UNIVERSITE CATHOLIQUE DE LOUVAIN

Institut de recherche multidisciplinaire pour la modélisation et l'analyse quantitative

Institut de statistique, biostatistique et sciences actuarielles

Errors-in-variables regressions to assess equivalence in method comparison studies

Ph.D. Thesis

Thèse présentée en vue de l'obtention du grade de
Docteur en Sciences (orientation statistique) par:
Bernard Francq

Membres du jury:
Prof. Bernadette Govaerts (Promoteur, UCL)
Prof. Ingrid Van Keilegom (Président du Jury, UCL)
Prof. Bruno Boulanger (Arlenda, ULg)
Prof. Mia Hubert (KUL)
Prof. Catherine Legrand (UCL)
Dr Pierre Pestiaux (Total)

Louvain-la-Neuve, Octobre 2013.

Acknowledgments

This thesis would not be what it is now if I had not been surrounded by outstanding people.
I would first like to thank my supervisor Professor Bernadette Govaerts. Thank you for the autonomy, your trust in me and the encouragements you gave me during these last years. Thank you for your comments, suggestions, ideas, checking for spelling mistakes for the papers we wrote and for this thesis.
I also want to thank Dr Pierre Pestiaux for your sympathetic ear for my results, our fruitful collaboration and your comments and suggestions to enhance this thesis.
I would also like to thank the members from my committee: Professor Mia Hubert, Professor Catherine Legrand and Professor Ingrid Van Keilegom. They gave me several good advices on my work and my research, thank you for your reading of the manuscript and your comments, suggestions to improve this thesis, ideas, checking for spelling mistakes. I would like also to acknowledge Professor Bruno Boulanger, member of the jury, for your reading of the manuscript and your comments, suggestions to improve this thesis.
Thanks to ISBA: the staff, secretariat, IT, colleagues and friends.
Finally, thank to Alain Guillet, Dr ès R, for the tips you gave me in R.

Contents

Introduction

The needs of the industries or laboratories to quickly assess the quality of products or samples leads to the development and improvement of new measurement methods or devices sometimes 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]. This means that there is no bias between these methods (the measurements provided by two devices or methods only differ because of the random measurement errors) or that these measurement methods could be interchangeable (the differences between the measurements provided by two devices are not meaningful in practice).
It is then essential for the chemical or clinical laboratories to assess the uncertainties of the measures and to control the reliability of measurements. The quality of the measures is also crucial for the customers (patients, researchers, therapists, doctors, industries,...) in order to carry out the right decisions and to properly interpret the analyzes.

To monitor and inspect the analytical methods or devices, four different methodologies coexist.
First, the validation of analytical methods which is the final step in the development of a new method. The goal of the validation is to attest that the analytical procedure will provide results as close as possible (according to given criteria) to the true but unknown quantities for its whole range of measurability. The linearity of the method is also assessed with the validation.
Second, the transfert of analytical methods attests that the results between an analytical method used in routine by a laboratory are similar to those obtained by the same procedure applied in a new laboratory or transferred to another one.
Third, the interlaboratories studies consists to compare the analytical measures provided by different laboratories on the same samples. It is obviously essential to proof that the decision carried out by analytical measures does not rely to the chosen laboratory. Indeed, different analytical laboratories must provide measures as close as possible to each other.

Four, the ***method comparison studies***, also called the ***equivalence*** of analytical methods, focus on the comparison of different analytical methods. When several new measurement methods are developed, each pair can be compared or each new measurement method can potentially be compared to the reference method.
These four methodologies, based on continuous variables, are quite common in the following fields: pharmaceutical, chemistry, biology, medicine, agronomy or engineering. They have several similarities but they are nearly always discussed separately in the literature. This thesis deals with the fourth methodology, the method comparison studies (or equivalence), and more precisely the comparison of two measurement devices.
The literature does not provide a clear and unique definition of the equivalence concept. Statistical equivalence approach will, in priority in this thesis, test whether two devices are equivalent notwithstanding the errors of measurement or whether there is a bias between them. Additionally, a statistical test can be performed to compare the accuracies of the two methods.
Practical equivalence approach will not focus on statistical parameters (bias and variance) but will consider two methods equivalent when one device can be substituted by the other one without affecting the decision taken from the measurement result.

Statistical approaches in method comparison studies

Different approaches are proposed in the literature to deal with method comparison studies and equivalence:

1. Correlation coefficients or t-test for paired data
   Some authors [2,3] assess the equivalence between two measurement methods by computing the well-known Pearson correlation coefficient r (or its square: the coefficient of determination) or a t-test for paired data. Unfortunately, the correlation coefficient is not useful to test the equivalence. It is, indeed, usually close to 1 in method comparison studies. More appropriate coefficients exist in the literature [4–6] and will be described briefly in Chapter 1 after the presentation of the general model with standardized notations. The uselessness of paired t-tests (or Wilcoxon signed rank test) will also be described in Chapter 1.

2. Errors-in-variables regressions
   The approach based on a regression analysis (a linear functional relationship [7]) on the measures provided by both devices is widely applied and focuses on the parameter estimators and their confidence intervals [8]. This approach compares the measures by estimating a regression line in a classical (X, Y) plot by taking into account the measurement errors in both axes. Different regression techniques are compared in Chapter 2

(comparison of bias, coverage probabilities under equivalence or not, ...). New diagrams are also proposed to assess the equivalence between two devices under homoscedasticity. These diagrams are very useful when the measurement errors are unknown. The comparison of the different regressions and these new diagrams are also discussed by Francq and Govaerts [9]. Hyperbolic confidence bands are also proposed and shown to be equivalent to the simultaneous confidence interval for the regression parameters in Chapter 3. These hyperbolic confidence bands are also discussed under homoscedasticity or heteroscedasticity by Francq and Govaerts [10]. Under heteroscedasticity, the estimated variances (point by point) can be modeled (variances profiles) in order to improve the coverage probabilities. This improvement will also be discussed in Chapter 3, as well as the robustness of the regressions.

3. Bland and Altman approach
   The most known and widely used approach is certainly the one proposed by Bland and Altman which focuses directly on the differences between two measurement methods [11–13]. This method was initially proposed by Tukey with a Tukey mean-difference plot and popularized in medical statistics by J. Martin Bland and Douglas G. Altman [14]. Usually, an agreement interval is computed on the differences and compared to an acceptance interval to assess whether the differences are clinically important or not (acceptable or not). If the differences are not clinically important, then the two measurement methods can be considered as interchangeable. Chapter 4 presents the coverage probabilities under equivalence of the agreement interval. Additionally, β expectation and $\beta - \gamma$ content tolerance intervals are also introduced as a better alternative to the agreement interval. A robust estimator of the variance of the single differences will also be given. Usually, these intervals are computed and displayed horizontally in a Bland and Altman plot. Ideally, as a proportional bias can exist between the two devices, these intervals should be computed around a regression line. By default, Bland and Altman propose to estimate a regression line by OLS [11]. A new asymptotic unbiased regression line is proposed in Chapter 4 by using a correlated-errors-in-variables regression based on existing errors-in-variables regressions and the coverage probabilities are provided under homoscedasticity. New predictive intervals are also proposed in Chapter 4 in order to predict a single measure or a single difference. The main results of Chapter 4 were published by Francq and Govaerts [15].

