Bootstrap Method for Measurement Error Model

by **Q. T. Thach** and **N. G. N. Prasad**

Q. T. Thach
Department of Community Medicine
The University of Hong Kong
Patrick Manson Building South Wing
7 Sassoon Road
Hong Kong

N. G. N. Prasad
Department of Mathematical Sciences
University of Alberta
Edmonton, Alberta, T6G–2G1
Canada

In this paper, we investigate the classical and a weighted bootstrap methods for the structural relationship model with known variance ratio for an arbitrary error distribution. Our bootstrap procedures perform well even in comparison with the normal theory estimates in normal situations, i.e., they have better coverage accuracy than the normal approximation.

1 Introduction

It has been generally recognized that true measurements of a characteristic are often difficult to observe but they are observed with measurement errors. In view of this, considerable effort has been expended on the development of methods of analyzing data which are contaminated with measurement errors. In this chapter, we consider a model that is related to measurement errors. This model can be viewed as a generalization of the simple regression model, which takes into account random measurement errors on both the dependent and independent variables.

More precisely, we assume that two random variables U_1 and U_2 are observed subject to measurement error, both are related by $U_2 = \alpha + \beta U_1$, where α and β are unknown parameters. Further, we assume that actual observed values are $X = U_1 + \delta$ and $Y = U_2 + \varepsilon$. The pairs $(\delta_i, \varepsilon_i)$ are independently distributed for different i's, and δ_i and ε_i may or may not be independent of each other although they are independent of (U_{1i}, U_{2i}). When U_1 and U_2 are assumed to be unknown constants, the model is known as a functional model, whereas if the U_{1i}'s are independent random variables with the same distribution the model is known as a structural model. Fuller (1987) provided more details on these models.

It is well–known (see Kendall and Stuart, 1979 and Riersøl, 1950) that if δ and ε are normally distributed, then α and β are unidentifiable if and only if U_1 and U_2 are constants or U_1 and U_2 are normally distributed. Hence, assuming normal errors and without further information, α and β cannot be estimated in the functional model or in the structural model. However, if the error variance ratio $\lambda^2 = \sigma_\varepsilon^2/\sigma_\delta^2$ is known then both α and β can be estimated consistently when $\sigma_{U_1}^2$, σ_δ^2 and σ_ε^2 are finite. The above assumption is often satisfied if X and Y represent similar characteristics measured in the same units, in which case $\lambda^2 = 1$. In other instances, information from another independent sample, such as a preliminary study, often provides a suitable value for λ^2. In the functional case where the U_{1i}'s are true unknown values, Solari (1969) pointed out that the solution of maximum likelihood equations was a saddle point. Birch (1964) and Barnett (1967) obtained the maximum likelihood solution when both σ_δ^2 and σ_ε^2 were known. For a comprehensive coverage of this work and related topics on this subject, see for example Fuller (1987), Gleser (1981), and Chan and Mak (1983) on a multivariate model. Lindley and El-Sayyad (1968) and Zellner (1971) considered a Bayesian approach to these models.

Much of the interest has been focused on the estimation and testing procedure of β. Little investigation has been devoted to bootstrap procedure for estimating the standard error and confidence interval for β. The sampling distribution of the regression estimator $\hat{\beta}$ is skewed (see Anderson and Sawa, 1982). As a result, the large sample normal approximation as well as the likelihood ratio chi–square approximation performs poorly for small samples. In contrast, the bootstrap sampling distribution incorporates the skewness of the true sampling distributions. This feature is referred to as the second-order correctness of the bootstrap. Babu and Bai (1992) obtained a two–term Edgeworth expansion for $\hat{\beta}$ for a linear functional error–in–variables model. They showed that by using these expansions the bootstrap approximation of the sampling distribution was superior to the classical normal approximation. Linder and Babu (1994) proposed a bootstrap procedure based on the residuals for functional measurement error model with known error variance ratio and symmetric errors. However, their method is cumbersome, in part, because it involves calculations of correction factors so that the first two moments of the bootstrap estimator $\hat{\beta}^*$ match with the usual estimates of the first two moments of $\hat{\beta}$. Moreover, implementation of this approach requires a different correction factor for each parameter. Kelly (1984) considered the structural model with known error variance ratio from the influence function of β and obtained an estimate of the variance of $\hat{\beta}$. She noted that this method performed poorly, as the influence function estimate of the variance was biased.

In this paper, we investigate the classical and a weighted bootstrap methods for the structural relationship model with known variance ratio for an arbitrary error distribution. Wu (1986) first proposed the weighted bootstrap in the context of the classical regression problem. In this procedure i.i.d. $\{t_i, i = 1, 2, ..., n\}$ observations are drawn from an external population having mean 0 and variance 1, independent of the original data. For the second order accuracy of the bootstrap estimator based on this method, Liu (1988) suggested another restriction on the external population, namely that third central moment of t_i must also be equal 1. This paper is divided into three main sections. In Section 2, the method of moments is used to estimate the parameters and some preliminary results on asymptotic properties of the estimators are also given. Section 3 reviews the Linder and Babu method and describes the proposed bootstrap methods along with their asymptotic properties. In Section 4, the results of a simulation study are given.

2 The structural model

Consider the structural equations model for n random vectors $\mathbf{Z}_i = (X_i, Y_i)^T$. It is assumed that for each $i = 1 \ldots n$, we have

$$\mathbf{Z}_i = \begin{pmatrix} X_i \\ Y_i \end{pmatrix} = \begin{pmatrix} U_{1i} \\ U_{2i} \end{pmatrix} + \begin{pmatrix} \delta_i \\ \varepsilon_i \end{pmatrix} = \mathbf{U}_i + \boldsymbol{\xi}_i, \tag{2.1}$$

$$U_{2i} = \alpha + \beta U_{1i}, \tag{2.2}$$

where the $\mathbf{U}_i$'s are independently distributed with mean vector $\boldsymbol{\mu}$ and covariance matrix $\boldsymbol{\Gamma}_{\mathbf{U}}$, with

$$\boldsymbol{\mu} = \begin{pmatrix} \mu_1 \\ \mu_2 \end{pmatrix} \quad \text{and} \quad \boldsymbol{\Gamma}_{\mathbf{U}} = \begin{pmatrix} \sigma_{U_1}^2 & \beta\sigma_{U_1}^2 \\ & \beta^2\sigma_{U_1}^2 \end{pmatrix}. \tag{2.3}$$

The random variables $\boldsymbol{\xi}_i$'s are i.i.d. with mean vector $\mathbf{0}$ and covariance matrix

$$\boldsymbol{\Gamma}_{\boldsymbol{\xi}} = \begin{pmatrix} \sigma_\delta^2 & 0 \\ & \sigma_\varepsilon^2 \end{pmatrix}. \tag{2.4}$$

such that

$$\lambda^2 = \sigma_\varepsilon^2/\sigma_\delta^2 \text{ is known.} \tag{2.5}$$

Further, we assume that for each i,

$$\mathbf{U}_i \text{ and } \boldsymbol{\xi}_i \text{ are independent.} \tag{2.6}$$

Let F denote the common distribution of the $\mathbf{Z}_i$'s. By (2.1)–(2.6), the mean vector $\boldsymbol{\mu}_F$ and covariance matrix $\boldsymbol{\Gamma}_F$ are, respectively, given by

$$\boldsymbol{\mu}(F) = \begin{pmatrix} \mu_X(F) \\ \mu_Y(F) \end{pmatrix} = \begin{pmatrix} \mu_1 \\ \alpha + \beta\mu_1 \end{pmatrix} \tag{2.7}$$

and

$$\boldsymbol{\Gamma}(F) = \begin{pmatrix} \sigma_{XX}(F) & \sigma_{XY}(F) \\ & \sigma_{YY}(F) \end{pmatrix} = \begin{pmatrix} \sigma_{U_1}^2 + \sigma_\delta^2 & \beta\sigma_{U_1}^2 \\ & \beta^2\sigma_{U_1}^2 + \sigma_\varepsilon^2 \end{pmatrix}. \tag{2.8}$$

If we substitute for U_1 and U_2 from (2.1) into (2.2), we obtain

$$Y_i = \alpha + \beta X_i + \varepsilon_i - \beta\delta_i, \quad i = 1, \ldots, n. \tag{2.9}$$

This is not a classical regression model since here, X is a random variable which is correlated with the error term $(\varepsilon - \beta\delta)$. From (2.3)–(2.8), we have

$$Cov_F(X, \varepsilon - \beta\delta) = -\beta\sigma_\delta^2, \tag{2.10}$$

which is 0 only if $\sigma_\delta^2 = 0$, the case corresponding to the simple regression situation, or in the trivial case $\beta = 0$. Thus, the existence of errors in both U_1 and U_2 poses a problem quite distinct from that of conventional regression model.

2.1 The method of moments estimators

The parameter vector $\boldsymbol{\theta} = (\alpha, \beta)^T$ can be written as a functional of the unknown distribution function F (see Kelly, 1984) by letting

$$\alpha = \alpha(F) = \mu_Y(F) - \beta(F)\mu_X(F), \tag{2.11}$$

$$\begin{aligned}\beta &= \beta(F)\\ &= \frac{1}{2\sigma_{XY}(F)}[\sigma_{YY}(F) - \lambda^2\sigma_{XX}(F) + \{[\sigma_{XY}(F) - \lambda^2\sigma_{XX}(F)]^2 + 4\lambda^2\sigma_{XY}^2(F)\}^{1/2}],\end{aligned} \tag{2.12}$$

where by definition,

$$\mu_Y(F) = \int\int y dF(x,y),$$
$$\sigma_{XX}(F) = \int\int x^2 dF(x,y) - \left[\int\int x dF(x,y)\right],$$

and the other quantities are defined in a similar fashion. Note that $\beta(F)$ may be rewritten as

$$\beta(F) = h(F) + [h^2(F) + \lambda^2]^{1/2}, \tag{2.13}$$

with

$$h = h(F) = \frac{1}{2\sigma_{XY}(F)}\{\sigma_{YY}(F) - \lambda^2\sigma_{XX}(F)\}. \tag{2.14}$$

Here and in what follows, for any sequences $\{H_i\}$ and $\{R_i\}$, we use the notation

$$\bar{H} = n^{-1}\sum_{i=1}^{n} H_i, \quad S_{HR} = n^{-1}\sum_{i=1}^{n}(H_i - \bar{H})(R_i - \bar{R}). \tag{2.15}$$

Let F_n denote the sample distribution function corresponding to F. Denote the sample mean and covariance matrix, respectively, by

$$\boldsymbol{\mu}(F_n) = \begin{pmatrix} \bar{X} \\ \bar{Y} \end{pmatrix} \text{ and } \quad \boldsymbol{\Gamma}(F_n) = \begin{pmatrix} S_{XX} & S_{XY} \\ & S_{YY} \end{pmatrix}. \tag{2.16}$$

Under the model defined by (2.1)–(2.6), the method of moments estimator of $\boldsymbol{\theta}(F) = (\alpha(F), \beta(F))^T$ is given by

$$\boldsymbol{\theta}(F_n) = (\alpha(F_n), \beta(F_n))^T, \tag{2.17}$$

where

$$\hat{\alpha} = \alpha(F_n) = \bar{Y} - \beta(F_n)\bar{X}, \tag{2.18}$$

$$\hat{\beta} = \beta(F_n) = h(F_n) + (h^2(F_n) + \lambda^2)^{1/2}, \tag{2.19}$$

with

$$\hat{h} = h(F_n) = \frac{1}{2S_{XY}}\{S_{YY} - \lambda^2 S_{XX}\}. \tag{2.20}$$

By the law of large numbers, $\hat{\alpha}$ and $\hat{\beta}$ are consistent estimators for α and β, respectively, for all distribution functions F with finite second moments. When F is bivariate normal, $\boldsymbol{\theta}(F_n)$ is the maximum likelihood estimator of $\boldsymbol{\theta}(F)$ (see Kendall and Stuart, 1979).

