Research Paper ∎

# Bootstrap in Errors-in-Variables Regressions Applied to Methods Comparison Studies

**Bernard G. Francq**

**Abstract.** In method comparison studies, the measurements taken by two methods are compared to assess whether they are equivalent. If there is no analytical bias between the methods, they should provide the same results on average notwithstanding the measurement errors. This equivalence can be assessed with regression techniques by taking into account the measurement errors. Among them, the paper focuses on Deming Regression (DR) and Bivariate Least-Squares regression (BLS). The confidence intervals (CI's) of the regression parameters are useful to assess the presence or absence of bias. These CI's computed by errors-in-variables regressions are approximate (except the one for slope estimated by DR), which leads to coverage probabilities lower than the nominal value. Six bootstrap approaches and the jackknife are assessed in the paper as means to improve the coverage probabilities of the CI's.

# Uporaba zankanja v regresiji za spremenljivke z merskimi napakami v študijah primerjave metod

**Izvleček.** V študijah primerjave metod primerjamo meritve z dvema metodama, da bi ocenili, ali sta ekvivalentni. Če nobena od metod ni pristranska, moramo z njima v povprečju dobiti enake rezultate ne glede na napake merjenja. Tovrstno ekvivalentnost lahko preverjamo z regresijskimi pristopi, ki upoštevajo merske napake. Prispevek se osredotoča na Demingovo regresijo (DR) in bivariatno regresijo po metodi najmanjših kvadratov. Z intervali zaupanja (IZ) za regresijske parametre lahko ocenimo, ali je prisotna pristranost. IZ so pri regresiji za spremenljivke z merskimi napakami le približni (razen za ocenjeni naklon pri DR), zato je dejanska stopnja zaupanja nižja od deklarirane. Prispevek primerja šest oblik zankanja in metodo pipca kot pristope za izboljšanje ustreznosti stopnje zaupanja IZ.

Institucija avtorja / Author's institution: Université Catholique de Louvain, Institut de Statistique, Biostatistique et sciences Actuarielles, Louvain, Belgium.

Kontaktna oseba / Contact person: Bernard G. Francq, ISBA, Voie du Roman Pays 20 bte L1.04.01, B-1348 Louvain-la-Neuve. e-pošta / e-mail: bernard.g.francq@uclouvain.be.

## Introduction

The needs of the industries and laboratories to quickly assess the quality of products or samples leads to the development and improvement of new measurement methods that should be faster, easier to handle, less expensive or more accurate than the reference method. These alternative methods should ideally lead to results comparable to those obtained by a standard method [1]. Ideally, there should be no bias between the two methods, i.e., the measurement methods should be interchangeable.

Different approaches are proposed in the literature to deal with method comparison studies. The most widely known and used is the approach proposed by Bland and Altman, which focuses directly on the differences between two measurement methods [2-4]. The approach based on regression analysis (a linear functional relationship [5]) is also widely applied; it focuses on the parameter estimates and their confidence intervals (CI's) [6]. This paper deals with the regression approach. In order to statistically test the equivalence between two measurement methods, a certain characteristic of a sample can be measured by the two methods in the experimental domain of interest. The pairs of measurements taken by the reference method and the alternative one can be modelled by a regression line and the parameter estimates used to test the equivalence. Obtaining an intercept significantly different from zero in such regression indicates a systematic analytical bias between the methods, and a slope significantly different from one indicates a proportional bias [6]. To perform the regression correctly it is essential to take into account the errors in both variables (i.e., dimensions, axes) and the heteroskedasticity if necessary [6]. Various types of regressions exist to tackle this problem [7]; this paper focuses on the Deming Regression (DR) and Bivariate Least Square (BLS), as well as the basic Ordinary Least Square (OLS) regression.

It is known that the coverage probabilities of the approximate confidence intervals computed by DR

or BLS can be lower than the nominal level especially when the ratio of the measurement errors' variances is lower than one. In the paper, different bootstrap procedures are briefly explained and assessed with simulations in order to improve these coverage probabilities and thus obtain more precise confidence intervals. The systolic blood pressure data set published by Bland and Altman [2] is used to illustrate these techniques.

## How to test the equivalence?

In the systolic blood pressure data [2], simultaneous measurements were made using a sphygmomanometer and a semi-automatic blood pressure monitor. The Bland and Altman approach focuses on "practical" equivalence to assess whether the observed differences between the two measurement methods are meaningful or not in practice. The present paper focuses on "strict" or "statistical" equivalence. The bias between the two devices is considered because the two devices should provide equal (equivalent) measures notwithstanding the errors of measurement.

The standard design in method comparison studies is to measure each specimen/subject once using both devices/methods. However, with such design it is not possible to estimate the variances of measurement errors, as explained below.

### The general model

To compare two measurement methods, a parameter of interest is measured on $N$ sampling units ($i = 1, 2, \ldots, N$) by both methods [10-12]:
$$X_{ij} = \xi_i + \tau_{ij}; Y_{ij} = \eta_i + \nu_{ij}, \qquad (1)$$
where $X_{ij}(j = 1, 2, \ldots, n_{X_i})$ and $Y_{ij}(j = 1, 2, \ldots, n_{Y_i})$ are the repeated measures for unit $i$ by methods $X$ and $Y$, respectively, and $n_{X_i}$ and $n_{Y_i}$ are the number of repeated measures of unit $i$ by each method. $\xi_i$ and $\eta_i$ are the true but unobservable values of the parameter of interest for both methods, which are assumed to be linked

by a linear relationship [10-12]:

$$\eta_i = \alpha + \beta \xi_i. \tag{2}$$

The means of the repeated measures for the unit $i$ are given by $X_i$ and $Y_i$:

$$X_i = \frac{1}{n_{X_i}} \sum_{j=1}^{n_{X_i}} X_{ij} \text{ and } Y_i = \frac{1}{n_{Y_i}} \sum_{j=1}^{n_{Y_i}} Y_{ij}; \tag{3}$$