The general goal of this thesis is to improve and compare the errors-in-variables regressions approach and the Bland and Altman approach. As the first one focuses on confidence intervals while the second focuses on agreement or predictive intervals, they are not easily comparable. Indeed, the main difference

is that a confidence interval collapses to itself when the sample size increases while agreement or tolerance (predictive) intervals move closer to the concerned quantiles. In order to confront these two approaches, it is then compulsory to develop confidence intervals within a Bland and Altman plot and predictive intervals in a (X, Y) plot. The last chapter of this thesis will, then, compare both approaches and will conclude whether it is better to regress in a Bland and Altman plot or in a (X, Y) plot. It will also conclude whether it is better to apply an agreement interval or predictive interval in a Bland and Altman plot or in a (X, Y) plot.

Examples of data set in method comparison studies

Many data sets dealing with method comparison studies can be found in the literature. Moreover, the statistical consulting service (SMCS) of the Université Catholique de Louvain (UCL) is regularly consulted in this context. This section presents a list of examples from these sources, with their design and references when available. Note that the measurement errors are usually unknown in practice. The two first examples are analyzed in this thesis.

- Systolic Blood Pressure data (SBP data)
  In the systolic blood pressure data presented and analyzed by Bland and Altman [11], simultaneous measurements were performed using a sphygmomanometer by two observers (denoted J and R) and a semi-automatic blood pressure monitor (denoted S). The design is quite simple: three sets of readings were made in quick succession by both devices (and both observers) on each patient (85 patients, 9 measures per patient). This data set will be a leitmotiv in this thesis. This data set is characterized by the presence of many outliers (mainly vertical outliers) and possible heteroscedasticity.
  Other data sets are also provided by Bland and Altman [11] (but will not be analyzed in this thesis):
  - Measurements of plasma volume expressed as a percentage of normal in 99 subjects (unreplicated data), using two alternative sets of normal values due to Nadler and Hurley
  - Fat content of human milk (45 samples, unreplicated data) determined by enzymic procedure for the determination of triglycerides and measured by the Standard Gerber method
  - Cardiac data provided by two methods, radionuclide ventriculography (RV) and impedance cardiography (IS), for 12 subjects with unequal numbers of replicates
- Arsenate ion in natural river water (AsO_4 data)
  In the arsenate ion in natural river water data [16], 30 pairs of measures

are provided by 2 methods: firstly, a continuous selective reduction and atomic absorption spectrometry and secondly, a non-selective reduction, cold trapping and atomic emission spectrometry. The mean measures with their standard errors of the mean are given and analyzed in the literature [16,17] (the detailed data are not given in the literature). Lower concentrations are more frequent than higher concentrations. This data set is characterized by measurement errors that increase with measured concentrations for both devices (this is common in chemistry) and some good leverage points (without a logarithmic transformation).

- Bone cyst fluid volume
  In the bone cyst fluid volume data [18], 29 bone cysts are measured by an usual manual method (manual contour tracing) and by a semiautomatic segmentation method designed to easily measure the volume of a bone cyst from magnetic resonance imaging (MRI). Two independent and trained observers performed each two manual segmentations retrospectively for each cyst and also two semiautomatic segmentations (2 measures per cyst per observer). This data set was provided by the 'Cliniques universitaires Saint-Luc' for a statistical consultation through the SMCS and analyzed by B. Francq. This data set is characterized by the heteroscedasticity.

- Pigmentation of skin
  This data set was provided by a physiotherapist at UCL. The skin pigmentation of 11 patients was measured after healing of a scar by two colorimeters: the Minolta CR-300 and the Gardner SpectroGuide. These data were analyzed in the advanced master thesis of Francq [19] and is characterized by unknown measurement errors and the absence of replicated data (actually, the devices measure three times in quick succession but display only the average of the three measures).

- Compressive force on the concrete
  This data set was provided in the context of a thesis in engineering at UCL. The compressive force on 17 samples of concrete was measured in two different laboratories: LEMSC and CEPESI. These data were analyzed by Francq and the results described briefly in a consultation report. This data set is characterized by unknown measurement errors with unreplicated data and two outliers (good leverage points).

- Blood analysis by the red-cross
  The red-cross provided to the SMCS several data sets and some of them were analyzed by Francq and described briefly in a consultation report.

  - Comparison of the concentration of antibody ANTI-HBs in 29 positives blood samples (the negatives results are not compared) pro-

vided by two devices (with one or two days delays): Axsym (routine device) and Architect (new device). This data set is characterized by unknown measurement errors with unreplicated data and log-normal distributions for X and Y.

- Comparison of three hematology devices: Cell-Dyn (routine device) and two new and identical devices (located in different cities) Sysmex (A1609 and A1615) to measure white blood cell (WBC), red blood cell (RBC), platelets (PLT), hemoglobin (HGB), percentage of hematocrit and mean corpuscular volume in 119 blood samples. This data set is characterized by unknown measurement errors and unreplicated data.
- Comparison of two new devices (Sysmex) to the same device in the main laboratory of the red-cross to measure and assess the equivalence with the previously listed variables. There are 20 blood samples, unknown measurement errors and unreplicated data.
- Comparison of 34 new portable devices (Hemocue) to the fixed laboratory of the red-cross: each portable device is compared to the laboratory one (34 comparisons) for a total of 1211 blood samples but the sample sizes varies from a portable device to another one (from 3 to 89). Replicated data are available. This data set is obviously characterized by the number of equivalence test to perform and the need for multiple testing adjustments. However, this thesis deals with the equivalence of two measurement methods.

Design in method comparison studies

In the above examples, measurements are performed using (at least) two measurement methods in order to study whether these two devices are equivalent or not. This thesis will not deal in detail with the design of experiments in method comparison studies although there are few articles on this topic [20]. However, the most classical design consists of measuring each sample (or subject) once by both devices. Unfortunately, whith unreplicated data, it is not possible to estimate the measurement error variances if needed and a design with replicated data is more suitable. If the accuracy of a reference measurement method (gold standard) is known, each sample (or subject) can, more simply, be measured only once by this method and several times by the new method. Different observers or operators can, sometimes, perform the measurements for both devices (i.e. Bone cyst fluid volume data) or for one device (i.e. SBP data). In the first case, the observers could be compared like two different devices. In both cases, the design structure can be taken into account with an ANOVA analysis in order to better estimate the measurement error variances to finally compare the devices. In the systolic blood pressure data,

the design is quite simple: three sets of readings were made by both devices (and both observers for the manual device) on each patient.

