{\beta}_1$		
Sample Size	$h_1(x)$	$h_2(x)$	L-SIMEX	$h_1(x)$	$h_2(x)$	L-SIMEX
500	0.090	0.615	0.35	0.299	0.178	0.319
600	0.087	0.585	0.316	0.294	0.136	0.312
700	0.077	0.627	0.279	0.289	0.146	0.282
800	0.074	0.579	0.284	0.289	0.155	0.289
900	0.070	0.565	0.278	0.249	0.152	0.288

weight functions h are used to see their effects on the estimation of β_0 and β_1,

$$h_1(x) = \begin{pmatrix} 1 \\ x^3 \end{pmatrix}, \quad h_2(x) = \begin{pmatrix} x^2 \\ x^3 \end{pmatrix}.$$

The values of the true parameters β_0, β_1 and σ_u^2 are chosen to be 1's, and the logistic distribution has location parameter 0 and scale parameter 1. The sample sizes n are chosen to be 500, 600,700,800 and 900, and the simulation is replicated 200 times for each sample size. The MSE is used to evaluate the performance of the revised moment estimation procedures. The simulation results are shown in Table 5. Although L-SIMEX estimates have larger MSEs, in particular, when estimating β_1, it still provides reasonably good estimates for the parameters. Since L-SIMEX only needs a solution for the empirical version of the moment condition $Eh(X)(Y - g(X; \beta)) = 0$, which is usually much simpler than the one based on the revised moment condition, so it remains a competitive estimation procedure over the revised moment estimation methods, in particular, if the empirical version of the revised moment condition is hard to solve.

Acknowledgments

The research is supported by the collaborative grants from NSF DMS 1205271, 1205276. The authors would like to thank Professor Raymond J. Carroll for allowing us to use the Framingham Heart Study data set which he obtained from The National Heart, Lung, and Blood Institute. The authors also thank the two referees for their constructive comments.

References

[1] Carroll, R. J. and Hall, P. (1988). Optimal rates of convergence for deconvolving a density. *Journal of the American Statistical Association*. **83** 1184–1186. MR0997599

[2] Carroll, R. J., Küchenhoff, H., Lombard, F., and Stefanski, L. A. (1996). Asymptotics for the SIMEX estimator in nonlinear measurement error models. *Journal of the American Statistical Association*. **91** 242–250. MR1394078

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

[4] Dwork, C., Mcsherry, F., Nissim, K, and Smith, A. (2006). Calibrating noise to sensitivity in private data analysis. *Theory of Cryptography, Lecture Notes in Computer Science*. **3876** 265–284. MR2241676

[5] Fan, J. (1991a). On the optimal rates of convergence for nonparametric deconvolution problems. *Annals of Statistics*. **19** 1257–1272. MR1126324

[6] Fan, J. (1991b). Asymptotic normality for deconvolving kernel density estimators. *Sankhya, Series A*. **53** 97–110. MR1177770

[7] Fan, J. (1991c). Global behavior of deconvolution kernel estimates. *Statistica Sinica*. **1** 541–551. MR1130132

[8] Furman, E. (2008). On a multimvariate gamma distribution. *Statistics and Probability Letters*. **78** 2353–2360. MR2462670

[9] Holm, H. and Alouini, M. S. (2004). Sum and difference of two squared corrected Nakagami variates in connection with the McKay distribution. *IEEE Transactions on Communications*. **52** 1367–1376.

[10] Hong, H. and Tamer, E. (2003). A simple estimator for nonlinear error in variable models. *Journal of Econometrics*. **117** 1–19. MR2002282

[11] Kotz, S., Kozubowski, T., and Podgórski, K. (2001). *The Laplace Distribution and Generalizations*. Boston: Birkhäuser. MR1935481

[12] Liu, M. C. and Taylor, R. L (1989). A consistent nonparametric density estimator for the deconvolution problem. *Canadian Journal of Statistics*. **17** 399–410. MR1047309

[13] Mckenzie, H., Jerde, C., Visscher, D., Merrill, E., and Lewis, M. (2009). Inferring linear feature use in the presence of GPS measurement error. *Environ. Ecol. Stat.* **16** 531–546. MR2565367

[14] Mathai, A. M. and Moschopoulos, P. G. (1991). On multivariate Gamma. *Journal of Multivariate Analysis*. **39** 135–153. MR1128677

[15] Mathai, A. M. and Moschopoulos, P. G. (1992). A form of multivariate Gamma distribution. *Ann. Inst. Statist. Math.* **44** 97–106. MR1165574

[16] Purdom, E. and Holmes, S. (2005). Error distribution for gene expression data. *Statistical Applications in Genetics and Molecular Biology*. **4**, Article 16. MR2170432

[17] Rousson, V. and Dodge, Y. (1999). The complication of the fourth central moment. *The American Statistician*. **53** 267–269. MR1711547

[18] Polzehl, J. and Zwanzig, S. (2004). SIMEX and TLS: An equivalence result. *Preprint Weiererstrass Insttitute Berlin, ISSN0946-8633*. **No 999**. MR2101546

[19] Stefanski, L. A. and Carroll, R. J. (1990). Deconvoluting kernel density estimators *Statistics*. **21** 165–184. MR1054861

[20] Stefanski, L. A. and Cook, J. R. (1995). Simulation-Extrapolation: The measurement error jackknife. *Journal of the American Statistical Association*. **90** 1247–1256. MR1379467

[21] Volkmer, H. (2014). A characterization of the normal distribution. *To appear*.

[22] Wang, H., Stefanski, L., and Zhu, Z. (2012). Corrected-loss estimation for quantile regression with covariate measurement errors. *Biometrika*. **99** 405–421. MR2931262