$\tau_{ij}$ and $\nu_{ij}$ are the measurement errors, which are supposed to be independent and normally distributed (with constant variances under homoskedasticity):

$$\begin{pmatrix} \tau_{ij} \\ \nu_{ij} \end{pmatrix} \sim N \left( \begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_{\tau_i}^2 & 0 \\ 0 & \sigma_{\nu_i}^2 \end{pmatrix} \right). \tag{4}$$

Hence, the means of the repeated measures are also normally distributed around $\xi_i$ or $\eta_i$:

$$\begin{pmatrix} X_i \\ Y_i \end{pmatrix} \sim N \left( \begin{pmatrix} \xi_i \\ \eta_i \end{pmatrix}, \begin{pmatrix} \sigma_{\tau_i}^2/n_{X_i} & 0 \\ 0 & \sigma_{\nu_i}^2/n_{Y_i} \end{pmatrix} \right). \tag{5}$$

If the variances $\sigma_{\tau_i}^2$ and $\sigma_{\nu_i}^2$ are unknown, they can be estimated with repeated measures; otherwise, these variances are unknown and inestimable. The estimates of $\sigma_{\tau_i}^2$ and $\sigma_{\nu_i}^2$ are given by $s_{\tau_i}^2$ and $s_{\nu_i}^2$:

$$s_{\tau_i}^2 = \frac{1}{n_{X_i}-1} \sum_{j=1}^{n_{X_i}} \left( X_{ij} - X_i \right)^2 \text{ and}$$

$$s_{\nu_i}^2 = \frac{1}{n_{Y_i}-1} \sum_{j=1}^{n_{Y_i}} \left( Y_{ij} - Y_i \right)^2. \tag{6}$$

In further explanations, the following notation will also be used:

$$\bar{X} = \frac{1}{N} \sum_{i=1}^{N} X_i \text{ and } \bar{Y} = \frac{1}{N} \sum_{i=1}^{N} Y_i;$$

$$s_{xx} = \sum_{i=1}^{N} (X_i - \bar{X})^2; \; S_{yy} = \sum_{i=1}^{N} (Y_i - \bar{Y})^2 \text{ and}$$

$$s_{xy} = \sum_{i=1}^{N} (X_i - \bar{X})(Y_i - \bar{Y}).$$

## The homoskedastic model

Under homoskedasticity, the measurement errors variances are constant through the domain of interest $\left( \sigma_{\tau_i}^2 = \sigma_\tau^2 \text{ and } \sigma_{\nu_i}^2 = \sigma_\nu^2 \; \forall i \right)$. Moreover, a constant number of replicates will be assumed $(n_{X_i} = n_X \text{ and } n_{Y_i} = n_Y \; \forall i)$ to prevent the model from becoming heteroskedastic even if the accuracies of the measurement methods are constant. Under homoskedasticity, the variances

$s_{\tau_i}^2$ and $s_{\nu_i}^2$ are estimates of $\sigma_\tau^2$ and $\sigma_\nu^2$ and the "overall" estimates for $\sigma_\tau^2$ and $\sigma_\nu^2$ are given by $s_\tau^2$ and $s_\nu^2$:

$$s_\tau^2 = \frac{\sum_{i=1}^{N}(n_{X_i}-1)s_{\tau_i}^2}{\left(\sum_{i=1}^{N} n_{X_i}\right)-N} \text{ and } s_\nu^2 = \frac{\sum_{i=1}^{N}(n_{Y_i}-1)s_{\nu_i}^2}{\left(\sum_{i=1}^{N} n_{Y_i}\right)-N}, \tag{7}$$

or with constant repeated measures:

$$s_\tau^2 = \frac{\sum_{i=1}^{N} s_{\tau_i}^2}{N} \text{ and } s_\nu^2 = \frac{\sum_{i=1}^{N} s_{\nu_i}^2}{N}. \tag{8}$$

## How to test the equivalence?

If the two measurement methods are equivalent, they should give the same results for a given sample notwithstanding the measurement errors. In the model notation, method equivalence means that $\xi_i = \eta_i \; \forall i$ [6,13]. In practice, due to the measurement errors, these parameters are unobservable and the equivalence test will be based on the following regression model:

$$Y_i = \alpha + \beta X_i + \varepsilon_i \text{ with } \varepsilon_i \sim N(0, \sigma_{\varepsilon_i}^2) \text{ and}$$

$$\sigma_{\varepsilon_i}^2 = \frac{\sigma_{\nu_i}^2}{n_{Y_i}} + \beta^2 \frac{\sigma_{\tau_i}^2}{n_{X_i}}, \tag{9}$$

where the intercept $\alpha$ and the slope $\beta$ are estimated respectively by $\hat{\alpha}$ and $\hat{\beta}$. This regression model is applied on the averages of repeated measures because individual measures cannot be paired.

The estimated parameters $\hat{\alpha}$ and $\hat{\beta}$ provide the information to assess the equivalence. An intercept significantly different from 0 means that there is a constant bias between the two measurement methods, and a slope significantly different from 1 means that there is a proportional bias between the two measurement methods [6]. Therefore, the following two-sided hypothesis will be used to test method equivalence:

$$H_o: \alpha = 0 \; ; H_1: \alpha \neq 0 \text{ and}$$

$$H_o: \beta = 1 \; ; H_1: \beta \neq 1. \tag{10}$$

The null hypothesis $H_o: \alpha = 0$ is rejected if 0 is not included in the confidence interval (CI) for $\alpha$ and the null hypothesis $H_o: \beta = 1$ is rejected if 1 is not included in the CI for $\beta$. The joint CI is not considered in this paper.

# OLS regression versus errors-in-variables regressions

This section briefly reviews the formulas for the estimation of a regression line by means of the commonly used Ordinary Least Squares (OLS) regression when $X$ is observed without errors. Next, the formulas for two errors-in-variables regressions are provided – the Deming Regression and the Bivariate Least Squares regression. Note that in practice $\sigma_\tau^2$ or/and $\sigma_\nu^2$ can be estimated with replicated data and replaced by $s_\tau^2$ or/and $s_\nu^2$ if needed.