Chapter 1

Model and goal of methods equivalence testing

Contents

1.1 The general model

To compare two measurement methods, we suppose that a parameter of interest is measured on N samples ($i = 1, 2, ..., N$) or subjects by both methods [21–23]:

$$X_{ij} = \xi_i + \tau_{ij},\ Y_{ik} = \eta_i + \nu_{ik} \tag{1.1}$$

$X_{ij} (j = 1, 2, ..., n_{X_i})$ and $Y_{ik} (k = 1, 2, ..., n_{Y_i})$ are the repeated measures for sample i by methods X and Y respectively. n_{X_i} and n_{Y_i} are the number of repeated measures of sample i by each method. ξ_i and η_i are the true but unobservable value of the parameter of interest for both methods, assumed to be linked by a linear regression [21–23]:

$$\eta_i = \alpha + \beta\xi_i \tag{1.2}$$

Note that a lack of fit test can be realized in practice to assess this assumption [24,25]. τ_{ij} and ν_{ik} are the measurement errors supposed independent and normally distributed (with constant variances under homoscedasticity):

$$\begin{pmatrix} \tau_{ij} \\ \nu_{ik} \end{pmatrix} \sim iN\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma^2_{\tau_i} & 0 \\ 0 & \sigma^2_{\nu_i} \end{pmatrix}\right) \tag{1.3}$$

X_i and Y_i are the means of the repeated measures for a given sample:

$$X_i = \frac{1}{n_{X_i}} \sum_{j=1}^{n_{X_i}} X_{ij} \text{ and } Y_i = \frac{1}{n_{Y_i}} \sum_{k=1}^{n_{Y_i}} Y_{ik} \tag{1.4}$$

These means are also normally distributed around ξ_i or η_i:

$$\begin{pmatrix} X_i \\ Y_i \end{pmatrix} \sim N\left(\begin{pmatrix} \xi_i \\ \eta_i \end{pmatrix}, \begin{pmatrix} \frac{\sigma^2_{\tau_i}}{n_{X_i}} & 0 \\ 0 & \frac{\sigma^2_{\nu_i}}{n_{Y_i}} \end{pmatrix}\right) \tag{1.5}$$

If the variances $\sigma^2_{\tau_i}$ and $\sigma^2_{\nu_i}$ are unknown, they can be estimated if repeated measures are available, otherwise, these variances are inestimable. The estimators of $\sigma^2_{\tau_i}$ and $\sigma^2_{\nu_i}$ are given by $S^2_{\tau_i}$ and $S^2_{\nu_i}$:

$$S^2_{\tau_i} = \frac{1}{n_{X_i} - 1} \sum_{j=1}^{n_{X_i}} (X_{ij} - X_i)^2 \text{ and } S^2_{\nu_i} = \frac{1}{n_{Y_i} - 1} \sum_{k=1}^{n_{Y_i}} (Y_{ik} - Y_i)^2 \tag{1.6}$$

In further sections, the following notations will also be used:

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

$$S_{xx} = \sum_{i=1}^{N} (X_i - \overline{X})^2, S_{yy} = \sum_{i=1}^{N} (Y_i - \overline{Y})^2 \text{ and } S_{xy} = \sum_{i=1}^{N} (X_i - \overline{X})(Y_i - \overline{Y}).$$

1.2 The homoscedastic model

In the homoscedastic model, we will consider that both measurement methods have a constant accuracy through the domain of interest ($\sigma^2_{\tau_i} = \sigma^2_\tau$ and $\sigma^2_{\nu_i} = \sigma^2_\nu$ $\forall i$). Moreover, as the regression will be applied to the mean measures Y_i with respect to X_i, the number of replicates is, here, constant ($n_{X_i} = n_X$ and $n_{Y_i} = n_Y$ $\forall i$) in order to prevent the model to become heteroscedastic even if the accuracies of measurement methods are constant.

In the case of homoscedasticity, the variances $S^2_{\tau_i}$ and $S^2_{\nu_i}$ are estimates of σ^2_τ and σ^2_ν and we can define global estimates for σ^2_τ and σ^2_ν by S^2_τ and S^2_ν:

$$S^2_\tau = \frac{\sum_{i=1}^N (n_{X_i} - 1) S^2_{\tau_i}}{(\sum_{i=1}^N n_{X_i}) - N} \text{ and } S^2_\nu = \frac{\sum_{i=1}^N (n_{Y_i} - 1) S^2_{\nu_i}}{(\sum_{i=1}^N n_{Y_i}) - N}$$

or with constant number of repeated measures ($n_{X_i} = n_X$ and $n_{Y_i} = n_Y \ \forall i$), we have :

$$S^2_\tau = \frac{\sum_{i=1}^N S^2_{\tau_i}}{N} \text{ and } S^2_\nu = \frac{\sum_{i=1}^N S^2_{\nu_i}}{N} \tag{1.7}$$

1.3 How to test the equivalence?

1.3.1 Correlation coefficients or t-test paired data

As mentioned in the introduction, some authors [3] assess the equivalence between two measurement methods by computing the well-known Pearson correlation coefficient R (or its square: the determination coefficient). Unfortunately, the estimated correlation coefficient r is not appropriate to assess the equivalence [2,26], as well as the intraclass correlation coefficient [27]. Recall that the Pearson correlation coefficient is estimated by the following formula:

$$r = \frac{S_{xy}}{\sqrt{S_{xx} S_{yy}}} \tag{1.8}$$

It is wrong to believe that the higher is this coefficient the better is the equivalence. Moreover, as reminded by Bland and Altman [11], r is often much higher than 0.9 or even close to 1 when comparing the measures provided by two devices on the same samples (or patients). To illustrate the inadequacy of the correlation coefficient, we designed three small and fictive data sets for which five samples were measured once by device X and once by device Y. These data are given in Table 1.1 and displayed in Figure 1.1 with the equivalence line $Y = X$.

Data set 1 was created with the following X-values: 10, 12, 14, 16, 18 and the Y-values were created by adding 2 to the X-values. It is, then, obvious that $r = 1$ but X and Y are not equivalent as there is a systematic bias equal to two. It is also obvious that the PValue of a t-test paired data (or Wilcoxon signed rank test) is equal to 0 (or not computable).

Data set 2 was created with the following X-values: 11, 13, 15, 17, 19 and the Y-values are the reverse of these values. It is, then, obvious that $r = -1$ but X and Y are not equivalent as they 'cross' (this is a didactic example). It is also obvious that the PValue of a t-test paired data (or Wilcoxon signed rank test) is equal to 1.