Theorem 2.1. *Let the model defined by (2.1)–(2.6) hold with known error variance ratio $\lambda^2 > 0$ and suppose that X and Y have finite sixth moment. Then*

(i)

$$E_F(\hat{\beta} - \beta) = -\frac{\beta}{2n\mu_{11}^2\sqrt{h^2 + \lambda^2}}\left\{\mu_{13} - \lambda^2\mu_{31} - 2h\mu_{22}\right\} + O(n^{-2}). \tag{2.21}$$

(ii)

$$E_F(\hat{\beta} - \beta)^2 = \frac{\beta^2}{4n\mu_{11}^2(h^2 + \lambda^2)}\left\{\mu_{04} + \lambda^4\mu_{40} + 2\mu_{22}(2h^2 - \lambda^2) - 4h(\mu_{13} - \lambda^2\mu_{31})\right\} + O(n^{-2}). \tag{2.22}$$

(iii) If the joint distribution of X and Y are symmetric, i.e., $\mu_{30} = \mu_{03} = \mu_{12} = \mu_{21} = 0$, then,

$$E_F(\hat{\beta} - \beta)^3 = \frac{1}{n^2}\left(\frac{\beta}{2\mu_{11}\sqrt{h^2+\lambda^2}}\right)^3 \{\mu_{06} - \lambda^6\mu_{60} - 3\lambda^2(\mu_{24} - \lambda^2\mu_{42}) \\ -6h[\mu_{15} + \lambda^4\mu_{51} - 2\lambda^2\mu_{33}] + 12h^2(\mu_{24} - \lambda^2\mu_{42}) - 8h^3\mu_{33}\} \\ + O(n^{-3}) \tag{2.23}$$

where $\mu_{uv} = E(X - \mu_X)^u(Y - \mu_Y)^v$.

Proof. See Appendix A. □

The expression in (2.22) can also be derived from the influence function for $\hat{\beta}$; the details are given in Kelly (1984). To get an idea of the magnitude of the bias of $\hat{\beta}$, we consider that the population follows a standard bivariate normal distribution, so that

$$\phi(x, y) = \frac{1}{2\pi\sqrt{1-\rho^2}} \exp\left[-\frac{1}{2(1-\rho^2)}(x^2 - 2\rho xy + y^2)\right]; \tag{2.24}$$

and

$$\mu_{21} = \mu_{12} = \mu_{30} = \mu_{03} = 0, \quad \mu_{20} = \mu_{02} = 1, \quad \mu_{31} = \mu_{13} = 3\rho, \\ \mu_{22} = 1 + 2\rho^2, \quad \mu_{40} = \mu_{04} = 3, \quad \mu_{24} = \mu_{42} = 3 + 12\rho^2 \text{ and } \mu_{33} = 9\rho + 6\rho^3.$$

This yields

$$E_F(\hat{\beta} - \beta) = \frac{\beta h}{\sqrt{h^2+\lambda^2}}\left(\frac{1}{n} - \frac{4}{n^2}\right)\left(\frac{1-\rho^2}{\rho^2}\right) + O(n^{-3}). \tag{2.25}$$

Denoting the first and second order approximation of the relative bias of $\hat{\beta}$ by $B_1(\hat{\beta})$ and $B_2(\hat{\beta})$, respectively, we have

$$B_1(\hat{\beta}) = \frac{E_F(\hat{\beta}) - \beta}{\beta} \doteq \frac{h}{n\sqrt{h^2+\lambda^2}}\left(\frac{1-\rho^2}{\rho^2}\right). \tag{2.26}$$

To a second approximation, the relative bias of $\hat{\beta}$ can, therefore, be expressed as

$$B_2(\hat{\beta}) \doteq B_1(\hat{\beta})\left(1 - \frac{4}{n}\right). \tag{2.27}$$

Equation (2.27) shows that the contribution of the second and third order terms to the relative bias of $\hat{\beta}$ is $4/n$ times the value of the latter to a first approximation. Unless n is small, the contribution can be considered negligible.

Comparing (2.21) and (2.22), we see that both the bias and the variance of $\hat{\beta}$ are of order n^{-1}. Hence, for n sufficiently large, the bias is negligible as compared to the standard error which is of the order $n^{-1/2}$.

The exact sampling behavior of the estimator $\hat{\beta}$ defined in (2.19) cannot be obtained easily. Therefore, it seems necessary to use large sample theory to develop an approximation of the distribution of $\hat{\beta}$. We now give the asymptotic normal distribution of the estimators for the slope $\hat{\beta}$ and intercept $\hat{\alpha}$, under the general structural linear relationship (2.1)–(2.6).

Theorem 2.2. *Let the model defined by (2.1)–(2.6) hold with known error variance ratio $\lambda^2 > 0$ and X and Y have finite sixth moment. Then, as $n \to \infty$,*

1. $n^{1/2}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta}) \xrightarrow{d} N(\mathbf{0}, \boldsymbol{\Sigma})$, *for* $\hat{\boldsymbol{\theta}} = (\hat{\alpha}, \hat{\beta})^T$, *where*

$$\boldsymbol{\Sigma} = \begin{bmatrix} \sigma_\delta^2(\beta^2 + \lambda^2) + \mu_1^2\Sigma_{22} & -\mu_1\Sigma_{22} \\ -\mu_1\Sigma_{22} & \Sigma_{22} \end{bmatrix},$$

and

$$\Sigma_{22} = \frac{\beta^2}{4\mu_{11}^2(h^2 + \lambda^2)} \left\{\mu_{04} + \lambda^4\mu_{40} + 2\mu_{22}(2h^2 - \lambda^2) - 4h(\mu_{13} - \lambda^2\mu_{31})\right\}.$$

2. *Furthermore,* $\hat{\boldsymbol{\Sigma}}$ *converges in probability to* $\boldsymbol{\Sigma}$, *where*

$$\hat{\boldsymbol{\Sigma}} = \begin{bmatrix} \hat{\sigma}_\delta^2(\hat{\beta}^2 + \lambda^2) + \hat{\mu}_1^2\hat{\Sigma}_{22} & -\hat{\mu}_1\hat{\Sigma}_{22} \\ -\hat{\mu}_1\hat{\Sigma}_{22} & \hat{\Sigma}_{22} \end{bmatrix},$$

$$\hat{\sigma}_\delta^2 = \sum_{i=1}^{n} e_i^2/(n(\hat{\beta}^2 + \lambda^2)), \quad e_i = Y_i - \hat{\alpha} - \hat{\beta}X_i, \quad \hat{\mu}_1 = n^{-1}\sum_{i=1}^{n} \hat{U}_{1i},$$

$$\hat{\Sigma}_{22} = \frac{\hat{\beta}^2}{4\hat{\mu}_{11}^2(\hat{h}^2 + \lambda^2)} \left\{\hat{\mu}_{04} + \lambda^4\hat{\mu}_{40} + 2\hat{\mu}_{22}(2\hat{h}^2 - \lambda^2) - 4\hat{h}(\hat{\mu}_{13} - \lambda^2\hat{\mu}_{31})\right\},$$

with $\hat{U}_{1i}$ is defined in (3.1) and $\hat{\mu}_{rs}$ is a plug-in sample moment estimate, and is given by
$\hat{\mu}_{rs} = n^{-1}\sum_{i=1}^{n}(X_i - \bar{X})^r(Y_i - \bar{Y})^s$.

Proof. We note that S_{YY}, S_{XX} and S_{XY} are asymptotically unbiased and from Kendall and Stuart (1979), we have the following results

$$\begin{aligned} Var_F[S_{YY}] &= n^{-1}(\mu_{04} - \mu_{02}^2), & Var_F[S_{XX}] &= n^{-1}(\mu_{40} - \mu_{20}^2), \\ Var_F[S_{XY}] &= n^{-1}(\mu_{22} - \mu_{11}^2), & Cov_F[S_{YY}, S_{XY}] &= n^{-1}(\mu_{13} - \mu_{02}\mu_{11}), \\ Cov_F[S_{XX}, S_{XY}] &= n^{-1}(\mu_{31} - \mu_{20}\mu_{11}), & Cov_F[S_{YY}, S_{XX}] &= n^{-1}(\mu_{22} - \mu_{02}\mu_{20}). \end{aligned}$$

Because the sample moments are converging in probability to their respective population moments, we can expand $\hat{\beta}$ using a Taylor series expansion about β to obtain

$$\hat{\beta} = \beta + \beta(h^2 + \lambda^2)^{-1/2}(\hat{h} - h) + O_p(n^{-1}), \tag{2.28}$$

or equivalently,

$$n^{1/2}(\hat{\beta} - \beta) = n^{1/2}\beta(\hat{h} - h)(h^2 + \lambda^2)^{-1/2} + O_p(n^{-1/2}), \tag{2.29}$$

which implies that the limiting distribution of $n^{1/2}(\hat{\beta} - \beta)$ is the same as that of $n^{1/2}\beta(\hat{h} - h)(h^2 + \lambda^2)^{-1/2}$. The asymptotic variance of $\hat{h}$ is given by

$$\begin{aligned} Var_F(\hat{h}) = h^2 \Big\{ & \frac{1}{E_F^2[S_{YY} - \lambda^2 S_{XX}]} \big(Var_F[S_{YY}] - 2\lambda^2 Cov_F[S_{YY}, S_{XX}] \\ & + \lambda^4 Var_F[S_{XX}] \big) + \frac{Var[S_{XY}]}{E_F^2[S_{XY}]} - \frac{2}{E_F[S_{XY}] E_F[S_{YY} - \lambda^2 S_{XX}]} \\ & \big(Cov_F[S_{YY}, S_{XY}] - \lambda^2 Cov_F[S_{XX}, S_{XY}] \big) \Big\} \\ > \quad = {} & \frac{1}{4n\mu_{11}^4} \{ \mu_{11}^2 [\mu_{04} - \mu_{02}^2 - 2\lambda^2(\mu_{22} - \mu_{02}\mu_{20}) + \lambda^4(\mu_{40} - \mu_{20}^2)] \\ & + 4h^2\mu_{11}^2(\mu_{22} - \mu_{11}^2) - 2\mu_{11}(\mu_{02} - \lambda^2\mu_{20}) \\ & \times [\mu_{13} - \mu_{02}\mu_{11} - \lambda^2(\mu_{31} - \mu_{20}\mu_{11})] \} + O(n^{-2}) \\ = {} & \frac{1}{4n\mu_{11}^2} \big\{ \mu_{04} + \lambda^4\mu_{40} + 2\mu_{22}(2h^2 - \lambda^2) - 4h(\mu_{13} - \lambda^2\mu_{31}) \big\} + O(n^{-2}). \end{aligned}$$

Hence the asymptotic variance of $\hat{\beta}$ is given by

$$Var_F(\hat{\beta}) = \frac{\beta^2}{h^2 + \lambda^2} Var_F(\hat{h}) + O(n^{-2}).$$

Turning to $Var_F(\hat{\alpha})$, consider

$$\begin{aligned} \hat{\alpha} = \hat{Y} - \hat{\beta}\bar{X} &= \alpha + \beta U_{1i} + \bar{\varepsilon} - \hat{\beta}(\bar{U}_{1i} + \bar{\delta}) \\ &= \alpha - (\hat{\beta} - \beta)\mu_1 + \bar{v} + O_p(n^{-1}), \end{aligned}$$

where $\bar{v} = \bar{\varepsilon} - \beta\bar{\delta}$ and hence the distribution of $n^{1/2}(\hat{\alpha} - \alpha)$ has the same distribution as that of $n^{1/2}[\bar{v} - (\hat{\beta} - \beta)\mu_1)]$. This yields

$$\begin{aligned} Var_F(\hat{\alpha}) &= Var_F(\bar{v} - (\hat{\beta} - \beta)\mu_1) \\ &= \sigma_\delta^2(\beta^2 + \lambda^2) + \mu_1^2 Var_F(\hat{\beta}), \end{aligned} \tag{2.30}$$

and

$$\begin{aligned} Cov_F(\hat{\alpha}, \hat{\beta}) &= Cov_F[\bar{v} - (\hat{\beta} - \beta)\mu_1), (\hat{\beta} - \beta)\mu_1)] \\ &= -\mu_1 Var_F(\hat{\beta}). \end{aligned} \tag{2.31}$$

To show the asymptotic normality of $n^{1/2}(\hat{\boldsymbol{\theta}} - \boldsymbol{\theta})$ and the consistency of $\hat{\boldsymbol{\Sigma}}$, we note that $\hat{\boldsymbol{\theta}}$ is a continuous differentiable function of the U-statistics $(\bar{X}, \bar{Y}, S_{XX}, S_{YY}, S_{XY})$. Thus by Theorems 8 and 9 of Arvesen (1969), the desired results follow. □

Because $\hat{\Sigma}$ is a consistent estimator of Σ, it follows that

$$t = n^{1/2}\hat{\Sigma}_{22}^{-1/2}(\hat{\beta} - \beta) \tag{2.32}$$

is approximately distributed as a $N(0, 1)$ random variable. In practice it seems reasonable to approximate the distribution of (2.32) with the distribution of Student's t with $n - 2$ degree of freedom. Instead of using $n^{-1}\hat{\Sigma}_{22}$ to estimate $Var_F(\hat{\beta})$, one could use a jackknife procedure. The next section describes a jackknife procedure.