## Ordinary Least Squares (OLS) regression

The easiest way to estimate the parameters $\alpha$ and $\beta$ of model (9) under homoskedasticity is to apply the basic technique of OLS [12-13]. The OLS regression minimises the sum of squared vertical distances (residuals) between each point and the line as shown in Figure 1. The corresponding parameter estimators are given by the following formulas:

$$\hat{\beta}_{OLS} = \frac{s_{xy}}{s_{xx}} \text{ and } \hat{\alpha}_{OLS} = \bar{Y} - \hat{\beta}_{OLS}\bar{X}. \tag{11}$$
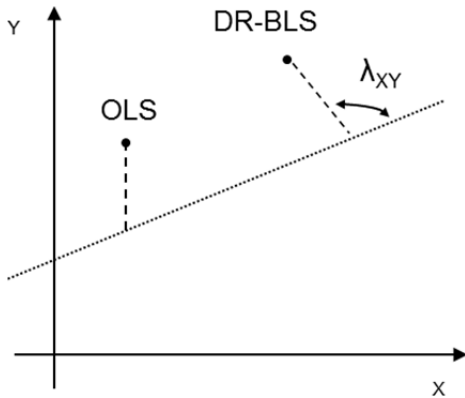


**Figure 1** Illustration of OLS and DR-BLS regressions criteria of minimisation.

Unfortunately, the OLS minimisation criterion does not take into account the errors in the independent variable [14]. OLS supposes that there is no error produced by the measurement method assigned to the X-axis, i.e., the $\tau_{ij}$ are supposed to be equal to zero or negligible. The corresponding estimates are therefore obviously biased [14].

Supposing that $\sigma_\tau^2 = 0$, the $100(1–\gamma)\%$ CI for $\beta$ is symmetric around $\hat{\beta}_{OLS}$ and is computed as [15]

$$CI(\beta_{OLS}): \hat{\beta}_{OLS} \pm t_{1-\frac{\gamma}{2};N-2}s_{\hat{\beta}_{OLS}} \tag{12}$$

with $s_{\hat{\beta}_{OLS}} = \sqrt{\frac{s_{OLS}^2}{s_{xx}}}$ and

$$s_{OLS}^2 = \frac{1}{N-2}\sum_{i=1}^{N}\left(Y_i - \hat{\alpha}_{OLS} - \hat{\beta}_{OLS}X_i\right)^2, \tag{13}$$

where $t_{1-\gamma/2;N-2}$ is the $100(1–\gamma/2)\%$ percentile of a $t$-distribution with $N-2$ degrees of freedom.

In the same way, the $100(1–\gamma)\%$ CI for $\alpha$ is symmetric around $\hat{\alpha}_{OLS}$ and can be computed as $CI(\alpha_{OLS}): \hat{\alpha}_{OLS} \pm t_{1-\frac{\gamma}{2};N-2}s_{\hat{\alpha}_{OLS}}$ with

$$s_{\hat{\alpha}_{OLS}} = \sqrt{s_{OLS}^2\left(\frac{1}{N} + \frac{\bar{X}^2}{s_{xx}}\right)}. \tag{14}$$

These CI's are exact under the assumptions of OLS, especially that of no errors in the X-values ($\sigma_\tau^2 = 0$) and normality of $\varepsilon_i$.

## Deming Regression (DR)

To take into account the errors in both variables, the following ratio between the two error variances can be computed:

$$\lambda_{XY} = \frac{\sigma_\nu^2/n_Y}{\sigma_\tau^2/n_X}. \tag{15}$$

It is the ratio of the errors' variance in the Y over the errors' variance in X.

The DR is the Maximum Likelihood (ML) solution of model (1) when $\lambda_{XY}$ is known [10]. In practice, $\lambda_{XY}$ can be estimated with replicated data.

The DR minimises the sum of the (weighted) squares of the oblique distances between each point to the line [11,16] as shown in Figure 1. The angle of the direction is related to $\lambda_{XY}$ and given by $-\lambda_{XY}/\hat{\beta}$ [11]. The ML estimators are:

$$\hat{\beta}_{DR} = \frac{s_{yy} - \lambda_{XY}s_{xx} + \sqrt{\left(s_{yy} - \lambda_{XY}s_{xx}\right)^2 + 4\lambda_{XY}s_{xy}^2}}{2s_{xy}} \text{ and }$$

$$\hat{\alpha}_{DR} = \bar{Y} - \hat{\beta}_{DR}\bar{X}. \tag{16}$$

The ratio $\lambda_{XY}$ is assumed to be constant by DR. This assumption is fulfilled under homoskedasticity and balanced design ($n_{X_i}$ and $n_{Y_i}$ constant).

Gillard and Iles [17-18] propose to compute the variance-covariance matrix of the estimators using the method of moments. When $\lambda_{XY}$ is assumed to be known, the variances of the estimators can be computed with the following formulas (modified to take into account the replicated data):

$$s_{\hat{\beta}_{DR}}^2 = \frac{s_{xx}s_{yy}-s_{xy}^2}{N\left(\frac{s_{xy}}{\hat{\beta}_{DR}}\right)^2},$$

$$s_{\hat{\alpha}_{DR}}^2 = \bar{X}^2 s_{\hat{\beta}_{DR}}^2 + \frac{\hat{\beta}_{DR}^2 \sigma_\tau^2/n_X + \sigma_v^2/n_Y}{N}. \qquad (17)$$

The approximate and symmetric CI for $\beta$ or $\alpha$ can be easily computed by associating a $t$-distribution to the standard error of the parameter because the estimators provided by ML are asymptotically normally distributed [19]:

$$\text{CI}(\beta_{DR}): \hat{\beta}_{DR} \pm t_{1-\frac{\gamma}{2};N-2}s_{\hat{\beta}} \text{ and}$$

$$\text{CI}(\alpha_{DLR}): \hat{\alpha}_{DR} \pm t_{1-\frac{\gamma}{2};N-2}s_{\hat{\alpha}}. \qquad (18)$$