Data set 3 was created under equivalence as explained later in Section 1.3.3

	Data set 1		Data set 2		Data set 3	
	X	Y	X	Y	X	Y
1	10	12	11	19	9.8	10.4
2	12	14	13	17	12.2	12.3
3	14	16	15	15	14.1	13.5
4	16	18	17	13	15.9	16.2
5	18	20	19	11	18.2	17.9
PVal	0		1		0.930	
r	1		-1		0.991	
r_c	0.8		-1		0.988	

Table 1.1: *Three fictive data sets to illustrate the uselessness of correlation coefficients or t-test paired data*

and $r = 0.991$ and the PValue of a t-test paired data is equal to 0.930.
Actually, a t-test for paired data (or Wilcoxon signed rank test) is not useful as it does not test the right hypothesis (hypothesis (1.9) given in further section). Indeed, this t-test assesses the following null hypothesis: $H_0 : \mu_X = \mu_Y$ where μ_X and μ_Y are the means of X and Y and have usually no interest in method comparison studies. More appropriate correlation coefficients exist in the literature but are rarely applied like the concordance correlation coefficient given by Lin [4,5] or the gold-standard correlation coefficient (which is related to the intraclass correlation coefficient) given by St. Laurent [6]. These coefficients will not be detailed in this thesis, except the concordance correlation coefficient that we computed for the three fictive data sets. This coefficient is defined as:

$$\rho_c = r \frac{2\sigma_\nu \sigma_\tau}{\sigma_\nu^2 + \sigma_\tau^2 + (\mu_X - \mu_Y)^2}$$

and the corresponding estimator is given by:

$$r_c = \frac{2(S_{xy}/N)}{S_{xx}/N + S_{yy}/N + (\overline{X} - \overline{Y})^2}$$

Additionally to r, r_c takes into account the distance of the estimated line to the identity line $Y = X$ (how far is the estimated line to the identity line). In data set 1, $r_c = 0.8$: indeed, $r = 1$ but we translated the line two units upper to the identity line and this translation penalizes the concordance correlation coefficient which decreases up to 20% in comparison to r. In data set 2, $r_c = -1$: the estimated line is, indeed, orthogonal to the identity line. In data set 3, $r_c = 0.988$ which is nearly equal to r: the estimated line is, indeed, very close to the identity line.

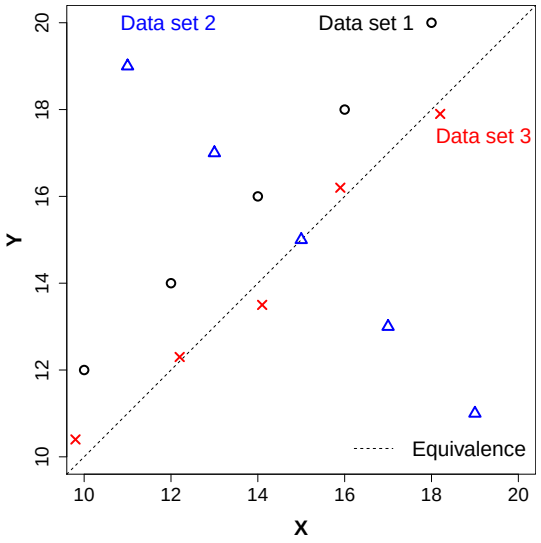

Figure 1.1: *Scatterplot of the three fictive data sets in Table 1.1 with the equivalence line* $Y = X$

Unfortunately, these coefficients are not appropriate as they often provide values too close to 1 and it is not clear to define a threshold to assess the equivalence. Moreover, statistical tests and confidence intervals are not easy to build on such correlation coefficients.

1.3.2 Strict versus flexible equivalence

When two measurement methods are equivalent, we can expect that they provide (on average) the same measure for a given sample notwithstanding the errors of measurement. Or, we can expect that the differences between the measures provided by both devices are not higher than a practical threshold of equivalence.
These two points of views are not exactly the same. Actually, the first point of view focuses directly on a strict equivalence between the devices. The second point of view, popularized by Bland and Altman [11], focuses directly on the differences between the measures with a flexible equivalence. With the first point of view, errors-in-variables regressions are applied in the (X, Y) space and confidence intervals computed on the regression coefficients while the Bland and Altman's approach focuses directly on the $Y - X$ differences and predictive or agreement intervals are applied.
Confidence intervals and predictive intervals are two different concepts in statis-

tics. Both intervals are computed, usually, with a 95% confidence level and 95% predictive level. But, when the sample size increases, a confidence interval collapses to itself such that the null hypothesis will always be rejected, while a predictive interval moves closer to population distribution quantiles: the lower bound moves closer to the 2,5% quantile ($q_{0,025}$) and the upper bound moves closer to the 97,5% quantile ($q_{0,975}$). Confidence intervals are usually applied to assess the uncertainties of estimated parameters or to carry out a hypothesis test. Predictive intervals are usually applied to known in which interval will lie a future value or a future measurement.

1.3.3 Statistical 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 notations, method equivalence means then [8,28]:

$$\xi_i = \eta_i \ \forall i \tag{1.9}$$

In practice, due to the measurement errors, these parameters are unobservable and the equivalence test will be based on the following regression model:

$$Y_i = \alpha + \beta X_i + \epsilon_i \text{ with } \epsilon_i \sim N(0, \sigma^2_{\epsilon_i}) \text{ and } \sigma^2_{\epsilon_i} = \frac{\sigma^2_{\nu_i}}{n_{Y_i}} + \beta^2 \frac{\sigma^2_{\tau_i}}{n_{X_i}}, \tag{1.10}$$

or under homoscedasticity:

$$Y_i = \alpha + \beta X_i + \epsilon_i \text{ with } \epsilon_i \sim N(0, \sigma^2_{\epsilon}) \text{ and } \sigma^2_{\epsilon} = \frac{\sigma^2_{\nu}}{n_{Y}} + \beta^2 \frac{\sigma^2_{\tau}}{n_{X}}, \tag{1.11}$$

where α, the intercept and β, the slope 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. A lot of practitioners wonder which measurement method to put on the X-axis or on the Y-axis. But, actually, the variables X and Y 'play similar roles' [29] and when the regression line is estimated adequately (by taking into account errors in both axis), the coordinate system should not matter [30].
The estimated parameters $\hat{\alpha}$ and $\hat{\beta}$ provide the information to assess the strict equivalence. Indeed, 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 [8]. So, the following two-sided hypotheses will be used to test method equivalence:

$$H_0^{\alpha} : \alpha = 0,\ H_1^{\alpha} : \alpha \neq 0 \text{ and } H_0^{\beta} : \beta = 1,\ H_1^{\beta} : \beta \neq 1 \tag{1.12}$$

The null hypothesis $H_0^\alpha : \alpha = 0$ is rejected if 0 is not included inside a confidence interval (CI) for α and the null hypothesis $H_0^\beta : \beta = 1$ is rejected if 1 is not included inside a CI for β. The two measurement methods can be called 'average equivalent' if the hypotheses $H_0^\beta : \beta = 1$ and $H_0^\alpha : \alpha = 0$ are not rejected [28]. However, these tests can also be applied jointly by checking whether the point $(1, 0)$ is included or not in a joint-CI for the regression coefficients $\theta = (\alpha, \beta)'$ which is, classically, a confidence ellipse:

$$H_0 : \theta = (0, 1)' \text{ and } H_1 : \theta \neq (0, 1)' \tag{1.13}$$

The separated CI will be firstly investigated, followed by the joint-CI.