2.2 Jackknife variance estimation

In this section, we give the jackknife variance estimator for $\hat{\boldsymbol{\theta}}$ (see Kelly, 1984). Let $\hat{\boldsymbol{\theta}}_{-i}$ be an estimator of $\boldsymbol{\theta}$ with the i-th observation (X_i, Y_i) omitted and define the pseudo values

$$\hat{\boldsymbol{\theta}}_i = n\hat{\boldsymbol{\theta}} - (n-1)\hat{\boldsymbol{\theta}}_{-i}. \tag{2.33}$$

The jackknife estimator of $\hat{\boldsymbol{\theta}}$ is

$$\hat{\boldsymbol{\theta}}_{(\cdot)} \quad = n^{-1}\sum \hat{\boldsymbol{\theta}}_i. \tag{2.34}$$

The jackknife estimator of the variance-covariance matrix of $\hat{\boldsymbol{\theta}}$ is

$$\hat{\boldsymbol{\Sigma}}_J = \frac{n-1}{n}\sum[\hat{\boldsymbol{\theta}}_i - \hat{\boldsymbol{\theta}}_{(\cdot)}][\hat{\boldsymbol{\theta}}_i - \hat{\boldsymbol{\theta}}_{(\cdot)}]^T. \tag{2.35}$$

Since $\hat{\alpha}$ and $\hat{\beta}$ are continuously differentiable functions of the U-statistics $(\bar{X}, \bar{Y}, S_{XX}, S_{YY}, S_{XY})$ and when $E_F(X^4) < \infty$ and $E_F(Y^4) < \infty$, by Theorem 9 of Arvessen (1969), we have

$$n\hat{\mathbf{\Sigma}}_J - \mathbf{\Sigma} \xrightarrow{p} 0 \quad \text{as} \quad n \to \infty. \tag{2.36}$$

3 The Linder and Babu method

Linder and Babu (1994) proposed a bootstrap method where resampling was done by taking a sample with replacement from the residuals and then repeating this a number of times to match the usual variance estimates. However, in such resampling method, one needs to modify the residuals and the usual bootstrap variance estimator.

Let $\hat{U}_{1i}$ and $\hat{U}_{2i}$ denote the fitted values U_{1i} and U_{2i}, respectively. We require, for every i, that $\lambda^2 = (Y_i - \hat{U}_{2i})^2/(X_i - \hat{U}_{1i})^2$, which in turn requires the redefinition of the fitted values,

$$\begin{aligned} \hat{U}_{1i} &= X_i + r_i/(\lambda + |\hat{\beta}|), \\ \hat{U}_{2i} &= \hat{\alpha} + \hat{\beta}\hat{U}_{1i} = Y_i - \lambda e_i/(\lambda + |\hat{\beta}|), \end{aligned} \tag{3.1}$$

where $e_i = Y_i - \hat{\alpha} - \hat{\beta}X_i$.

The residuals $(X_i - \hat{U}_{1i}, Y_i - \hat{U}_{2i})$ underestimate the true error, i.e. the mean squared of the residuals are asymptotically negatively biased for variances $\hat{\sigma}_\delta^2$ and $\hat{\sigma}_\varepsilon^2$, respectively. This results from the fact that

$$s^2 = \hat{\sigma}_\delta^2 = \sum_1^n e_i^2/(n(\lambda^2 + \hat{\beta}^2)) \tag{3.2}$$

is a consistent estimator for σ_δ^2, see Kendall and Stuart (1979). Hence, to estimate the error variances consistently, the residuals are adjusted by multiplying with the correction factor $d_n = (\lambda + |\hat{\beta}|)/(\lambda^2 + \hat{\beta}^2)^{1/2}$, resulting in a set of "pseudo" residuals

$$\begin{aligned} r_i &= -e_i/(\lambda^2 + \hat{\beta}^2)^{1/2}, \\ s_i &= \lambda e_i/(\lambda^2 + \hat{\beta}^2)^{1/2} = -\lambda r_i. \end{aligned} \tag{3.3}$$

We now describe the Linder and Babu bootstrap algorithm

1. Given data set $(X_i, Y_i), i = 1, \ldots, n$ and the estimator $\hat{\beta}$ as defined in (2.19) for slope, compute the fitted values $(\hat{U}_{1i}, \hat{U}_{2i}), i = 1, \ldots, n$ using (3.1).

2. Resample $\hat{\delta}_i^*$ with replacement from the set $c_n\{r_1, \ldots, r_n\}$, where r_i are given (3.3) and $c_n = \hat{\beta} S_{\hat{U}_1 \hat{U}_1} / S_{XY}$.

3. Independently of (2), resample $\hat{\varepsilon}_i^*$ with replacement from the set $c_n\{s_1, \ldots,$
$s_n\}$, where s_i are given by (3.3) and c_n is the same as in step (2).

4. Obtain bootstrap estimate $\hat{\beta}^*$ which is the analogue of $\hat{\beta}$ as in (2.19) from the replicate data (X_i^*, Y_i^*).

5. Repeat steps (1)–(5) above B times, where B is large, typically between 100 and 1,000.

6. Calculate estimates of $Var_F(\hat{\beta})$ by
$$var_*(\hat{\beta}^*) = (B-1)^{-1} \sum_{b=1}^{B} (\hat{\beta}_b^* - \hat{\beta})^2.$$

7. In the b-th run, calculate the bootstrap-t, defined as
$$t_b^* = (\hat{\beta}_b^* - \hat{\beta}) / \sqrt{\hat{\Phi}^*(\hat{\beta}^*)},$$
where $\hat{\beta}_b^*$ is the analogue of $\hat{\beta}$ computed in b-th bootstrap sample and $\hat{\Phi}^*(\hat{\beta}^*)$ is the bootstrap estimate of the variance of $\hat{\beta}^*$ given by
$$\hat{\Phi}^*(\hat{\beta}^*) = var_*(\hat{\beta}^*) + n^{-1}\hat{\Psi}$$
where $\hat{\Psi} = 4\lambda^6 \hat{\beta}^4 s^4 kur (\hat{\beta}^2 + \lambda^2)^{-4} \hat{v}^{-4}$ and
$$kur = \frac{n^{-1} SD^4 - 6\lambda^2 \hat{\beta}^2 s^4}{(\hat{\beta}^4 + \lambda^4) s^4} - 3$$
with $SD^4 = \sum_{i=1}^n e_i^4$, $\hat{v} = \hat{\mu}_{11}/\hat{\beta}$ and s^2 given by (3.2).

Remark: Linder and Babu (1994) noted that under this bootstrap procedure the usual variance estimator given in Step 6 above was not a proper estimator of $Var_F(\hat{\beta})$. Hence they suggested to use $\hat{\Phi}^*(\hat{\beta}^*) = var_*(\hat{\beta}^*) + n^{-1}\hat{\Psi}$. The computation of $\hat{\Psi}$ involves the fourth moment of residuals. To avoid this "after" correction, we propose two new bootstrap procedures where the usual bootstrap variance estimator is a valid estimator of $Var_F(\hat{\beta})$.

3.1 The proposed bootstrap procedure

The bootstrap procedure proposed here for the structural model differs from the classical method in that it does not resample the data (X_i, Y_i) directly. Instead it starts with an estimating function for the parameter and independently resamples residuals in that function. We assume the conditions in Theorem 2.1 hold and that $\hat{h}$ has a first order continuous derivative around $h = h(\mu_{02}, \mu_{20}, \mu_{11})$. By a multivariate Taylor series expansion of $\hat{h}$, we have

$$\begin{aligned}
\hat{h} &= h + \frac{1}{2}\mu_{11}^{-1}\{(\hat{\mu}_{02} - \mu_{02}) - \lambda^2(\hat{\mu}_{20} - \mu_{20}) - 2h(\hat{\mu}_{11} - \mu_{11})\} + O_p(n^{-1}) \\
&= h + \frac{1}{2}\mu_{11}^{-1}\{\hat{\mu}_{02} - \lambda^2\hat{\mu}_{20} - 2h\hat{\mu}_{11}\} + O_p(n^{-1}). \qquad (3.4)
\end{aligned}$$

We re-express (3.4) as

$$\begin{aligned}
\hat{h} &= h + \frac{1}{2n}\mu_{11}^{-1}\sum_{i=1}^{n}\{(Y_i - \bar{Y})^2 - \lambda^2(X_i - \bar{X})^2 - 2h(X_i - \bar{X})(Y_i - \bar{Y})\} + O_p(n^{-1}) \\
&= h + \frac{1}{2n}\mu_{11}^{-1}\sum_{i=1}^{n}\{(\varepsilon_i - \bar{\varepsilon})^2 - \lambda^2(\delta_i - \bar{\delta})^2 - 2h(\varepsilon_i - \bar{\varepsilon})(\delta_i - \bar{\delta}) \\
&\quad + 2\gamma U_{1i}(\varepsilon_i - \bar{\varepsilon}) - 2\gamma\beta U_{1i}(\delta_i - \bar{\delta})\} + O_p(n^{-1}) \\
&= h + \frac{1}{2n}\mu_{11}^{-1}\sum A_i + O_p(n^{-1}), \qquad (3.5)
\end{aligned}$$

where $A_i = (Y_i - \bar{Y})^2 - \lambda^2(X_i - \bar{X})^2 - 2h(X_i - \bar{X})(Y_i - \bar{Y})$ and $\gamma = \beta - h$.

Define

$$\begin{aligned}
\hat{A}_i &= (Y_i - \bar{Y})^2 - \lambda^2(X_i - \bar{X})^2 - 2\hat{h}(X_i - \bar{X})(Y_i - \bar{Y}) \\
&= (\hat{\varepsilon}_i - \bar{\hat{\varepsilon}})^2 - \lambda^2(\hat{\delta}_i - \bar{\hat{\delta}})^2 - 2\hat{h}(\hat{\varepsilon}_i - \bar{\hat{\varepsilon}})(\hat{\delta}_i - \bar{\hat{\delta}}) \\
&\quad + 2\hat{\gamma}\hat{U}_{1i}(\hat{\varepsilon}_i - \bar{\hat{\varepsilon}}) - 2\hat{\gamma}\hat{\beta}\hat{U}_{1i}(\hat{\delta}_i - \bar{\hat{\delta}}). \qquad (3.6)
\end{aligned}$$

where $\hat{\gamma} = \hat{\beta} - \hat{h}$. Equation (3.5) suggests that our bootstrap estimator for $\hat{h}$ is

$$\hat{h}^* = \hat{h} + \frac{1}{2n}\hat{\mu}_{11}^*\hat{\mu}_{11}^{-2}\sum_{i=1}^{n}\hat{A}_i^*, \qquad (3.7)$$

and using (2.12) that for $\hat{\beta}$ is

$$\hat{\beta}^* = \hat{h}^* + (\hat{h}^{2*} + \lambda^2)^{1/2}, \qquad (3.8)$$

where $\hat{\mu}_{11}^* = n^{-1}\sum_{i=1}^n (X_i^* - \bar{X})(Y_i^* - \bar{Y})$. Note that $\sum_{i=1}^n \hat{A}_i = 0$. We now describe the resampling algorithm for the slope parameter in the structural linear relationship (2.1)–(2.6).