For the slope $\beta$, an exact solution exists – the exact and asymmetric CI for $\beta$ can be computed as follows [11]:

$$\text{Exact-CI}(\beta_{DR}): \sqrt{\lambda_{XY}}\tan(\hat{\theta}_s \pm \phi) \text{ where}$$

$$\text{CI}(\theta_s): \hat{\theta}_s \pm \phi \qquad (19)$$

with $\hat{\beta}_{DR} = \tan\hat{\theta}$, $\hat{\theta} = \arctan\hat{\beta}_{DR}$,

$$\hat{\theta}_s = \arctan(\hat{\beta}_{DR}/\sqrt{\lambda_{XY}}) \text{ and} \qquad (20)$$

$$\phi =$$

$$\frac{1}{2}\arcsin\left( t_{1-\frac{\gamma}{2};N-2}\frac{2}{\sqrt{N-2}} * \right.$$

$$\left. \sqrt{\frac{\lambda_{XY}(s_{xx}s_{yy}-s_{xy}^2)}{\left(s_{yy}-\lambda_{XY}s_{xx}\right)^2+4\lambda_{XY}s_{xy}^2}} \right) \qquad (21)$$

**Bivariate Least Square regression (BLS)**

The BLS is a generic name but this paper refers to BLS as defined first by Lisý *et al.* [20] and later by other authors [6,21-23]. The BLS can take into

account error and heteroskedasticity in both variables and is usually explained in matrix notation [6,21-23]. Here, the formulas are given under homoskedaticity with replicated data. The estimates of the parameters (the **b** vector) are computed by iteration using the following formulas:

$$\mathbf{R}\hat{\mathbf{b}} = \mathbf{g} \qquad (22)$$

$$\frac{1}{w_{BLS}}\begin{pmatrix} N & \sum_{i=1}^{N}X_i \\ \sum_{i=1}^{N}X_i & \sum_{i=1}^{N}X_i^2 \end{pmatrix}\begin{pmatrix} \hat{\alpha}_{BLS} \\ \hat{\beta}_{BLS} \end{pmatrix} = \frac{1}{w_{BLS}} *$$

$$\begin{pmatrix} \sum_{i=1}^{N}Y_i \\ \sum_{i=1}^{N}\left(X_iY_i + \hat{\beta}_{BLS}\frac{\sigma_\tau^2}{n_X}\frac{(Y_i-\hat{\alpha}_{BLS}-\hat{\beta}_{BLS}X_i)^2}{w_{BLS}}\right) \end{pmatrix} \qquad (23)$$

$$\hat{\mathbf{b}} = \mathbf{R}^{-1}\mathbf{g}. \qquad (24)$$

Note that $w_{BLS}$, the weighting factor, is equal for each data point under homoskedasticity and equals the variance of the residuals. The vector $\hat{\mathbf{b}}$ provides the estimates $\hat{\beta}_{BLS}$ and $\hat{\alpha}_{BLS}$; under homoskedasticity it can be proven that $\hat{\beta}_{BLS} = \hat{\beta}_{DR}$ and $\hat{\alpha}_{BLS} = \hat{\alpha}_{DR}$.

Riu and Rius [22] propose the following variance-covariance matrix for the BLS parameters:

$$\mathbf{V(b)}_{BLS} = s_{BLS}^2 \mathbf{R}^{-1}, \qquad (25)$$

or equivalently

$$s_{\hat{\beta}_{BLS}}^2 = \frac{W_{BLS}Ns_{BLS}^2}{N\sum_{i=1}^{N}X_i^2-(\sum_{i=1}^{N}X_i)^2} \text{ and}$$

$$s_{\hat{\alpha}_{BLS}}^2 = \frac{W_{BLS}s_{BLS}^2\sum_{i=1}^{N}X_i^2}{N\sum_{i=1}^{N}X_i^2-(\sum_{i=1}^{N}X_i)^2}. \qquad (26)$$

The approximate and symmetric CI's for $\beta$ or $\alpha$ are then given by the following formulas [6]:

$$\text{CI}(\beta_{BLS}): \hat{\beta}_{BLS} \pm t_{1-\frac{\gamma}{2};N-2}s_{\hat{\beta}_{BLS}} \text{ and}$$

$$\text{CI}(\alpha_{BLS}): \hat{\alpha}_{BLS} \pm t_{1-\frac{\gamma}{2};N-2}s_{\hat{\alpha}_{BLS}}. \qquad (27)$$

# Bootstrap in errors-in-variables regressions

In this section, two well-known bootstrap procedures (bootstrapping the pairs and bootstrap on the residuals) are briefly explained, as well as the jackknife procedure [24]. These approaches are compared using simulations and real data.

## Jackknife

The jackknife is a simplified version of the bootstrap, applied by the MedCalc software in method comparison studies and sometimes suggested in the literature [25-27]. The main advantages are its simplicity and its fast algorithm. Figure 2 illustrates the jackknife procedure for the estimation of a regression line. First, the regression line is estimated with the initial sample (the "true" sample) to obtain the estimated values of the slope and the intercept, $\hat{\beta}$ and $\hat{\alpha}$. Then, each point in the scatterplot is removed alternately and for each step a new regression line is estimated. $N$ "pseudo"-regressions are therefore obtained, each with $N-1$ points. When the point $(X_i, Y_i)$ is removed, the estimated slope and intercept are given respectively by $\hat{\beta}_{-i}$ and $\hat{\alpha}_{-i}$. The jackknife estimators after $N$ steps are respectively given by $\hat{\beta}_{\text{jack}} = N\hat{\beta} - (N-1)/N \sum_{i=1}^{N} \hat{\beta}_{-i}$ and

$$\hat{\alpha}_{\text{jack}} = N\hat{\alpha} - (N-1)/N \sum_{i=1}^{N} \hat{\alpha}_{-i}. \tag{28}$$