1.3.4 Practical equivalence

Usually, in practice, the practitioners want to know whether the measurements methods are interchangeable without affecting the decision based on the measures. This means that the devices are allowed to be not strictly equivalent: they can differ up to a given threshold called, in this thesis, an acceptance interval: $[-\Delta, \Delta]$.
For example, with the systolic blood pressure data, differences up to 10 mmHg can be considered not meaningful or not clinically important [11]. As Bland and Altman focus directly on the observed differences between the measurements methods, the goal of testing the equivalence is to check that:

$$|Y_i - X_i| < \Delta, \forall i \ (i = 1, ..., N) \tag{1.14}$$

In a Bland and Altman's plot, the differences D_i are plotted on the Y-axis with respect to the averages M_i on the X-axis (under homoscedasticity and equal number of replicates):

$$D_i = Y_i - X_i \text{ and } M_i = \frac{X_i + Y_i}{2} \tag{1.15}$$

The choice to compute the differences $Y_i - X_i$ or $X_i - Y_i$ does not really matter. The mean and the variance of the differences can also be computed:

$$\overline{D} = \frac{1}{N} \sum_{i=1}^{N} D_i \text{ and } S_D^2 = \frac{1}{N-1} \sum_{i=1}^{N} \left(D_i - \overline{D}\right) \tag{1.16}$$

Horizontal agreement or tolerance intervals (formulae given in a further chapter) can be computed around $\overline{D}$ and ideally such intervals are included inside the acceptance interval $[-\Delta, \Delta]$. In that case, it can be concluded that the differences between the two devices are likely to be not clinically important and both methods are interchangeable. Otherwise, the differences can exceed the acceptable threshold and the devices cannot be interchangeable. If the

differences D_i increase or decrease with respect to M_i, then a regression line must be estimated and predictive or agreement intervals computed around the regression line.

Chapter 2

Errors-in-variables regressions in a (X, Y) plot

Contents

Various types of regression exist to deal with method comparison problem [31] and many papers were published these last decades with some by well-known statisticians as Bartlett, Wald,... [29–35]. A historical overview of the dedicated literature is given by Gillard and Iles [36]. However, the papers dealing with replicated data are rarer [21,22].
In this chapter, we propose to review, standardize the notations and summarize the equations for estimating the regression line of the homoscedastic model (1.11) $Y_i = \alpha + \beta X_i + \epsilon_i$. We consider the usual parametric regressions (the OLS lines, the Orthogonal Regression, the geometric Mean Regression, the Deming Regression and the BLS regression) with a clear criterion of minimization as shown in Figure 2.1. Additionally, the non-parametric regression given by Passing and Bablok [24] will also be considered as this regression was specifically designed to deal with method comparison studies and its statistical hypotheses. The goal of this chapter is to describe these regressions under homoscedasticity, to summarize the relationships, to compare the biases of the parameter estimates of model (1.11) and the coverage probabilities with simulations. The results of the simulations are displayed in charts, instead of large data sets, according to λ which is the ratio of the measurement errors variances. Finally, the regression's techniques will be applied to real data and diagrams will be proposed and described in order to summarize the information needed to assess a strict equivalence.

2.1 Presentation of seven regression approaches under homoscedasticity

2.1.1 Vertical Ordinary Least Squares regression - OLSv

The easiest way to estimate the parameters α and β of model (1.11) under homoscedasticity is to use the very well known, widely applied and old technique of Ordinary Least Squares [37,38]: OLS. The OLSv regression minimizes the sum of squares of the vertical distances (residuals) between each point and the line as shown in Figure 2.1. OLSv regression minimizes the following criterion:

$$C_{OLSv} = \sum_{i=1}^{N} (Y_i - \hat{\alpha} - \hat{\beta} X_i)^2$$

Another way to understand this criterion is to consider that it minimizes the sum of area of squares (or circles) for which the sides (or radius) are the vertical distances between each point and the line (Figure 2.1).
The corresponding estimators for α and β are given by:

$$\hat{\beta}_{OLSv} = \frac{S_{xy}}{S_{xx}} \text{ and } \hat{\alpha}_{OLSv} = \overline{Y} - \hat{\beta}_{OLSv} \overline{X} \tag{2.1}$$

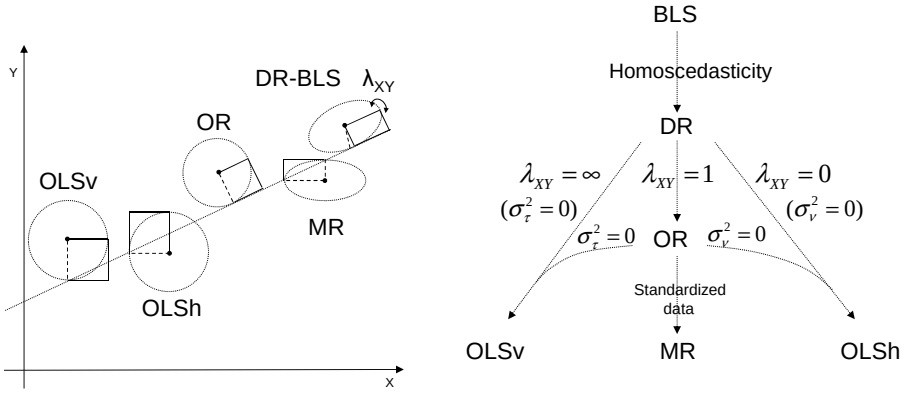

Figure 2.1: *The regression criteria (left) and relationships between them (right)*

Unfortunately, OLSv minimization criterion does not take into account the errors in the independent variable X [39]. In other words, the OLSv supposes that there is no error on the measurement method in the X-axis i.e., the τ_{ij} are supposed to be equal to zero or negligible (a rule to decide when these errors can be considered as negligible, will be given in Section 2.3). The corresponding estimates are therefore obviously biased [39] and the bias are related to the variance ratio: $\sigma_\tau^2/\sigma_\xi^2$ [40]. For large sample sizes, the estimated parameters provided by OLSv converge, as mentioned by Draper and Smith [40] or by Hartmann et al. [41], to:

$$\hat{\beta}_{OLSv} \xrightarrow{N\to\infty} \frac{1}{1+(\sigma_\tau^2/n_X)/\sigma_\xi^2}\beta \tag{2.2}$$

$$\hat{\alpha}_{OLSv} \xrightarrow{N\to\infty} \alpha + \frac{(\sigma_\tau^2/n_X)/\sigma_\xi^2}{1+(\sigma_\tau^2/n_X)/\sigma_\xi^2}\overline{X}\beta \tag{2.3}$$

2.1.2 Horizontal Ordinary Least Squares regression - OLSh

Instead of minimizing the sum of the squares of the distances between each point and the line in a vertical direction like the OLSv, these distances can be horizontally minimized (Figure 2.1) [39]. We will call this technique OLSh by analogy to the OLSv. Actually, OLSh is exactly equivalent to the OLSv when the X-axis and the Y-axis are reversed [42]. So, OLSh estimators can be computed by applying OLSv to the 'inverse' model:

$$X_i = \alpha^* + \beta^* Y_i + \epsilon_i^* \text{ with } \alpha^* = -\alpha/\beta \text{ and } \beta^* = 1/\beta \tag{2.4}$$

The corresponding OLSv estimators for α^* and β^* are:

$$\hat{\beta}^*_{OLSv} = \frac{S_{xy}}{S_{yy}} \text{ and } \hat{\alpha}^*_{OLSv} = \overline{X} - \hat{\beta}^*_{OLSv}\overline{Y}$$