1. Given a data set $(X_i, Y_i),\ i = 1, \ldots, n$ and the estimator (2.19) for the slope of the model, compute the fitted values $(\hat{U}_1, \hat{U}_2)$ using (3.1).

2. Resample $\hat{\delta}_i^*$, with replacement from the set $\{r_1, \ldots, r_n\}$, where r_i are given in (3.3).

3. Independently of step (2), resample $\hat{\varepsilon}_i^*$ at random with replacement from the set $\{s_1, \ldots, s_n\}$, where s_i are given in (3.3).

4. Compute $\hat{A}_i^* = (\hat{\varepsilon}_i^* - \bar{\hat{\varepsilon}})^2 - \lambda^2(\hat{\delta}_i^* - \bar{\hat{\delta}})^2 - 2\hat{h}(\hat{\varepsilon}_i^* - \bar{\hat{\varepsilon}})(\hat{\delta}_i^* - \bar{\hat{\delta}}) + 2\hat{\gamma}\hat{U}_{1i}(\hat{\varepsilon}_i^* - \bar{\hat{\varepsilon}}) - 2\hat{\gamma}\hat{\beta}\hat{U}_{1i}(\hat{\delta}_i^* - \bar{\hat{\delta}})$.

5. Compute

$$\hat{h}^* = \hat{h} + \frac{1}{2n}\hat{\mu}_{11}^*\hat{\mu}_{11}^{-2}\sum_{i=1}^n \hat{A}_i^*, \tag{3.9}$$

$$\hat{\beta}^* = \hat{h}^* + (\hat{h}^{2*} + \lambda^2)^{1/2}, \tag{3.10}$$

   where $\hat{\mu}_{11}^* = n^{-1}\sum_{i=1}^n (X_i^* - \bar{X})(Y_i^* - \bar{Y})$.

6. Repeat steps (1)–(5) above a large number of times, B, to obtain $\hat{\beta}_1^*, \ldots, \hat{\beta}_B^*$.

7. Calculate estimate of $Var_F(\hat{\beta})$ with

$$var_*(\hat{\beta}^*) = E_*(\hat{\beta}^* - E_*\hat{\beta}^*)^2, \tag{3.11}$$

   where E_* denotes expectation with respect to bootstrap sampling which can be approximated by

$$var_*(\hat{\beta}^*) = (B-1)^{-1}\sum_{b=1}^B (\hat{\beta}_b^* - \hat{\beta}_{(\cdot)})^2, \tag{3.12}$$

   where $\hat{\beta}_{(\cdot)}^* = \sum_{b=1}^B \hat{\beta}_b^*/B$.

8. In the b-th run, calculate the bootstrap-t:

$$t_b^* = \frac{\hat{\beta}_b^* - \hat{\beta}}{\sqrt{var_J^*(\hat{\beta}_b^*)}}, \quad b = 1, \ldots, B \tag{3.13}$$

where $var_J^*(\hat{\beta}_b^*)$ is the jackknife variance estimator of $\hat{\beta}_b^*$ given by

$$var_J^*(\hat{\beta}_b^*) = \frac{n-1}{n} \sum_{i=1}^{n} (\hat{\beta}_b^{*(i)} - \hat{\beta})^2 \tag{3.14}$$

with $\hat{\beta}_b^{*(i)}$ being the estimator of the same functional form as $\hat{\beta}_b^*$, but computed from the reduced sample of size $n-1$ obtained by omitting the i-th observation.

The asymptotic distribution of the bootstrap estimator proposed above is given in the following two theorems.

Theorem 3.1. *In the structural relationship (2.1)–(2.5), we assume* $E(X^6 + Y^6) < \infty$. *Then,* $n^{1/2}(\hat{\beta}^* - \hat{\beta})$ *converges in distribution to a normal random variable with zero mean and variance* $\hat{\Sigma}_{22}$, *where* $\hat{\beta}^*$ *is the bootstrap estimator of* $\hat{\beta}$ *resulting from the proposed resampling procedure in Section 3.1, and* $\hat{\Sigma}_{22}$ *as defined in (2.28).*

Proof. Let $\bar{\hat{A}}^* = n^{-1} \sum_{i=1}^{n} \hat{A}_i^*$ where $\hat{A}_i^*$ is defined in step (4) in the proposed bootstrap procedure above and write

$$\hat{h}^* = \hat{h} + \frac{\hat{\mu}_{11}^*}{\hat{\mu}_{11}} \frac{\bar{\hat{A}}^*}{2\hat{\mu}_{11}}. \tag{3.15}$$

We observe that $\bar{\hat{A}}^*$ is the mean of an i.i.d. sample with population mean $\bar{\hat{A}}$, where $\bar{\hat{A}} = n^{-1} \sum_{i=1}^{n} \hat{A}_i$. Then, by the central limit theorem, we have

$$\sqrt{n}(\bar{\hat{A}}^* - \bar{\hat{A}}) \xrightarrow{d} N(0, \hat{\sigma}_A^2), \tag{3.16}$$

with

$$\begin{aligned}
\hat{\sigma}_A^2 &= \frac{1}{n} \sum_{i=1}^{n} (\hat{A}_i - \bar{\hat{A}})^2 \\
&= \frac{1}{n} \sum_{i=1}^{n} \hat{A}_i^2 - \bar{\hat{A}}^2 \\
&= \hat{\mu}_{04} + \lambda^4 \hat{\mu}_{40} + 2\hat{\mu}_{22}(2\hat{h}^2 - \lambda^2) - 4\hat{h}(\hat{\mu}_{13} - \lambda^2 \hat{\mu}_{31}) + o_p(n^{-1})
\end{aligned} \tag{3.17}$$

By the law of large numbers $\hat{\mu}_{11}^* \xrightarrow{p} \hat{\mu}_{11}$ and by Slutsky's theorem, we have

$$\sqrt{n}(\hat{h}^* - \hat{h}) \stackrel{d}{=} \frac{1}{2\hat{\mu}_{11}} \sqrt{n}(\bar{\hat{A}}^* - \bar{\hat{A}}) \tag{3.18}$$

which implies that

$$\sqrt{n}(\hat{h}^* - \hat{h}) \xrightarrow{d} N(0, \frac{1}{4\hat{\mu}_{11}^2} \hat{\sigma}_A^2). \tag{3.19}$$

We conclude that $n^{1/2}(\hat{\beta}^* - \hat{\beta})$ converges to $N(0, \hat{\Sigma}_{22})$ in distribution by the delta method (see Bishop, Fienberg, and Holland, 1975). □

Theorem 3.2. *Assume the conditions of Theorem 3.1 hold. Then, under the proposed sampling procedure described in Section 3.1,*

(i) $E_F[E_*(\hat{\beta}^* - \hat{\beta})] = -E_F(\hat{\beta} - \beta) + O(n^{-2})$.
In addition, if the joint distribution of X and Y is symmetric, we have

(ii) $E_F[E_*(\hat{\beta}^* - \hat{\beta})^2] = E_F(\hat{\beta} - \beta)^2 + O(n^{-2})$ *and*

(iii) $E_F[E_*(\hat{\beta}^* - \hat{\beta})^3] = E_F(\hat{\beta} - \beta)^3 + O(n^{-3})$,

where E_* represents expectation with respect to the distribution induced by bootstrap sampling described in Section 3.1.

Proof. See Appendix B. □

3.2 The weighted bootstrap

Wu (1986) proposed a weighted bootstrap method in the context of classical regression Generally, the method entails first taking i.i.d. samples $\{t_i; i = 1, 2, ..., n\}$ from an external population having mean 0 and variance 1 and then generating bootstrap data by setting

$$y_i^* = \boldsymbol{x}_i^T \hat{\boldsymbol{\beta}} + t_i e_i, \quad i = 1, 2, \ldots, n, \tag{3.20}$$

where $\boldsymbol{x}_i$ is a $p \times 1$ deterministic vector, $\hat{\boldsymbol{\beta}}$ is the $p \times 1$ vector of least squares estimators of $\boldsymbol{\beta}$ and $e_i = y_i - \boldsymbol{x}_i^T \hat{\boldsymbol{\beta}}$. Liu (1988) suggested that another restriction needed to be imposed on t_i namely, $E(t_i^3) = 1$, to modify Wu's bootstrap procedure to share the usual second order asymptotic properties of the classical bootstrap. We begin by using the idea of the weighted bootstrap to construct a bootstrap procedure in the context of measurement error model. We describe below the weighted resampling algorithm for the slope parameter in the structural linear relationship (2.1)–(2.6).

1. Generate $D_i, i = 1, \ldots, n$; i.i.d. random variables with gamma distribution having density $g_D(x) = [p^q/(q-1)!]x^{q-1}e^{-px}I_{\{x>0\}}$, where $p = 2$ and $q = 4$.

2. Compute the bootstrap data, for $i = 1, \ldots, n$

$$\hat{\mu}_{02}^{*(i)} = \hat{\mu}_{02} + t_i[(Y_i - \bar{Y})^2 - \hat{\mu}_{02}], \tag{3.21}$$
$$\hat{\mu}_{20}^{*(i)} = \hat{\mu}_{20} + t_i[(X_i - \bar{X})^2 - \hat{\mu}_{20}], \tag{3.22}$$
$$\hat{\mu}_{11}^{*(i)} = \hat{\mu}_{11} + t_i[(X_i - \bar{X})(Y_i - \bar{Y}) - \hat{\mu}_{11}], \tag{3.23}$$

   where $t_i = D_i - E(D_i)$.

3. Obtain bootstrap estimates $\hat{\mu}_{20}^* = n^{-1}\sum_{i=1}^n \hat{\mu}_{20}^{*(i)}$, $\hat{\mu}_{02}^* = n^{-1}\sum_{i=1}^n \hat{\mu}_{02}^{*(i)}$ and $\hat{\mu}_{11}^* = n^{-1}\sum_{i=1}^n \hat{\mu}_{11}^{*(i)}$, computed from the bootstrap data $\hat{\mu}_{02}^{*(i)}$ $\hat{\mu}_{20}^{*(i)}$ and $\hat{\mu}_{11}^{*(i)}$, respectively.

4. Obtain bootstrap estimates $\hat{h}^*$ and $\hat{\beta}^*$ along the lines of $\hat{h}$ and $\hat{\beta}$.

5. Repeat steps (1)–(5) above a large number of times, B, to obtain $\hat{\beta}_1^*, \ldots, \hat{\beta}_B^*$.

6. Calculate estimate of standard error of $\hat{\beta}^*$ with

$$var(\hat{\beta}) = E_t(\hat{\beta}^* - E_t\hat{\beta}^*)^2, \tag{3.24}$$

   where E_t denotes expectation with respect to the weighted bootstrap sampling which can be approximated by

$$var_t(\hat{\beta}^*) = (B-1)^{-1}\sum_{b=1}^{B}(\hat{\beta}_b^* - \hat{\beta}_{(\cdot)})^2, \tag{3.25}$$

   where $\hat{\beta}_{(\cdot)}^* = \sum_{b=1}^B \hat{\beta}_b^*/B$.

7. In the b-th run, calculate the bootstrap-t, defined as

$$t_b^* = \frac{\hat{\beta}_b^* - \hat{\beta}}{\sqrt{var_t(\hat{\beta}_b^*)}}, \quad b = 1, \ldots, B, \tag{3.26}$$

where $var_t(\hat{\beta}_b^*)$ is the bootstrap variance applied to b-th bootstrap sample and given by

$$var_t(\hat{\beta}_b^*) = \frac{\hat{\beta}^2}{4n\hat{\mu}_{11}^2(\hat{h}^2 + \lambda^2)} \sum_{i=1}^{n} t_i^2 \hat{A}_i^2. \tag{3.27}$$

Asymptotic properties of the weighted bootstrap estimators are given in the next two theorems.