The CI's are computed by the jackknife procedure as follow:

$$\text{CI}(\beta): \hat{\beta}_{\text{jack}} \pm q_{1-\frac{\gamma}{2}} s_{\hat{\beta}_{\text{jack}}} \text{ and}$$

$$\text{CI}(\alpha): \hat{\alpha}_{\text{jack}} \pm q_{1-\frac{\gamma}{2}} s_{\hat{\alpha}_{\text{jack}}}, \tag{29}$$

where $q_{1-\gamma/2}$ is the $1-\gamma/2$ quantile of the standardized normal distribution, and

$$s_{\hat{\alpha}_{\text{jack}}}^2 = \frac{N-1}{N} \sum_{i=1}^{N} \left( \hat{\alpha}_{-i} - \frac{1}{N} \sum_{i=1}^{N} \hat{\alpha}_{-i} \right)^2 \text{ and}$$

$$s_{\hat{\beta}_{\text{jack}}}^2 = \frac{N-1}{N} \sum_{i=1}^{N} \left( \hat{\beta}_{-i} - \frac{1}{N} \sum_{i=1}^{N} \hat{\beta}_{-i} \right)^2. \tag{30}$$
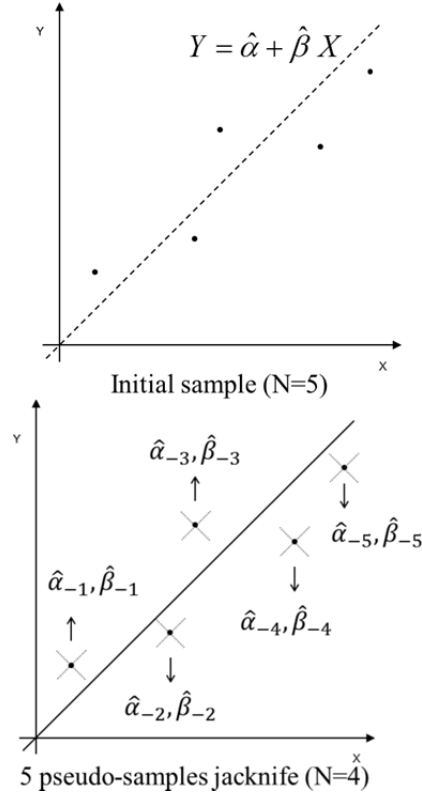


**Figure 2** Illustration of the jackknife procedure for the estimation of a regression line.
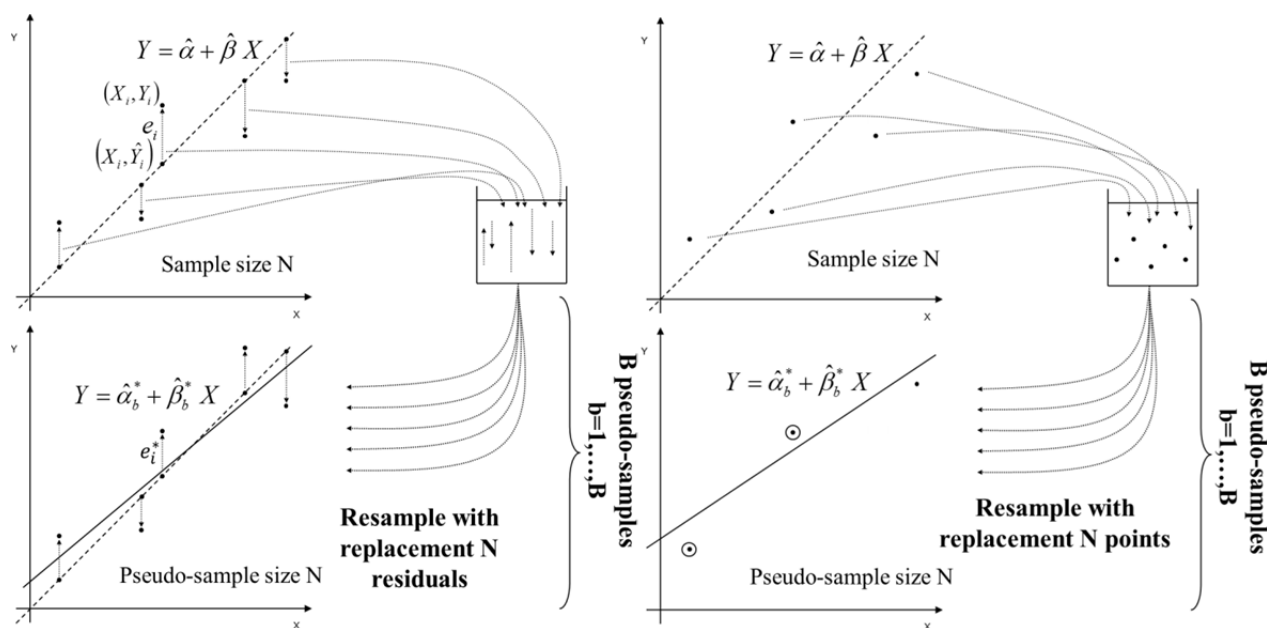
**Figure 3** Illustration of the bootstrapping the residuals procedure (left) and bootstrapping the pairs (right) for the estimation of a regression line (the circled point is a point resampled twice).