And the OLSh estimators for α and β become [40]:

$$\hat{\beta}_{OLSh} = \frac{1}{\hat{\beta}^*_{OLSv}} = \frac{S_{yy}}{S_{xy}} \text{ and } \hat{\alpha}_{OLSh} = -\frac{\hat{\alpha}^*_{OLSv}}{\hat{\beta}^*_{OLSv}} = \overline{Y} - \hat{\beta}_{OLSh}\overline{X} \tag{2.5}$$

By analogy to OLSv, another way to understand this criterion is to consider that it minimizes the sum of areas of squares (or circles) for which the sides (or radius) are the horizontal distances between each point and the line (Figure 2.1).
Note that the well-known coefficient of determination R^2 (the square of the Pearson correlation coefficient) can be related to the two estimated slopes given by OLSv and OLSh [42]:

$$R^2 = \hat{\beta}_{OLSv}\hat{\beta}^*_{OLSv} = \frac{\hat{\beta}_{OLSv}}{\hat{\beta}_{OLSh}}$$

It is obvious that, if all points are aligned, there is only one line passing through these points such that $\hat{\beta}_{OLSv} = \hat{\beta}_{OLSh}$ and $R^2 = 1$ and, at the opposite, these two lines are perpendicular when $R^2 = 0$ [42]. Unfortunately, by analogy to OLSv, OLSh does not take into account the errors in the dependent variable Y. In other words, the OLSh supposes that there is no error on the measurement method in the Y-axis [39], the ν_{ik} are supposed to be equal to zero or negligible. $\hat{\alpha}_{OLSh}$ and $\hat{\beta}_{OLSh}$ are therefore also biased estimators for α and β [39].

2.1.3 The geometric Mean Regression: MR

As OLS regressions cannot tackle the problem of the errors in both axes, another regression consists of minimizing the criterion C_{MR}: the sum of area of rectangles (or ellipses) built from the projections (half-axes of ellipses) of the points to the line in parallel to the axes as shown in Figure 2.1.

$$C_{MR} = \sum_{i=1}^{N} |X_i - X(Y_i)|\,|Y_i - Y(X_i)| = \sum_{i=1}^{N} \left|X_i - \frac{Y_i}{\hat{\beta}} + \frac{\hat{\alpha}}{\hat{\beta}}\right| \left|Y_i - \hat{\alpha} - \hat{\beta}X_i\right|$$

Sprent and Dolby call this regression the geometric Mean Regression [43] (MR) because it can be easily proved that its slope is the geometric mean of the slopes given by the two OLS regressions. The MR estimators are [43]:

$$\hat{\beta}_{MR} = \sqrt{\frac{S_{yy}}{S_{xx}}}\,\text{sign}(S_{xy}) = \sqrt{\hat{\beta}_{OLSv}\hat{\beta}_{OLSh}} \text{ and } \hat{\alpha}_{MR} = \overline{Y} - \hat{\beta}_{MR}\overline{X} \tag{2.6}$$

The most common name of this regression in French is 'régression des moindres rectangles' [44] but it seems that there is no consensus in English for its name. The following names can be found in the literature: 'least triangles' [45], 'standardized principal component analysis regression' [46], 'standard major axis' [47], 'reduced major axis regression' [47], 'allometry relationship' [47] or 'ordinary least products' [48]. We will come back to the reasons of these different names. The MR was first introduced to tackle the problem of estimating a regression line between two variables when it is not clear which is the dependent and independent variable. The MR can be applied when both variables are 'interdependent' [44] which is the case in our context. The results of two measurement methods are, indeed, compared, so neither of these two methods can be considered as the response of the other. Instead of estimating the slope by the geometric mean of the two OLS slopes, some authors propose to estimate the slope by the arithmetic mean of the two OLS slopes or their bisectrix [49]. These solutions are only compromises without a clear criterion of minimization and will not be discussed in this thesis.

2.1.4 Orthogonal Regression: OR

OR minimizes the criterion C_{OR}: the sum of the squares of orthogonal distances between each point and the line (instead of a parallel direction to an axis like in OLS) [50], which is the same than minimizing the sum of area of squares (or circles) tangent to the line (Figure 2.1):

$$C_{OR} = \sum_{i=1}^{N} \left(\left(X_i - \frac{Y_i + X_i/\hat{\beta} - \hat{\alpha}}{\hat{\beta} + 1/\hat{\beta}} \right)^2 + \left(Y_i - \hat{\alpha} - \frac{\hat{\beta} Y_i + X_i - \hat{\alpha}\hat{\beta}}{\hat{\beta} + 1/\hat{\beta}} \right)^2 \right)$$

This criterion is derived from the Pythagorean theorem. The corresponding OR estimators are [50]:

$$\hat{\beta}_{OR} = \frac{S_{yy} - S_{xx} + \sqrt{(S_{yy} - S_{xx})^2 + 4S_{xy}^2}}{2S_{xy}} \text{ and } \hat{\alpha}_{OR} = \overline{Y} - \hat{\beta}_{OR}\overline{X} \qquad (2.7)$$

The OR is probably the most applied regression in the context of (linear-) errors-in-variables regressions as it is a good compromise when OLS fail because of the errors in both axes. Unfortunately, OR is only valid when variability of errors in both axes are equal [28], in other words, when $\sigma_\tau^2 = \sigma_\nu^2$ with unreplicated data or $\sigma_\tau^2/n_X = \sigma_\nu^2/n_Y$ with replicated data as we regress the averages of measures for each given sample and measurement method. The OR is also called in the literature 'total least square regression' [50] or 'principal component analysis regression' [46].

2.1.5 Deming Regression: DR

MR and OR take into account errors in both axes but not the values of their variances. Actually, only the ratio between the two error variances is needed to improve the methodology. This ratio can be defined as follows:

$$\lambda_{XY} = \frac{\sigma_\nu^2/n_Y}{\sigma_\tau^2/n_X} \tag{2.8}$$

By default, we choose to compute the ratio of the variance in the Y-axis by the variance in the X-axis because this is the most common way to compute it (Linnet computes it inversely [51–54]). Usually, in the literature, the problem of replicated data is not presented in detail, so that λ is merely defined by $\lambda = \sigma_\nu^2/\sigma_\tau^2$ as a precision ratio as mentioned by Tan and Iglewicz [28].
The Deming regression DR is the Maximum Likelihood (ML) solution of model (1.11) when λ_{XY} is known as explained by Fuller [23] or Ripley and Thompson [16]. In practice, if λ_{XY} is unknown because σ_τ^2 or/and σ_ν^2 is/are unknown, they can be estimated from replicated data and replaced by S_τ^2 or/and S_ν^2. We will denote the estimators of λ by $\hat{\lambda}$ and λ_{XY} by $\hat{\lambda}_{XY}$.
The DR minimizes the criterion C_{DR}: the sum of the square of oblique distances between each point to the line [28,51]. The angle of the direction is related to λ_{XY} and given by $-\lambda_{XY}/\hat{\beta}$ [28]:

$$C_{DR} = \sum_{i=1}^{N} \left(\left(X_i - \frac{Y_i + \lambda_{XY} X_i/\hat{\beta} - \hat{\alpha}}{\hat{\beta} + \lambda_{XY}/\hat{\beta}} \right)^2 + \left(Y_i - \hat{\alpha} - \frac{\hat{\beta} Y_i + \lambda_{XY} X_i - \hat{\alpha}\hat{\beta}}{\hat{\beta} + \lambda_{XY}/\hat{\beta}} \right)^2 \right)$$