Theorem 3.3. *In the structural relationship (2.1)–(2.5), we assume* $E(X^6 + Y^6) < \infty$. *Then,* $n^{1/2}(\hat{\beta}^* - \hat{\beta})$ *converges in distribution to a normal random variable with zero mean and variance* $\hat{\Sigma}_{22}$, *where* $\hat{\beta}^*$ *is the bootstrap estimator of* $\hat{\beta}$ *resulting from the proposed resampling procedure in Section 3.2 and* $\hat{\Sigma}_{22}$ *is as defined in (2.28).*

Proof. Consider

$$\begin{aligned}
\hat{h}^* &= \hat{h} + \frac{1}{2\hat{\mu}_{11}^*}(\hat{\mu}_{02}^* - \lambda^2\hat{\mu}_{20}^*) - \frac{1}{2\hat{\mu}_{11}}(\hat{\mu}_{02} - \lambda^2\hat{\mu}_{20}) \\
&= \frac{1}{2\hat{\mu}_{11}^*}(\hat{\mu}_{02}^* - \lambda^2\hat{\mu}_{20}^* - 2\hat{h}\hat{\mu}_{11}^*) - \frac{1}{2\hat{\mu}_{11}}(\hat{\mu}_{02} - \lambda^2\hat{\mu}_{20} - 2\hat{h}\hat{\mu}_{11}),
\end{aligned}$$

or equivalently, we have

$$\begin{aligned}
\hat{h}^* - \hat{h} &= \frac{1}{2\hat{\mu}_{11}^*}(\hat{\mu}_{02}^* - \lambda^2\hat{\mu}_{20}^* - 2\hat{h}\hat{\mu}_{11}^*) - \frac{1}{2\hat{\mu}_{11}}(\hat{\mu}_{02} - \lambda^2\hat{\mu}_{20} - 2\hat{h}\hat{\mu}_{11}) \\
&= \frac{1}{2\hat{\mu}_{11}}\{(\hat{\mu}_{02}^* - \hat{\mu}_{02}) - \lambda^2(\hat{\mu}_{20}^* - \hat{\mu}_{20}) - 2\hat{h}(\hat{\mu}_{11}^* - \hat{\mu}_{11})\}\frac{\hat{\mu}_{11}}{\hat{\mu}_{11}^*} \\
&+ \frac{1}{2\hat{\mu}_{11}^*}(\hat{\mu}_{02} - \lambda^2\hat{\mu}_{20} - 2\hat{h}\hat{\mu}_{11}) - \frac{1}{2\hat{\mu}_{11}}(\hat{\mu}_{02} - \lambda^2\hat{\mu}_{20} - 2\hat{h}\hat{\mu}_{11}).
\end{aligned}$$

Since $\hat{\mu}_{11}^*$ is the mean of i.i.d samples, by the law of large numbers $\hat{\mu}_{11}^* \xrightarrow{p} \hat{\mu}_{11}$ and by Slutsky's theorem, we have

$$\sqrt{n}(\hat{h}^* - \hat{h}) \stackrel{d}{=} \frac{\sqrt{n}}{2\hat{\mu}_{11}}\{(\hat{\mu}_{02}^* - \hat{\mu}_{02}) - \lambda^2(\hat{\mu}_{20}^* - \hat{\mu}_{20}) - 2\hat{h}(\hat{\mu}_{11}^* - \hat{\mu}_{11})\} \tag{3.28}$$

and

$$E_t(\hat{\mu}^*_{02} - \hat{\mu}_{02})^2 = n^{-1}(\hat{\mu}_{04} - \hat{\mu}^2_{20}), \tag{3.29}$$
$$E_t(\hat{\mu}^*_{20} - \hat{\mu}_{20})^2 = n^{-1}(\hat{\mu}_{40} - \hat{\mu}^2_{02}), \tag{3.30}$$
$$E_t(\hat{\mu}^*_{11} - \hat{\mu}_{11}) = n^{-1}(\hat{\mu}_{22} - \hat{\mu}^2_{11}), \tag{3.31}$$
$$E_t(\hat{\mu}^*_{02} - \hat{\mu}_{02})(\hat{\mu}^*_{20} - \hat{\mu}_{20}) = n^{-1}(\hat{\mu}_{22} - \hat{\mu}_{02}\hat{\mu}_{20}), \tag{3.32}$$
$$E_t(\hat{\mu}^*_{02} - \hat{\mu}_{02})(\hat{\mu}^*_{11} - \hat{\mu}_{11}) = n^{-1}(\hat{\mu}_{13} - \hat{\mu}_{02}\hat{\mu}_{11}), \tag{3.33}$$
$$E_t(\hat{\mu}^*_{20} - \hat{\mu}_{20})(\hat{\mu}^*_{11} - \hat{\mu}_{11}) = n^{-1}(\hat{\mu}_{31} - \hat{\mu}_{20}\hat{\mu}_{11}). \tag{3.34}$$

We have

$$\sqrt{n}(\hat{h}^* - \hat{h}) \xrightarrow{d} N(0, \frac{1}{4\hat{\mu}^2_{11}}\hat{\sigma}^2_A), \tag{3.35}$$

with $\hat{\sigma}^2_A$ is given by (3.17). We conclude that $n^{1/2}(\hat{\beta}^* - \hat{\beta})$ converges in distribution to $N(0, \hat{\Sigma}_{22})$, where $\hat{\Sigma}_{22}$ is given by (2.28) by the delta method (see Bishop, Fienberg, and Holland, 1975). □

Theorem 3.4. *Assume the conditions of Theorem 3.3 hold. Then, under the proposed weighted sampling procedure described in Section 3.2,*

(i) $E_F[E_t(\hat{\beta}^* - \hat{\beta})] = E_F(\hat{\beta} - \beta) + O(n^{-2})$,

(ii) $E_F[E_t(\hat{\beta}^* - \hat{\beta})^2] = E_F(\hat{\beta} - \beta)^2 + O(n^{-2})$,

(iii) $E_F[E_t(\hat{\beta}^* - \hat{\beta})^3] = E_F(\hat{\beta} - \beta)^3 + O(n^{-3})$,

where E_t represents the expectation with respect to the distribution induced by bootstrap sampling described in Section 3.2.

Proof. See Appendix C. □

4 A simulation study

This section describes a simulation study which compares the performance of the proposed resampling methods to various other methods. Along the lines of simulation study done by Linder and Babu (1994), a total of 12 different cases are used to generate data sets from (2.1) and (2.2) with $n = 20, 30$, $\alpha = 1, \beta = 2$ and $\lambda^2 = 1$. The U_{1i}'s are generated according to the following four "design" distributions:

1. Uniform(1.5,8.5),
2. Normal(5,4),
3. (5-$\sqrt{2}$) + $\sqrt{2}W$, where W is chi-square(1),
4. $(2/3)N(4,2.5)$ + $(1/3)N(7,26)$ (mixture normal).

For each design, independent pairs of errors $(\delta_i, \varepsilon_i)$ are generated according to the following distributions:

1. $N(0, 0.48)$,
2. Double exponential(0.49) i.e. with density $f(\delta) = 1.02\exp(-|\delta|/0.49)$,
3. Contaminated normal: $0.1N(1.8, 0.84) + 0.9N(-0.2, 0.04)$
4. "Moderate" heteroscedastic normal: $(\delta_i, \varepsilon_i) \sim \{0.4N(0,1); i = 1, \ldots, 5\}$, $\{0.6N(0,1); i = 6, \ldots, 10\}$, $\{0.8N(0,1); i = 11, \ldots, 15\}$, $\{N(0,1); i = 16, ..., 20\}$ for $n = 20$ and $(\delta_i, \varepsilon_i) \sim \{0.4N(0,1); i = 1, \ldots, 6\}$, $\{0.6N(0,1); i = 7, \ldots, 12\}$, $\{0.8N(0,1); i = 13, \ldots, 19\}$, $\{N(0,1); i = 20, \ldots, 25\}$, $\{N(0,1); i = 26, \ldots, 30\}$ for $n = 30$.
5. "Heavy" heteroscedastic normal: $(\delta_i, \varepsilon_i) \sim \{N(0,1); i = 1, \ldots, 10\}$, $\{2.0N(0,1); i = 11, \ldots, 20\}$ for $n = 20$ and $(\delta_i, \varepsilon_i) \sim \{N(0,1); i = 1, \ldots, 15\}$, $\{2.0N(0,1); i = 16, \ldots, 30\}$ for $n = 30$.

For every case, Monte Carlo expectations are computed based on $N = 10,000$ simulations. Within each simulation, Monte Carlo expectations with respect to bootstrap are computed based on $B = 1,000$ bootstraps. The absolute bias of $\hat{\beta}$ and the confidence intervals for β are calculated as follows:

1. **Normal Approx. :** Absolute bias $= N^{-1}\sum_{n=1}^{N}|\hat{\beta}_n - \beta|$. Large sample confidence interval for the linear structural error model:

   $$\hat{\beta} \pm z_{1-\alpha/2}\sqrt{var_F(\hat{\beta})}$$

   where $z_{1-\alpha/2}$ is the $1-\alpha/2$ percentile of the standard normal distribution and $var(\hat{\beta}) = n^{-1}\hat{\Sigma}_{22}$ with $\hat{\Sigma}_{22}$ given by (2.28).

   For the next three bootstrap methods, the absolute bias of $\hat{\beta}$ is computed as $(NB)^{-1}\sum_{n=1}^{N}\sum_{b=1}^{B}|\hat{\beta}_{nb} - \beta|$.

2. **LINDER & BABU:** The $100(1-\alpha)$ % confidence interval for β is given by

$$\left[\hat{\beta} - t^{*(LB)}_{1-\alpha/2}\sqrt{var_F(\hat{\beta})}, \hat{\beta} - t^{*(LB)}_{\alpha/2}\sqrt{var_F(\hat{\beta})}\right],$$

where $t^{*(LB)}_{\alpha/2}$ and $t^{*(LB)}_{1-\alpha/2}$ are the percentiles of the histogram of the Studentized values

$$t^{*(LB)}_b = \frac{(\hat{\beta}^*_b - \hat{\beta})}{\sqrt{\hat{\Phi}^*(\hat{\beta}^*_b)}}, \quad b = 1, \ldots, B.$$

3. **PROPOSED METHOD 1:** The $100(1-\alpha)$ % confidence interval for β is given by

$$\left[\hat{\beta} - t^{*(1)}_{1-\alpha/2}\sqrt{var_F(\hat{\beta})}, \hat{\beta} - t^{*(1)}_{\alpha/2}\sqrt{var_F(\hat{\beta})}\right],$$

where $t^{*(1)}_{\alpha/2}$ and $t^{*(1)}_{1-\alpha/2}$ are the percentiles of the histogram of the Studentized values

$$t^{*(1)}_b = \frac{(\hat{\beta}^*_b - \hat{\beta})}{\sqrt{var^*_J(\hat{\beta}^*_b)}}, \quad b = 1, \ldots, B,$$

and $var^*_J(\hat{\beta}^*_b)$ is the jackknife variance estimator of $\hat{\beta}^*$ in the b-th bootstrap sample and given by (3.27).

4. **PROPOSED METHOD 2:** The $100(1-\alpha)$ % confidence interval for $\hat{\beta}$ is given by

$$\left[\hat{\beta} - t^{*(2)}_{1-\alpha/2}\sqrt{var_F(\hat{\beta})}, \hat{\beta} - t^{*(2)}_{\alpha/2}\sqrt{var_F(\hat{\beta})}\right],$$

where $t^{*(2)}_{\alpha/2}$ and $t^{*(2)}_{1-\alpha/2}$ are the percentiles of the histogram of the Studentized values

$$t^{*(2)}_b = \frac{(\hat{\beta}^*_b - \hat{\beta})}{\sqrt{var_t(\hat{\beta}^*_b)}}, \quad b = 1, \ldots, B,$$

and $var_t(\hat{\beta}^*_b)$ is the bootstrap variance estimator of $\hat{\beta}^*$ in the b-th bootstrap sample and given by (3.27).