### Bootstrapping the residuals

Figure 3 (left) illustrates the bootstrap procedure on the vertical residuals. First, the regression line is estimated with the initial sample to obtain the estimated values of the slope and the intercept, $\hat{\beta}$ and $\hat{\alpha}$. Then, the vertical residuals are computed: $e_i = Y_i - \hat{\alpha} - \hat{\beta}X_i = Y_i - \hat{Y}_i$ and these residuals are resampled: $e_i^*$ is the $i$<sup>th</sup> resampled bootstrap residual. These resampled residuals are added to the initial predicted values to get a pseudo-sample of size $N$ where the $i$<sup>th</sup> point is $\left(X_i, Y_i^* = \hat{Y}_i + e_i^*\right)$. This is repeated $B$ times ($b = 1, \dots, B$) and for each step the slope and the intercept are estimated (as well as their variances), respectively for the pseudo-sample $b$ by $\hat{\beta}_b^*$ (its variance being $s_{\hat{\beta}_b}^{*2}$) and $\hat{\alpha}_b^*$ (its variance being $s_{\hat{\alpha}_b}^{*2}$). For each $b$ step, the following standardised deviates are computed:

$$U_{\hat{\beta},b}^* = \frac{\hat{\beta}_b^* - \hat{\beta}}{s_{\hat{\beta}_b}^*} \text{ and } U_{\hat{\alpha},b}^* = \frac{\hat{\alpha}_b^* - \hat{\alpha}}{s_{\hat{\alpha}_b}^*}. \qquad (31)$$

At this point, two different approaches can be followed to compute a confidence interval: the bootstrap-$t$ or the percentile bootstrap. The percentile bootstrap is certainly the easiest

solution as the confidence interval is computed directly by the $\gamma/2$ and $1 - \gamma/2$ percentile of the empirical distribution (i.e., the $B$ values) of $\hat{\beta}_b^*$ or $\hat{\alpha}_b^*$. The confidence interval by the bootstrap-$t$ is computed as

$$\text{CI}(\beta): \hat{\beta} \pm u_{\hat{\beta},1-\frac{\gamma}{2}}^* s_{\hat{\beta}} \text{ and}$$

$$\text{CI}(\alpha): \hat{\alpha} \pm u_{\hat{\alpha},1-\frac{\gamma}{2}}^* s_{\hat{\alpha}} \qquad (32)$$

where $u_{\hat{\beta},1-\frac{\gamma}{2}}^*$ is the $1 - \gamma/2$ quantile of the $U_{\hat{\beta},b}^*$ values and $u_{\hat{\alpha},1-\frac{\gamma}{2}}^*$ is the $1 - \gamma/2$ quantile of the $U_{\hat{\alpha},b}^*$ values.

### Bootstrapping the pairs

Figure 3 (right) illustrates the technique of bootstrapping the pairs. First, the regression line is estimated with the initial sample to obtain the estimated values of the slope and the intercept, $\hat{\beta}$ and $\hat{\alpha}$. Then, the points $(X_i, Y_i)$ are resampled where $(X_i^*, Y_i^*)$ is the $i$<sup>th</sup> resampled point. This is repeated $B$ times and for each step, as explained in the previous section, $\hat{\beta}_b^*$ ($s_{\hat{\beta}_b}^{*2}$) and $\hat{\alpha}_b^*$ ($s_{\hat{\alpha}_b}^{*2}$) are computed. For each $b$ step, the following values

are obtained:

$$W_{\hat{\beta},b}^* = \frac{\hat{\beta}_b^* - \hat{\beta}}{s_{\hat{\beta}_b}^*} \text{ and } W_{\hat{\alpha},b}^* = \frac{\hat{\alpha}_b^* - \hat{\alpha}}{s_{\hat{\alpha}_b}^*}. \tag{33}$$

As previously explained, the percentile bootstrap or the bootstrap-$t$ can be applied on the $W_{\hat{\beta},b}^*$ or $W_{\hat{\alpha},b}^*$ values to obtain the confidence interval for $\beta$ or $\alpha$.

## Coverage probabilities of the bootstrap procedures

In order to compare the coverage probabilities of the confidence intervals provided by the DR and BLS regressions and the bootstrap procedures presented in the previous sections, $10^4$ samples were simulated with $N = 10$ and 50 with unreplicated data ($n_X = n_Y = 1$, $\lambda = \lambda_{XY}$ known) under equivalence ($\alpha = 0, \beta = 1, \eta_i = \xi_i$) for the values of $\eta_i$ and $\lambda_{XY}$ described in Francq and Govaerts [19]. For each simulated sample, the CI is computed by DR and BLS and with the bootstrap procedures described in the previous sections (with $B = 500$). Note that the percentile bootstrap provides the same results for DR and BLS because $\hat{\beta}_{DR} = \hat{\beta}_{BLS}$ and $\hat{\alpha}_{DR} = \hat{\alpha}_{BLS}$, whereas the bootstrap-$t$ provides different CI's for DR and BLS because the variances of the parameters are taken into account (and are computed differently for DR and BLS). Finally, the coverage probabilities of the slopes (with a 95% nominal level) are computed for a given $\lambda_{XY}$.