The DR estimators are [28]:

$$\hat{\beta}_{DR} = \frac{S_{yy} - \lambda_{XY} S_{xx} + \sqrt{(S_{yy} - \lambda_{XY} S_{xx})^2 + 4\lambda_{XY} S_{xy}^2}}{2S_{xy}} \text{ and } \hat{\alpha}_{DR} = \overline{Y} - \hat{\beta}_{DR}\overline{X} \tag{2.9}$$

λ_{XY} is supposed constant in DR, an assumption fulfilled in this section since we suppose homoscedasticity. It can also be the case with heteroscedasticity if this one is identical in both axes such that the ratios of variances computed point by point are constant [52]:

$$\lambda_{XY_i} = (\sigma_{\nu_i}^2/n_{Y_i})/(\sigma_{\tau_i}^2/n_{X_i}) = \lambda_{XY} \forall i \tag{2.10}$$

The DR is also called Constant Variances Ratio regression [31].
Some other comments are interesting to provide about DR:

1. In the literature, the model is called functional when the terms ξ_i and η_i are constant and structural when these terms are random, full details are given by Fuller [23]. But, 'Any distinction between the linear functional

and structural relationships is mostly a theoretical one' as argued by Tan and Iglewicz [28]. Moreover, Jones confirms that 'both models provide the same estimates of α and β under the assumptions of normality and λ known' [42]. This is why we did not make the distinction between these two cases in Chapter 1.

2. The maximum likelihood estimation method provides five equations with six unknown parameters and therefore one of the parameters must be assumed to be known to get a solution [42]. However, the most useful (for practical or theoretical reasons [28]) situation is to consider λ_{XY} known but other cases are discussed in the literature [23,55].

3. Fuller [23] gives another formula to compute $\hat{\beta}_{DR}$ but this formula is equivalent to the one given here. The formula given by Linnet [51–54] looks also different but actually is exactly the same than ours, this is because he defines λ as $\sigma_\tau^2/\sigma_\nu^2$.

4. Mandel proposes a regression to apply in the context of inter-laboratories studies [56]. His regression is equivalent to DR but additionally it can take into account the correlation between the errors term. In the context of measurement methods, Francq [19] proved that when errors τ_{ij} and ν_{ik} are uncorrelated (and so the correlation term set to zero), his regression is equivalent to DR (see further section).

2.1.6 Bivariate Least Square regression: BLS

Bivariate Least Square regression, BLS, is a generic name but this thesis refers to the papers published firstly by Lisý et al. [57] and later by other authors [58–60]. BLS takes into account errors and heteroscedasticity in both axes and is written usually in matrix notation. We present here its formula in the case of homoscedasticity and with replicated data. BLS minimizes the criterion C_{BLS}: the sum of weighted residuals:

$$C_{BLS} = \frac{1}{W_{BLS}} \sum_{i=1}^{N} (Y_i - \hat{\alpha} - \hat{\beta} X_i)^2 = (N-2) S_{BLS}^2$$

with

$$W_{BLS} = \sigma_\epsilon^2 = \frac{\sigma_\nu^2}{n_Y} + \hat{\beta}^2 \frac{\sigma_\tau^2}{n_X}$$

Practically, the estimators of the parameters (the b vector) are computed by iterations with the following formula:

$$b = \begin{pmatrix} \hat{\alpha}_{BLS} \\ \hat{\beta}_{BLS} \end{pmatrix} = R^{-1} g \tag{2.11}$$

where

$$R = \frac{1}{W_{BLS}} \begin{pmatrix} N & \sum_{i=1}^{N} X_i \\ \sum_{i=1}^{N} X_i & \sum_{i=1}^{N} X_i^2 \end{pmatrix}$$

and

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

If σ_τ^2 or/and σ_ν^2 is/are unknown, they can be estimated with replicated data and replaced by S_τ^2 or/and S_ν^2. Moreover, even if the terms σ_τ^2 and σ_ν^2 are present separately in the formula, only the ratio λ_{XY} matters (in other words, the BLS line is the same for different values of σ_τ^2 and σ_ν^2 when λ_{XY} is constant).

2.1.7 Passing and Bablok regression: PB

Passing and Bablok propose a non-parametric regression specifically designed to deal with method comparison studies [24] where the slopes, β_{ij}^{PB}, of the straight lines between any two points, (X_i, Y_i) and (X_j, Y_j), are computed:

$$\beta_{ij}^{PB} = \frac{Y_i - Y_j}{X_i - X_j} \text{ for } 1 \leq i < j \leq n \tag{2.12}$$

Identical pairs of measurements with $X_i = X_j$ and $Y_i = Y_j$ dot not contribute to the estimation of the slope β. Any β_{ij}^{PB} with a value of -1 is also disregarded (for reasons of symmetry) as those with a value of 0 or $\pm\infty$ [24]. However, the occurrence of these special cases has a probability of zero (in theory) since X and Y are continuous variables (these cases must be discarded in practice). It follows that there are N^{PB} slopes β_{ij}^{PB} with $N^{PB} \leq C_N^2 = N(N-1)/2$. After sorting the β_{ij}^{PB}, the ranked sequence is obtained:

$$\beta_{(1)}^{PB} \leq ... \leq \beta_{(s)}^{PB} \leq ... \leq \beta_{(N^{PB})}^{PB}$$

The values of β_{ij}^{PB} are not independent, therefore, Passing and Bablok propose to compute the number of values of β_{ij}^{PB} lower than -1:

$$K = \#\{s | \beta_{(s)}^{PB} < -1\}$$

and to use K as an offset. The Passing and Bablok estimator for the slope, β, is then given by $\hat{\beta}_{PB}$, the shifted median of the slopes β_{ij}^{PB}:

$$\hat{\beta}_{PB} = \begin{cases} \beta_{((N^{PB}+1)/2+K)}^{PB} & \text{if } N^{PB} \text{ is odd,} \\ 0.5 \left(\beta_{(N^{PB}/2+K)}^{PB} + \beta_{(N^{PB}/2+1+K)}^{PB} \right) & \text{if } N^{PB} \text{ is even.} \end{cases} \tag{2.13}$$

The estimation of α requires that at least one half of the points is located above or on the regression line (and at least one half below or on the line). Therefore, the estimator of the intercept is given by $\hat{\alpha}_{PB}$ [24]:

$$\hat{\alpha}_{PB} = \text{median}\{Y_i - \hat{\beta}_{PB} X_i\} \tag{2.14}$$

This non-parametric regression given by Passing and Bablok assumes that the random error terms come from the same type of distribution, so not necessarily normal distributions. The variances, $\sigma^2_{\tau_i}$ and $\sigma^2_{\nu_i}$, do not need to be constant within the sampling range but should remain proportional [24].