For each of $N = 10,000$ simulations, we compute 90%, 95% and 99% confidence limits for β and their lower and upper tail frequencies (in percents). The tail frequencies represent tail probabilities and hence, the coverage probabilities of the confidence intervals. We also compute the confidence lengths (median). The following summaries are reported in Tables 1–14 and are calculated according to the above methods.

LOW: Error rate in the lower tail defined by

Normal Approx. : $\sum_{i=1}^{N} I_{L_{iN}}(\beta)/N$, where $L_{iN} = (\hat{\beta}_i - z_{\alpha/2}\sqrt{var_F(\hat{\beta}_i)}, \infty)$. The three bootstrap methods are given by $\sum_{i=1}^{N} I_{L_i}(\beta)/N$, where $L_i = (\hat{\beta}_i - t^*_{1-\alpha/2}\sqrt{var_F(\hat{\beta}_i)}, \infty)$ and $I_A(\cdot)$ in an indicator function defined by

$$I_A(x) = \begin{cases} 1 & x \in A \\ 0 & \text{otherwise.} \end{cases}$$

UP: Error rate in the upper tail defined by

Normal Approx. : $\sum_{i=1}^{N} I_{U_{iN}}(\beta)/N$, where $U_{iN} = (-\infty, \hat{\beta}_i + z_{\alpha/2}\sqrt{var_F(\hat{\beta}_i)})$. The three bootstrap methods are given by $\sum_{i=1}^{N} I_{U_i}(\beta)/N$, where $U_i = (-\infty, \hat{\beta}_i - t^*_{\alpha/2}\sqrt{var_F(\hat{\beta}_i)})$.

CP: The coverage probability defined by

Normal Approx. : $\sum_{i=1}^{N} I_{C_{iN}}(\beta)/N$, where $C_{iN} = (\hat{\beta}_i - z_{\alpha/2}\sqrt{var_F(\hat{\beta}_i)}, \hat{\beta}_i + z_{\alpha/2}\sqrt{var_F(\hat{\beta}_i)})$. The three bootstrap methods are given by $\sum_{i=1}^{N} I_{C_i}(\beta)/N$ where $C_i = (\beta \in (\hat{\beta}_i - t^*_{1-\alpha/2}\sqrt{var_F(\hat{\beta}_i)}, \hat{\beta}_i - t^*_{\alpha/2}\sqrt{var_F(\hat{\beta}_i)})$ with $t^*_{1-\alpha/2} = t^{*(LB)}_{1-\alpha/2}, t^{*(1)}_{1-\alpha/2}$ and $t^{*(2)}_{1-\alpha/2}$ and $t^*_{\alpha/2} = t^{*(LB)}_{\alpha/2}, t^{*(1)}_{\alpha/2}$ and $t^{*(2)}_{\alpha/2}$.

LGT: The median length of confidence intervals.

4.1 Summary of findings and conclusions

Table 15 reports the summary of the findings from Tables 1–14. For $n = 20$ the coverage probabilities for 90% are in the range of 81.23–87.55% for Normal Approx., 86–89.94% for Linder and Babu's method, 89.93-91.78% for proposed method 1 and 89.65-93.56% for proposed method 2. For 95%, they are in the range of 87.33–92.64% for Normal Approx., 92.86–95.16% for Linder and Babu's method, 94.98–97.32% for proposed method 1 and 93.60–97.03% for proposed method 2. For 99% coverage probability, they are in the range of 94.65–97.66% for Normal Approx., 98.22–99.07% for Linder and Babu's method, 98.88–99.33% for proposed method 1 and 97.72–99.54% for proposed method 2. Similarly, the coverage probabilities for $n = 30$ are in the same range.

The obvious conclusion to be drawn from Tables 11–14 is that the traditional large sample intervals are not corrected for the skewness of the distribution of $\hat{\beta}$. Its coverage probabilities are understated by their respective nominal rates. Our bootstrap procedures perform well throughout even in comparison with the normal theory estimates in normal situations, i.e., they have better coverage accuracy than the normal approximation. The tail errors rates show that all bootstrap methods result in heavier upper tail indicating a skewed distribution of $\hat{\beta}$ with a long tail to the left. This suggests that the use of bootstrap histograms to construct confidence interval is more appropriate. The weighted bootstrap tends to have inflated coverage probabilities and have long lengths, the reason being that the jackknife is not resistant to extreme values and perhaps data should be trimmed before jackknifing. It is not surprising that Linder and Babu's method does well here since they applied a correction factor in the bootstrap which makes their methods less appealing than ours, see remark on page 12. Another disadvantage of this method is that it is substantially more computer intensive than our proposed methods .

For the case of heteroscedastic errors, we present simulation results only for the case of uniform design with normal error distribution. These results appear to be robust against heteroscedascity and are enough to suggest that our methods are ahead compared to others in attaining their respective nominal coverage levels, as Tables 13–14 show.

Table 1: Comparison of tail coverage, coverage (%), length of confidence intervals and absolute bias of $\hat{\beta}$. The design is uniform with normal error distribution for N=10,000 simulations and B=1,000 bootstrap samples.

METHOD	n	$1-\alpha$	LOW	UP	CP	LGT	BIAS
NORMAL	20	0.90	4.43	8.76	86.81	0.539	0.139
APPROX.		0.95	2.39	5.76	91.85	0.642	0.139
		0.99	0.56	2.51	96.93	0.844	0.139
	30	0.90	4.76	7.59	87.65	0.441	0.112
		0.95	2.41	4.58	93.01	0.526	0.112
		0.99	0.55	2.04	97.41	0.691	0.112
LINDER	20	0.90	3.29	7.58	89.13	0.583	0.196
& BABU		0.95	1.70	3.75	94.55	0.718	0.196
		0.99	0.48	0.82	98.70	1.025	0.196
	30	0.90	3.03	8.24	88.73	0.455	0.163
		0.95	1.54	4.30	94.16	0.554	0.163
		0.99	0.39	1.00	98.61	0.769	0.163
PROPOSED	20	0.90	1.34	7.56	91.10	0.643	0.196
METHOD 1		0.95	0.55	3.82	95.63	0.794	0.196
		0.99	0.11	0.87	99.02	1.122	0.196
	30	0.90	1.29	8.64	90.07	0.492	0.163
		0.95	0.46	4.52	95.02	0.601	0.163
		0.99	0.07	1.20	98.73	0.833	0.163
PROPOSED	20	0.90	2.67	7.25	90.08	0.605	0.132
METHOD 2		0.95	1.17	4.79	94.04	0.714	0.132
		0.99	0.26	1.98	97.76	0.946	0.132
	30	0.90	3.19	6.92	89.89	0.474	0.107
		0.95	1.49	4.38	94.13	0.556	0.107
		0.99	0.30	1.92	97.78	0.723	0.107

Table 2: Comparison of tail coverage, coverage (%), length of confidence intervals and absolute bias of $\hat{\beta}$. The design is uniform with double exponential error distribution for N=10,000 simulations and B=1,000 bootstrap samples.

METHOD	n	$1-\alpha$	LOW	UP	CP	LGT	BIAS
NORMAL	20	0.90	5.00	8.57	86.43	0.526	0.143
APPROX.		0.95	2.46	5.43	92.11	0.627	0.143
		0.99	0.60	2.08	97.32	0.824	0.143
	30	0.90	4.71	7.39	87.90	0.435	0.115
		0.95	2.31	4.48	93.21	0.518	0.115
		0.99	0.46	1.53	98.01	0.681	0.115
LINDER	20	0.90	3.81	7.29	88.90	0.563	0.199
& BABU		0.95	1.87	3.42	94.71	0.686	0.199
		0.99	0.45	0.64	98.91	0.966	0.199
	30	0.90	3.17	7.86	88.97	0.446	0.165
		0.95	1.59	4.14	94.27	0.541	0.165
		0.99	0.41	0.69	98.90	0.740	0.165
PROPOSED	20	0.90	1.44	7.11	91.45	0.623	0.199
METHOD 1		0.95	0.60	3.23	96.17	0.761	0.199
		0.99	0.11	0.56	99.33	1.065	0.199
	30	0.90	1.25	8.00	90.75	0.486	0.165
		0.95	0.49	4.26	95.25	0.589	0.165
		0.99	0.11	0.78	99.11	0.807	0.165
PROPOSED	20	0.90	2.65	6.85	90.50	0.590	0.132
METHOD 2		0.95	1.20	4.18	94.62	0.694	0.132
		0.99	0.33	1.49	98.18	0.920	0.132
	30	0.90	3.08	6.51	90.41	0.469	0.108
		0.95	1.51	4.07	94.42	0.550	0.108
		0.99	0.22	1.29	98.49	0.714	0.108

Table 3: Comparison of tail coverage, coverage (%), length of confidence intervals and absolute bias of $\hat{\beta}$. The design is uniform with contaminated normal error distribution for N=10,000 simulations and B=1,000 bootstrap samples.

METHOD	n	$1-\alpha$	LOW	UP	CP	LGT	BIAS
NORMAL	20	0.90	6.72	6.99	86.29	0.069	0.018
APPROX.		0.95	3.82	4.17	92.01	0.082	0.018
		0.99	1.36	1.50	97.14	0.108	0.018
	30	0.90	6.74	6.49	86.77	0.058	0.015
		0.95	3.61	3.74	92.65	0.069	0.015
		0.99	1.05	1.28	97.67	0.091	0.015
LINDER	20	0.90	4.90	5.16	89.94	0.077	0.022
& BABU		0.95	2.49	2.35	95.16	0.094	0.022
		0.99	0.60	0.60	98.80	0.133	0.022
	30	0.90	5.15	5.29	89.56	0.063	0.019
		0.95	2.45	2.69	94.86	0.076	0.019
		0.99	0.65	0.59	98.76	0.105	0.019
PROPOSED	20	0.90	3.85	4.69	91.46	0.082	0.075
METHOD 1		0.95	1.95	2.10	95.95	0.100	0.075
		0.99	0.45	0.46	99.09	0.142	0.075
	30	0.90	4.33	4.91	90.76	0.065	0.070
		0.95	2.05	2.45	95.50	0.079	0.070
		0.99	0.43	0.55	99.02	0.109	0.070
PROPOSED	20	0.90	4.97	5.38	89.65	0.076	0.016
METHOD 2		0.95	3.05	3.35	93.60	0.089	0.016
		0.99	1.14	1.14	97.72	0.117	0.016
	30	0.90	5.26	5.59	89.15	0.062	0.014
		0.95	3.07	3.23	93.70	0.073	0.014
		0.99	0.93	1.17	97.90	0.095	0.014

Table 4: Comparison of tail coverage, coverage (%), length of confidence intervals and absolute bias of $\hat{\beta}$. The design is normal with normal error distribution for N=10,000 simulations and B=1,000 bootstrap samples.

METHOD	n	$1-\alpha$	LOW	UP	CP	LGT	BIAS
NORMAL	20	0.90	5.30	9.83	84.87	0.543	0.148
APPROX.		0.95	2.96	6.74	90.30	0.647	0.148
		0.99	0.83	3.15	96.02	0.850	0.148
	30	0.90	5.45	8.50	86.05	0.440	0.115
		0.95	2.89	5.57	91.54	0.525	0.115
		0.99	0.79	2.20	97.01	0.689	0.115
LINDER	20	0.90	3.41	7.63	88.96	0.628	0.207
& BABU		0.95	1.87	3.97	94.16	0.768	0.207
		0.99	0.54	0.76	98.70	1.080	0.207
	30	0.90	3.20	8.69	88.11	0.470	0.168
		0.95	1.78	4.40	93.82	0.575	0.168
		0.99	0.53	0.92	98.55	0.808	0.168
PROPOSED	20	0.90	1.43	7.75	90.82	0.674	0.207
METHOD 1		0.95	0.75	3.99	95.26	0.836	0.207
		0.99	0.12	0.81	99.07	1.192	0.207
	30	0.90	1.57	8.94	89.49	0.508	0.168
		0.95	0.63	4.75	94.62	0.623	0.168
		0.99	0.10	1.06	98.84	0.872	0.168
PROPOSED	20	0.90	2.59	7.21	90.20	0.659	0.141
METHOD 2		0.95	1.20	4.50	94.30	0.791	0.141
		0.99	0.22	1.52	98.26	1.110	0.141
	30	0.90	3.22	6.93	89.85	0.503	0.110
		0.95	1.40	4.29	94.31	0.597	0.110
		0.99	0.34	1.42	98.24	0.807	0.110

Table 5: Comparison of tail coverage, coverage (%), length of confidence intervals and absolute bias of $\hat{\beta}$. The design is normal with double exponential error distribution for N=10,000 simulations and B=1,000 bootstrap samples.