Figure 4 displays the coverage probabilities with respect to $\lambda_{XY}$ (which is graphed on a logarithmic scale) for $N = 10$ (left) and $N = 50$ (right). The exact formula for the CI for the slope by DR obviously provides the best coverage probabilities; the approximate ones provided by DR or BLS are slightly lower, especially for $N = 10$ and also when $\lambda_{XY} < 1$ for the BLS with $N = 10$ or $N = 50$. As expected, the jackknife approach provides coverage probabilities closer to the nominal level for $N = 50$ as the number of pseudo-samples is higher (the estimator obtained is therefore more precise). The coverage probabilities provided by bootstrapping the residuals collapse drastically when $\lambda_{XY}$ decreases and when $N$ increases, because the randomness of the errors in X is not taken into account by bootstrapping the vertical residuals. Lastly, the coverage probabilities provided by bootstrapping the pairs are very close to those obtained by the bootstrap-$t$ technique on DR or BLS, while the percentile technique is slightly worse. When $N$ increases, the three bootstrap techniques on the pairs move closer to each other and closer to the nominal level. It is noteworthy that bootstrapping the pairs can provide better coverage probabilities than the BLS formula, especially when $\lambda_{XY} < 1$ and $N = 50$.
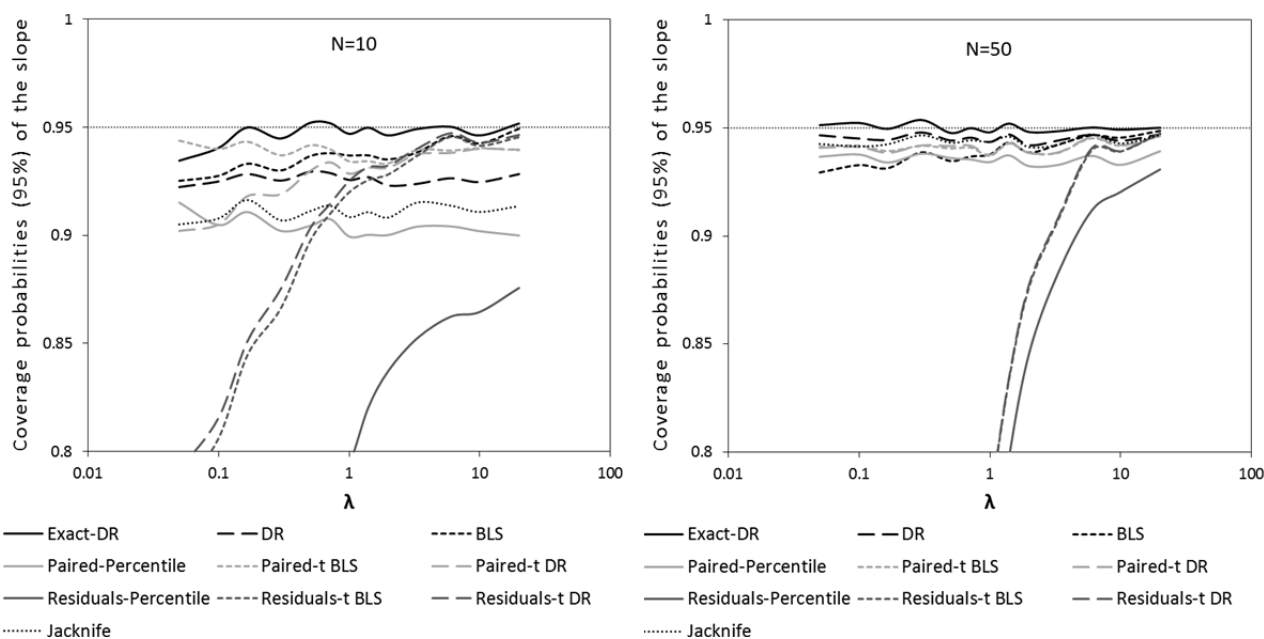
**Figure 4** Coverage probabilities of the CI for the slope (β), for $N = 10$ (left) or $N = 50$ (right) related to λ ($= \lambda_{XY}$ with $n_X = n_Y$) in a logarithmic scale, for the Deming Regression (DR) with its exact formula or approximate one, the Bivariate Least Square regression (BLS) with its approximate formula, two bootstrap procedures (on the pairs or residuals) split into three approaches (percentile, bootstrap-t on DR and bootstrap-t on BLS) and the jackknife.

## Application

In the systolic blood pressure data [2], simultaneous measurements were made by two observers (denoted J and R) using a sphygmomanometer and a semi-automatic blood pressure monitor (denoted S) for 85 patients. The systolic blood pressure was measured three times per patient by S and three times per patient by J (R is not considered here; for a brief overview of other designs and approaches see a recent example of a method comparison study from the field of rehabilitation [28]).

If the mean measurements given by S are assigned to the Y-axis and J to the X-axis, then it follows that the estimated value of λ ($=\lambda_{XY}$ as $n_X = n_Y$) is 2.223, and therefore $\hat{\beta}_{DR} = \hat{\beta}_{BLS} = 0.956$ and $\hat{\alpha}_{DR} = \hat{\alpha}_{BLS} = 21.230$ [19,29]. Figure 5 illustrates the different CI's for $\beta$ computed using the exact and approximate DR formula, the approximate BLS formula, the jackknife procedure, and six bootstrap approaches (percentile method, bootstrap-t on DR or BLS, bootstrapping the pairs

and bootstrapping the residuals). The exact DR formula provides a slightly asymmetric CI while the approximate DR and BLS CI's are symmetric. These three CI's are similar although the BLS one is slightly narrower. The CI obtained by jackknife is narrower than those obtained without resampling but the estimated slope is very similar to the previous ones. The bootstrap-t on either DR or BLS also yields very similar CI's while the percentile method provides a slightly wider CI and higher estimate. As expected, bootstrapping the residuals provides a shifted CI (upwards for the percentile method and downwards for the bootstrap-t). As explained in the previous section, the coverage probabilities of the bootstrap on the residuals collapse drastically and the CI's are therefore wrong because the randomness of the errors in the X variable is not taken into account.

The hypothesis $H_o: \beta = 1$ is not rejected for the CI's computed directly by DR or BLS, the jackknife or by bootstrapping the pairs. On the other hand, this hypothesis is erroneously rejected for the

bootstrap on the residuals (except for the percentile method) as these CI's are shifted.
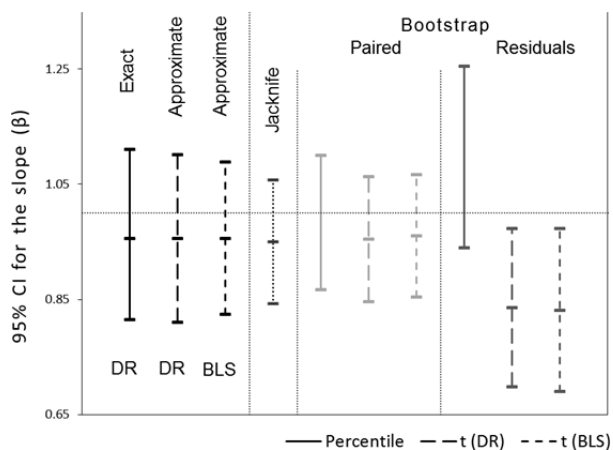


**Figure 5** The CI for the slope (β) for the Systolic Blood Pressure data computed by Deming Regression (DR) with exact and approximate formula, the Bivariate Least Squares regression (BLS) with approximate formula, two bootstrap procedures (on the pairs and on the residuals) split into three approaches (percentile, bootstrap-t on DR and bootstrap-t on BLS), and the jackknife.