METHOD	n	$1-\alpha$	LOW	UP	CP	LGT	BIAS
NORMAL APPROX.	20	0.90	4.94	9.36	85.70	0.549	0.155
		0.95	2.45	6.09	91.46	0.655	0.155
		0.99	0.71	2.52	96.77	0.860	0.155
	30	0.90	5.12	7.73	87.15	0.439	0.116
		0.95	2.68	4.80	92.52	0.524	0.116
		0.99	0.59	1.61	97.80	0.688	0.116
LINDER & BABU	20	0.90	3.39	7.23	89.38	0.608	0.220
		0.95	1.65	3.62	94.73	0.750	0.220
		0.99	0.40	0.60	99.00	1.074	0.220
	30	0.90	3.35	8.03	88.62	0.462	0.172
		0.95	1.69	3.96	94.35	0.561	0.172
		0.99	0.44	0.72	98.84	0.778	0.172
PROPOSED METHOD 1	20	0.90	1.38	7.01	91.61	0.603	0.220
		0.95	0.65	3.50	95.85	0.733	0.220
		0.99	0.15	0.63	99.22	1.053	0.220
	30	0.90	1.39	8.22	90.39	0.502	0.172
		0.95	0.55	4.17	95.28	0.610	0.172
		0.99	0.10	0.71	99.19	0.844	0.172
PROPOSED METHOD 2	20	0.90	2.05	6.21	91.74	0.673	0.153
		0.95	0.98	3.64	95.38	0.810	0.153
		0.99	0.15	1.17	98.68	1.145	0.153
	30	0.90	2.69	5.81	91.50	0.501	0.112
		0.95	1.14	3.35	95.51	0.597	0.112
		0.99	0.23	0.99	98.78	0.802	0.112

Table 6: Comparison of tail coverage, coverage (%), length of confidence intervals and absolute bias of $\hat{\beta}$. The design is normal with contaminated normal error distribution for N=10,000 simulations and B=1,000 bootstrap samples.

METHOD	n	$1-\alpha$	LOW	UP	CP	LGT	BIAS
NORMAL	20	0.90	6.97	7.83	85.20	0.072	0.019
APPROX.		0.95	4.26	4.96	90.78	0.086	0.019
		0.99	1.47	2.01	96.52	0.113	0.019
	30	0.90	6.59	7.55	85.86	0.058	0.015
		0.95	3.89	4.54	91.57	0.070	0.015
		0.99	1.21	1.66	97.13	0.091	0.015
LINDER	20	0.90	4.71	5.42	89.87	0.084	0.018
& BABU		0.95	2.42	2.80	94.78	0.104	0.018
		0.99	0.58	0.51	98.91	0.149	0.018
	30	0.90	4.70	5.82	89.48	0.065	0.019
		0.95	2.25	2.87	94.88	0.080	0.019
		0.99	0.66	0.72	98.62	0.111	0.019
PROPOSED	20	0.90	3.74	4.75	91.51	0.090	0.050
METHOD 1		0.95	1.76	2.39	95.85	0.111	0.050
		0.99	0.34	0.46	99.20	0.161	0.050
	30	0.90	4.14	5.39	90.47	0.068	0.037
		0.95	1.93	2.60	95.47	0.083	0.037
		0.99	0.50	0.68	98.82	0.117	0.037
PROPOSED	20	0.90	4.60	5.32	90.08	0.085	0.018
METHOD 2		0.95	2.77	3.34	93.89	0.102	0.018
		0.99	0.72	1.05	98.23	0.139	0.018
	30	0.90	4.72	5.51	89.77	0.059	0.014
		0.95	2.63	3.09	94.28	0.071	0.014
		0.99	0.86	0.96	98.18	0.100	0.014

Table 7: Comparison of tail coverage, coverage (%), length of confidence intervals and absolute bias of $\hat{\beta}$. The design is chi-square with normal error distribution for N=10,000 simulations and B=1,000 bootstrap samples.

METHOD	n	$1-\alpha$	LOW	UP	CP	LGT	BIAS
NORMAL	20	0.90	7.68	9.89	82.43	0.364	0.106
APPROX.		0.95	4.87	7.00	88.13	0.434	0.106
		0.99	1.58	3.43	94.99	0.571	0.106
	30	0.90	7.10	9.62	83.28	0.319	0.090
		0.95	4.17	6.34	89.49	0.380	0.090
		0.99	1.40	2.91	95.69	0.499	0.090
LINDER	20	0.90	3.91	6.51	89.55	0.457	0.142
& BABU		0.95	1.94	3.35	94.71	0.565	0.142
		0.99	0.37	0.76	98.87	0.812	0.142
	30	0.90	3.33	7.67	89.00	0.376	0.126
		0.95	1.76	4.03	94.21	0.462	0.126
		0.99	0.39	0.92	98.69	0.648	0.126
PROPOSED	20	0.90	2.45	6.20	91.35	0.502	0.142
METHOD 1		0.95	1.06	3.15	97.32	0.629	0.142
		0.99	0.21	0.68	99.11	0.914	0.142
	30	0.90	2.31	7.74	89.95	0.404	0.126
		0.95	1.06	4.19	94.75	0.501	0.126
		0.99	0.25	1.03	98.72	0.709	0.126
PROPOSED	20	0.90	2.62	4.94	92.44	0.555	0.116
METHOD 2		0.95	1.13	2.67	96.20	0.706	0.116
		0.99	0.12	0.57	99.31	1.198	0.116
	30	0.90	2.78	5.12	92.10	0.442	0.091
		0.95	1.18	2.77	96.05	0.547	0.091
		0.99	0.15	0.65	99.20	0.828	0.091

Table 8: Comparison of tail coverage, coverage (%), length of confidence intervals and absolute bias of $\hat{\beta}$. The design is chi-square with double exponential error distribution for N=10,000 simulations and B=1,000 bootstrap samples.

METHOD	n	$1-\alpha$	LOW	UP	CP	LGT	BIAS
NORMAL APPROX.	20	0.90	7.29	9.83	82.88	0.345	0.102
		0.95	4.46	6.61	88.93	0.411	0.102
		0.99	1.55	3.00	95.45	0.540	0.102
	30	0.90	7.07	8.61	84.32	0.292	0.084
		0.95	4.12	5.49	90.39	0.348	0.084
		0.99	1.32	2.03	96.65	0.457	0.084
LINDER & BABU	20	0.90	4.04	6.43	89.53	0.419	0.136
		0.95	2.22	3.20	94.58	0.517	0.136
		0.99	0.59	0.65	98.76	0.742	0.136
	30	0.90	3.39	6.49	90.12	0.340	0. 117
		0.95	1.72	3.13	95.15	0.417	0. 117
		0.99	0.40	0.53	99.07	0.584	0. 117
PROPOSED METHOD 1	20	0.90	2.76	5.79	91.45	0.459	0.136
		0.95	1.29	2.95	95.76	0.574	0.136
		0.99	0.33	0.56	99.11	0.835	0.136
	30	0.90	2.66	6.20	91.14	0.361	0.117
		0.95	1.05	3.27	95.68	0.449	0.117
		0.99	0.23	0.59	99.18	0.642	0.117
PROPOSED METHOD 2	20	0.90	2.77	4.33	92.90	0.509	0.121
		0.95	1.39	2.42	96.19	0.639	0.121
		0.99	0.24	0.53	99.23	1.042	0.121
	30	0.90	2.73	3.90	93.37	0.418	0.091
		0.95	1.20	1.72	97.08	0.523	0.091
		0.99	0.15	0.26	99.59	0.812	0.091

Table 9: Comparison of tail coverage, coverage (%), length of confidence intervals and absolute bias of $\hat{\beta}$. The design is chisquare with contaminated normal error distribution for N=10,000 simulations and B=1,000 bootstrap samples.

METHOD	n	$1-\alpha$	LOW	UP	CP	LGT	BIAS
NORMAL	20	0.90	9.24	9.53	81.23	0.048	0.014
APPROX.		0.95	6.16	6.51	87.33	0.057	0.014
		0.99	2.52	2.83	94.65	0.075	0.014
	30	0.90	7.84	8.10	84.06	0.040	0.011
		0.95	5.04	4.87	90.09	0.047	0.011
		0.99	2.06	1.66	96.28	0.062	0.011
LINDER	20	0.90	4.96	5.13	89.91	0.063	0.018
& BABU		0.95	2.41	2.47	95.12	0.078	0.018
		0.99	0.51	0.42	99.07	0.111	0.018
	30	0.90	4.77	4.78	90.45	0.048	0.014
		0.95	2.49	2.07	95.44	0.059	0.014
		0.99	0.63	0.37	99.00	0.082	0.014
PROPOSED	20	0.90	3.91	4.58	91.51	0.069	0.049
METHOD 1		0.95	1.87	2.05	96.08	0.088	0.049
		0.99	0.39	0.32	99.29	0.128	0.049
	30	0.90	4.20	4.47	91.33	0.051	0.048
		0.95	2.04	1.95	96.01	0.036	0.048
		0.99	0.52	0.34	99.14	0.090	0.048
PROPOSED	20	0.90	3.76	3.96	92.28	0.076	0.016
METHOD 2		0.95	1.76	1.67	96.57	0.097	0.016
		0.99	0.26	0.20	99.54	0.169	0.016
	30	0.90	4.01	3.85	92.14	0.054	0.011
		0.95	2.14	1.81	96.05	0.067	0.011
		0.99	0.47	0.30	99.23	0.100	0.011

Table 10: Comparison of tail coverage, coverage (%), length of confidence intervals and absolute bias of $\hat{\beta}$. The design is mixture normal with normal error distribution for N=10,000 simulations and B=1,000 bootstrap samples.

METHOD	n	$1-\alpha$	LOW	UP	CP	LGT	BIAS
NORMAL	20	0.90	5.78	10.04	84.18	0.538	0.148
APPROX.		0.95	3.34	6.81	89.85	0.642	0.148
		0.99	1.00	3.20	95.80	0.843	0.148
	30	0.90	5.52	8.43	86.05	0.441	0.117
		0.95	2.88	5.59	91.53	0.526	0.117
		0.99	0.77	2.28	96.95	0.691	0.117
LINDER	20	0.90	3.65	7.77	88.85	0.616	0.208
& BABU		0.95	2.01	3.88	94.11	0.765	0.208
		0.99	0.49	0.74	98.77	1.111	0.208
	30	0.90	2.98	8.46	88.56	0.475	0.169
		0.95	1.61	4.35	94.04	0.581	0.169
		0.99	0.48	0.90	98.62	0.819	0.169
PROPOSED	20	0.90	1.77	7.81	90.42	0.680	0.208
METHOD 1		0.95	0.74	4.03	95.23	0.845	0.208
		0.99	0.14	0.73	99.13	1.210	0.208
	30	0.90	1.45	8.76	89.79	0.513	0.169
		0.95	0.64	4.60	94.76	0.632	0.169
		0.99	0.17	1.07	98.76	0.885	0.169
PROPOSED	20	0.90	2.65	6.84	90.51	0.675	0.142
METHOD 2		0.95	1.26	4.24	94.50	0.818	0.142
		0.99	0.22	1.27	98.51	1.178	0.142
	30	0.90	3.03	6.64	90.33	0.511	0.111
		0.95	1.40	4.00	94.60	0.610	0.111
		0.99	0.39	1.42	98.19	0.828	0.111

Table 11: Comparison of tail coverage, coverage (%), length of confidence intervals and absolute bias of $\hat{\beta}$. The design is mixture normal with error double exponential distribution for N=10,000 simulations and B=1,000 bootstrap samples.

METHOD	n	$1-\alpha$	LOW	UP	CP	LGT	BIAS
NORMAL APPROX.	20	0.90	5.90	9.43	84.67	0.519	0.147
		0.95	3.15	6.18	90.67	0.618	0.147
		0.99	0.69	2.48	96.83	0.813	0.147
	30	0.90	5.40	7.96	86.64	0.428	0.115
		0.95	2.87	4.61	92.52	0.510	0.115
		0.99	0.61	1.54	97.85	0.670	0.115
LINDER & BABU	20	0.90	3.85	7.12	89.03	0.581	0.207
		0.95	1.92	3.56	94.52	0.715	0.207
		0.99	0.31	0.71	98.98	1.027	0.207
	30	0.90	3.39	8.05	88.56	0.453	0.166
		0.95	1.83	3.68	94.49	0.550	0.166
		0.99	0.47	0.67	98.86	0.762	0.166
PROPOSED METHOD 1	20	0.90	1.74	6.92	91.34	0.639	0.207
		0.95	0.74	3.36	95.90	0.789	0.207
		0.99	0.07	0.60	99.33	1.126	0.207
	30	0.90	1.65	8.16	90.19	0.490	0.166
		0.95	0.71	3.77	95.52	0.597	0.166
		0.99	0.14	0.75	99.11	0.823	0.166
PROPOSED METHOD 2	20	0.90	2.48	6.05	91.47	0.648	0.144
		0.95	1.10	3.48	95.42	0.783	0.144
		0.99	0.20	1.02	98.78	1.125	0.144
	30	0.90	2.72	5.54	91.74	0.492	0.110
		0.95	1.15	3.15	95.70	0.586	0.110
		0.99	0.30	0.84	98.86	0.794	0.110

Table 12: Comparison of tail coverage, coverage (%), length of confidence intervals and absolute bias of β. The design is mixture normal with contaminated error distribution for N=10,000 simulations and B=1,000 bootstrap samples.

METHOD	n	$1-\alpha$	LOW	UP	CP	LGT	BIAS
NORMAL	20	0.90	7.32	7.80	84.88	0.071	0.019
APPROX.		0.95	4.35	4.93	90.72	0.084	0.019
		0.99	1.64	1.97	96.39	0.111	0.019
	30	0.90	6.15	6.58	87.27	0.061	0.016
		0.95	3.63	4.03	92.34	0.073	0.016
		0.99	1.40	1.36	97.24	0.095	0.016
LINDER	20	0.90	4.81	5.25	89.94	0.083	0.024
& BABU		0.95	2.32	2.69	94.99	0.102	0.024
		0.99	0.52	0.53	98.95	0.147	0.024
	30	0.90	4.42	5.20	90.38	0.068	0.020
		0.95	2.36	2.60	95.04	0.083	0.020
		0.99	0.59	0.50	98.91	0.115	0.020
PROPOSED	20	0.90	3.61	4.61	91.78	0.088	0.093
METHOD 1		0.95	1.83	2.17	96.00	0.110	0.093
		0.99	0.31	0.42	99.27	0.159	0.093
	30	0.90	3.83	4.88	91.29	0.070	0.066
		0.95	1.95	2.37	95.68	0.086	0.066
		0.99	0.44	0.43	99.13	0.121	0.066
PROPOSED	20	0.90	4.66	5.26	90.08	0.085	0.017
METHOD 2		0.95	2.68	3.05	94.27	0.101	0.017
		0.99	0.85	0.99	98.16	0.139	0.017
	30	0.90	4.66	4.96	90.38	0.069	0.015
		0.95	2.60	2.82	94.58	0.081	0.015
		0.99	0.83	0.87	98.30	0.109	0.015

Table 13: Comparison of tail coverage, coverage (%), length of confidence intervals and absolute bias of β. The design is uniform with "moderate" heteroscedastic normal error distribution for N=10,000 simulations and B=1,000 bootstrap samples.

METHOD	n	$1-\alpha$	LOW	UP	CP	MED-LGT	BIAS
NORMAL APPROX.	20	0.90	4.72	7.73	87.55	0.395	0.102
		0.95	2.45	4.91	92.64	0.471	0.102
		0.99	1.01	2.40	97.66	0.619	0.102
	30	0.90	5.59	7.72	86.69	0.343	0.090
		0.95	2.91	4.80	92.29	0.408	0.090
		0.99	0.68	1.78	97.54	0.537	0.090
LINDER & BABU	20	0.90	4.20	7.35	88.45	0.418	0.131
		0.95	2.24	3.81	93.85	0.511	0.131
		0.99	0.66	0.88	98.46	0.715	0.131
	30	0.90	3.67	7.71	88.62	0.362	0.100
		0.95	1.99	3.98	94.03	0.439	0.100
		0.99	0.56	1.05	98.39	0.601	0.100
PROPOSED METHOD 1	20	0.90	2.45	7.07	90.48	0.452	0.131
		0.95	1.05	3.70	95.25	0.554	0.131
		0.99	0.28	0.84	98.88	0.776	0.131
	30	0.90	2.29	7.77	89.94	0.382	0.100
		0.95	1.05	4.14	94.81	0.463	0.100
		0.99	0.15	1.13	98.72	0.636	0.100
PROPOSED METHOD 2	20	0.90	4.39	2.05	93.56	0.422	0.093
		0.95	2.12	0.85	97.03	0.494	0.093
		0.99	0.49	0.23	99.28	0.648	0.093
	30	0.90	3.82	6.82	89.36	0.367	0.084
		0.95	2.12	4.22	93.66	0.431	0.084
		0.99	0.42	1.55	98.03	0.559	0.084

Table 14: Comparison of tail coverage, coverage (%), length of confidence intervals and absolute bias of β. The design is uniform with "heavy" heteroscedastic normal error distribution for N=10,000 simulations and B=1,000 bootstrap samples.

METHOD	n	$1-\alpha$	LOW	UP	CP	MED-LGT	BIAS
NORMAL	20	0.90	4.19	9.94	85.87	0.847	0.234
APPROX.		0.95	2.23	6.82	90.95	1.009	0.234
		0.99	0.38	3.03	96.59	1.326	0.234
	30	0.90	3.80	9.00	87.20	0.685	0.180
		0.95	1.58	6.08	92.34	0.817	0.180
		0.99	0.21	2.51	97.28	1.073	0.180
LINDER	20	0.90	4.35	8.79	86.86	0.862	0.402
& BABU		0.95	2.60	4.54	92.86	1.060	0.402
		0.99	0.85	0.93	98.22	1.520	0.402
	30	0.90	2.63	10.23	87.14	0.680	0.236
		0.95	1.49	5.65	92.86	0.827	0.236
		0.99	0.39	1.37	98.24	1.140	0.236
PROPOSED	20	0.90	1.08	8.99	89.93	0.984	0.397
METHOD 1		0.95	0.45	4.57	94.98	1.199	0.397
		0.99	0.07	0.77	99.16	1.669	0.397
	30	0.90	0.50	10.69	88.81	0.761	0.236
		0.95	0.17	5.97	93.86	0.924	0.236
		0.99	0.01	1.47	98.52	1.264	0.236
PROPOSED	20	0.90	1.08	8.99	89.93	0.975	0.227
METHOD 2		0.95	0.45	4.57	94.98	1.159	0.227
		0.99	0.07	0.77	99.16	1.571	0.227
	30	0.90	1.90	8.21	89.89	0.743	0.173
		0.95	0.66	5.46	93.88	0.876	0.173
		0.99	0.06	2.23	97.71	1.149	0.173

Table 15: Summary of coverage (%), length of confidence intervals and absolute bias of $\hat{\beta}$ for Tables 4.1–4.14.

METHOD	n	$1-\alpha$	CP		LGT		BIAS	
			MIN	MAX	MIN	MAX	MIN	MAX
NORMAL APPROX.	20	0.90	81.23	87.55	0.048	0.847	0.014	0.234
		0.95	87.33	92.64	0.057	1.009	0.014	0.234
		0.99	94.65	97.66	0.075	1.326	0.014	0.234
	30	0.90	83.28	87.90	0.040	0.685	0.010	0.180
		0.95	89.49	93.21	0.047	0.817	0.010	0.180
		0.99	95.69	98.01	0.062	1.073	0.010	0.180
LINDER & BABU	20	0.90	86.86	89.94	0.063	0.862	0.180	0.402
		0.95	92.86	95.16	0.078	1.060	0.180	0.402
		0.99	98.22	99.07	0.111	1.520	0.180	0.402
	30	0.90	87.14	90.45	0.480	0.680	0.140	0.236
		0.95	92.86	95.44	0.059	0.827	0.140	0.236
		0.99	98.24	99.07	0.082	1.140	0.140	0.236
PROPOSED METHOD 1	20	0.90	89.93	91.78	0.069	0.984	0.490	0.397
		0.95	94.98	97.32	0.088	1.199	0.490	0.397
		0.99	98.88	99.33	0.128	1.669	0.490	0.397
	30	0.90	88.81	91.33	0.051	0.761	0.037	0.236
		0.95	93.86	96.01	0.036	0.924	0.037	0.236
		0.99	98.52	99.19	0.090	1.264	0.037	0.236
PROPOSED METHOD 2	20	0.90	89.65	93.56	0.076	0.975	0.016	0.227
		0.95	93.60	97.03	0.089	1.159	0.016	0.227
		0.99	97.72	99.54	0.117	1.571	0.016	0.227
	30	0.90	89.15	93.37	0.054	0.743	0.011	0.173
		0.95	93.66	97.08	0.067	0.876	0.011	0.173
		0.99	97.71	99.59	0.095	1.149	0.011	0.173

References

Anderson, T. W., and Sawa, T. (1982). Exact and approximate distributions of the maximum likelihood estimators of a slope coefficient. *J. Royal Statist. Soc., Ser B*, **44**, 52–62.

Arvesen, J. N. (1969). Jackknifing U-statistics. *Ann. Statist.*, **40**, 2076–2100.

Babu, G. J., and Bai, Z. (1992). Edgeworth expansions for errors-in-variables models. *J. Multivariate Anal.*, **42**, 226–224.

Barnett, V. D. (1967). A note on linear structural relationships when both residual variances are known. *Biometrika*, **54**, 670–672.

Birch, M. (1964). A note on the maximum likelihood estimation of a linear structural relationship. *J. Amer. Statist. Assoc.*, **59**, 1175–1178.

Bishop, Y. M. M., Fienberg, S. E., and Holland, P. W. (1975). *Discrete Multivariate Analysis: Theory and Practice*, MIT Press, Cambridge.

Chan, N. N., and Mak, T. K. (1983). Estimation of multivariate linear functional relationships. *Biometrika*, **70**, 263–267.

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

Gleser, L. J. (1981). Estimation in a multivariate "error in variables" regression model: large sample results. *Ann. Statist.*, **9**, 24–44.

Kelly, G. (1984). The influence function in the errors in variable problem. *Ann. Statist.*, **12**, 87–100.

Kendall, M. G., and Stuart, A. (1979). *The Advanced Theory of Statistics*, 4th edition, volume 2, Griffin, London.

Linder, E., and Babu, G. J. (1994). Bootstrapping the linear functional relationship with known error variance ratio. *Scand. J. Statist.*, **21**, 21–39.

Lindley, D. V., and El-Sayyad, G. M. (1968). The Bayesian estimation of a linear functional relationship. *J. R. Statist. Soc. B*, **30**, 190–202.

Liu, R. Y. (1988). Bootstrap procedure under some non-i.i.d. models. *Ann. Statist.*, **14**, 1697–1708.

Riersøl, O. (1950). Identifiability of a linear relation between variables which are subject to error. *Econometrica*, **18**, 375–389.

Solari, M. E. (1969). The maximum likelihood solution of the problem of estimating a linear functional relationship. *J. R. Statist. Soc. B*, **31**, 372–375.

Wu, C. F. J. (1986). Jackknife, bootstrap and other resampling methods in regression analysis. *Ann. Statist.*, **14**, 1261–1350.

Zellner, A. (1971). *An Introduction to Bayesian Inference in Econometrics*, John Wiley and Sons, New York.