## Conclusion

Six different bootstrap procedures were compared in order to improve the coverage probabilities of the approximate confidence intervals for the parameters of the DR and BLS regressions. The bootstrap-t on DR or BLS provides very similar results. These two regressions are actually confounded under homoskedasticity and the variances of the parameters, though computed differently, are similar in practice. The jacknife is a simple method but its coverage probabilities are lower than the nominal level for small sample sizes, and its CI may therefore be too narrow in practice. Bootstrapping the residuals is not recommended as the coverage probabilities collapse and the CI's are shifted in practice. Bootstrapping the pairs is recommended to improve the coverage probabilities especially when the ratio of the measurement errors' variances is less than one. It can provide better coverage

probabilities than the approximate CI computed directly by DR or BLS. Moreover, this bootstrap approach takes into account the measurement errors in both variables.

**References**

1.  Westgard JO, Hunt MR: Use and interpretation of common statistical tests in method-comparison studies. *Clin Chem* 1973; 19: 49-57.
2.  Bland JM, Altman DG: Measuring agreement in method comparison studies. *Stat Methods Med Res* 1999; 8: 135-160.
3.  Altman DG, Bland JM: Measurement in medicine: the analysis of method comparison studies. *Statistician* 1983; 32: 307-317.
4.  Bland JM, Altman DG: Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986; 1(8476): 307-310.
5.  Lindley DV: Regression Lines and the Linear Functional Relationship. *J Roy Statist Soc Suppl* 1947; 9(2): 218-244.
6.  Martinez A, Del Rio FJ, Riu J, Rius FX: Detecting proportional and constant bias in method comparison studies by using linear regression with errors in both axes. *Chemom Intell Lab Syst* 1999; 49(2): 181-195.
7.  Riu J, Rius FX: Univariate regression models with errors in both axes. *J Chemometr* 1995; 9: 343–362.
8.  Madansky A: The fitting of straight lines when both variables are subject to error. *J Am Statist Assoc* 1959; 54(285): 173-205.
9.  Barnett VD: Fitting straight lines – the linear functional relationship with replicated observations. *J Roy Statist Soc Ser C* 1970; 19(2): 135-144.
10. Fuller WA: *Measurement error models*. New York 1987: Wiley.
11. Tan CY, Iglewicz B: Measurement-Methods Comparisons and Linear Statistical Relationship. *Technometrics* 1999; 41(3): 192-201.
12. Legendre AM: *Nouvelles méthodes pour la détermination des orbites des comètes, Appendice sur*

*la méthode des moindres carrés*. Paris 1805: Firmin-Didot.

13. Gauss CF: *Theoria motus corporum coelestium in sectionibus conicis solem ambientum*. 1809 (translated by Davis CH, New York 1963: Dover).

14. Cornbleet PJ, Gochman N: Incorrect least-squares regression coefficients in method-comparison analysis. *Clin Chem* 1979; 25(3): 432-438.

15. Dagnelie P: Statistique théorique et appliquée. Tome 2. Inférence statistique à une et à deux dimensions. Bruxelles 2011: De Boeck.

16. Linnet K: Necessary sample size for method comparison studies based on regression analysis. *Clin Chem* 1999; 45(6): 882-894.

17. Gillard JW, Iles TC: *Variance covariance matrices for linear regression with errors in both variables*. Cardiff 2006: Cardiff University School of Mathematics Technical Report.

18. Gillard JW, Iles TC: *Method of moments estimation in linear regression with errors in both variables*. Cardiff 2005: Cardiff University School of Mathematics Technical Report.

19. Francq BG, Govaerts BB: Measurement methods comparison with errors-in-variables regressions. From horizontal to vertical OLS regression, review and new perspectives. *Chemom Intell Lab Syst* 2014; 134: 123-139.

20. Lisý JM, Cholvadová A, Kutej J: Multiple straight-line least-squares analysis with uncertainties in all variables. *Computers Chem* 1990; 14(3): 189-192.

21. Del Río FJ, Riu, J, Rius, FX: Prediction intervals in linear regression taking into account errors on both axes. *J Chemometr* 2001; 15: 773–788.

22. Riu, J, Rius, FX: Assessing the accuracy of analytical methods using linear regression with errors in both axes. *Anal Chem* 1996; 68: 1851-1857.

23. Martinez A, Riu, J, Rius FX: Evaluating bias in method comparison studies using linear regression with errors in both axes. *J Chemometr* 2002; 16: 41–53.

24. Efron B, Tibshirani RJ: *An introduction to the bootstrap*. New York 1993: Chapman & Hall.

25. Armitage P, Berry G, Matthews JNS: *Statistical methods in medical research* (4th ed). Oxford 2002: Blackwell Science.

26. Linnet K: Performance of Deming regression analysis in case of misspecified analytical error ratio in method comparison studies. *Clin Chem* 1998; 44(5): 1024-1031.

27. Linnet K: Estimation of the linear relationship between the measurements of two methods with proportional errors. *Statist Med* 1990; 9: 1463–1473.

28. Vidmar G, Burger H, Erjavec T: Options for Comparing Measurement Agreement between Groups: Exercise Testing as Screening for Ability to Walk After Transfemoral Amputation. *Inf Med Slov* 2010; 15(2): 10-20.

29. Francq BG, Govaerts BB: Hyperbolic confidence bands of errors-in-variables regression lines applied to method comparison studies. *J Soc Fr Stat* 2014; 155(1): 23-45.