2.2 Relationships between the seven regressions

All six parametric regressions are closely linked and this section summarizes the existing relationships. Figure 2.1 (right) illustrates some of them.

The non-parametric regression given by Passing and Bablok has weaker assumptions than the parametric ones. However, these assumptions are 'sufficient for reliable parameter estimations and hypothesis testing when the slope, β, is close to 1' [24].
All six parametric regressions pass through the mean point $(\overline{X}, \overline{Y})$. BLS is the most general regression and includes the five others parametric regressions. Indeed, under normality and homoscedasticity, BLS is equivalent to DR to estimate the parameters [17]. Actually, under homoscedasticity, the BLS iterative formula estimation leads to the DR solution (see the proof given by Francq [19]).
It is obvious that DR is equivalent to OR when $\lambda_{XY} = 1$ (both measurement methods have the same accuracy and $n_X = n_Y$ [28] or when $\sigma^2_\tau / n_X = \sigma^2_\nu / n_Y$) for the estimation of α and β. It can also be proved that the DR is equivalent to OLSv when $\lambda_{XY} = \infty$ ($\sigma^2_\tau = 0$, no errors in the X-axis) and to OLSh when $\lambda_{XY} = 0$ ($\sigma^2_\nu = 0$, no error in the Y-axis) [28]. The OR and MR are identical when the data are standardized and, in this case, their estimated slopes are equal to 1. It is also obvious that OR is the major axis of a PCA analysis [46] as the goal of a PCA is to find a (first) axis (a line) which maximizes the variance of orthogonal projections of the points to this axis.
Note that the OLSv (OLSh) is also the maximum likelihood solution of model 1.11 when the measurement error variance in the X-axis (Y-axis) is equal to zero. Likewise, PCA can be considered to be the maximum likelihood solution when the measurement error variances have the same normal distribution (independent and identically distributed) [61]. PCA and OR are, then, equivalent in the bivariate case (for a zero intercept). However, Wentzell et al. proposed an errors-in-variables modeling method, called the maximum likelihood principal component analysis (MLPCA) [61], based on a generalization of the classical

PCA and singular value decomposition. In MLPCA, the measurement error structure is also assumed to be known in theory (OR assumes λ_{XY} to be equal to 1) or estimable in practice. The variances can have different normal distribution and an intercept can be added to the model if needed. Actually, the random errors in MLPCA are assumed to be normally distributed around the true measurements with known variances and covariance structure in virtually any form. These error terms can, then, be correlated but the error covariance matrix must be estimable [61,62]. Therefore, MLPCA 'includes' the classical PCA and DR. As explained in Section 2.1.6, BLS regression can take into account the heteroscedasticity in both axes. Moreover, it will be explained in a further chapter of this thesis that BLS regression can take into account the correlation between the error terms if needed. It is, then, expected that MLPCA and BLS give identical estimated parameters. Nevertheless, BLS is based on a least square estimation method while MLPCA is based on maximum likelihood estimation. However, 'it has been recognized that the least squares estimation method of fitting the best straight line to data points' (with normally distributed errors) 'yields identical results for the slope and intercept as does the method of maximum likelihood' [63]. These two methodologies can also give identical results for the standard errors in slope and intercept [63]. Schuermans et al. discuss in detail the equivalence between orthogonal regression and MLPCA [64]. Note that the weighted total least square (WTLS) is a generalization of the total least square regression where appropriate diagonal scaling matrices can be considered in order to maintain consistency when the errors are independent but unequally sized [64]. Many other type of orthogonal regression (or 'total least square' (TLS) regression) are available in the literature such as the generalized TLS when the errors are also correlated, or the restricted TLS to allow the incorporation of equality constraints, the mixed LS-TLS when some of the variables are measured without error or the regularized TLS method to improve the robustness of the TLS [64]. Finally, despite the fact that PCA and linear regression are closely related or sometimes equivalent [65] as explained above, the regressions usually focus on the parameters and their CI and/or on predictive intervals while the (ML)PCA usually highlights the subspace defined by the model and the relationships among several variables.
Furthermore, note that OLS regressions, MR and OR are equivalent if the estimated line is $Y = X$. This line is, indeed, the bisectrix with a 45° angle and in that case: the vertical distances built for the OLSv between each point and this line have the same length than the horizontal distances of the OLSh and the sides of the rectangles built with the MR (which are squares in this case), and equal (at a constant) to the squares of the OR.
We explained in section 1.3.3 that there is sometimes confusion about which measurement method to put on the X-axis or Y-axis. This choice is important for OLS regressions because they are not invariant to axes switching (because OLS takes into account the errors in only one axis). At the opposite,

for OR and MR, this choice is not important because the estimated line is exactly the same whatever the coordinate system is. Indeed, the estimated slopes $\hat{\beta}_{OR,(X,Y)}$ or $\hat{\beta}_{MR,(X,Y)}$ computed in a (X, Y) axes are mathematically equal to their analogues in a (Y, X) axes such as $\hat{\beta}_{OR,(X,Y)} = 1/\hat{\beta}_{OR,(Y,X)}$ or $\hat{\beta}_{MR,(X,Y)} = 1/\hat{\beta}_{MR,(Y,X)}$. Like OR and MR, DR and BLS regressions can also take into account axes switching by adapting the value of λ_{XY}. It is also irrelevant for PB regression 'which one of the two methods is denoted X or Y' [24].

2.3 Comparison of bias estimators under equivalence

In order to compare the estimators $\hat{\alpha}$ and $\hat{\beta}$ given by the seven regressions, we simulated 100000 samples:

- with $N = 10,\ 20$ and 50
- with unreplicated data: $n_X = n_Y = 1$ ($\lambda = \lambda_{XY}$)
- under equivalence: $\alpha = 0$, $\beta = 1$, $\eta_i = \xi_i$
- the η_i were fixed at equally spaced points between 10 and 20 in order to allow the comparison of the different values of N (this comparison would be difficult with random values of η_i)
- for values of λ_{XY} given in Table 2.1 (where σ_ν^2 varies from 0.1 to 2 and σ_τ^2 inversely from 2 to 0.1 providing 13 values of λ from 0.05 to 20)

For each simulated data set, $\hat{\alpha}$ and $\hat{\beta}$ were computed for the seven regressions according to formulae given in Section 2.1, with λ known (σ_ν^2 and σ_τ^2 known). Thereafter, the mean of the 100000 values of $\hat{\alpha}$ and $\hat{\beta}$ were computed per value λ_{XY} and regression techniques and displayed with smoothed lines in Figure 2.2 with respect to λ_{XY} in logarithmic scale.

We also simulated replicated data in such a way that the λ_{XY} values are identical to those of the unreplicated case:

- with equal number of repetitions: $n_X = n_Y = 2$ and $n_X = n_Y = 3$ ($\lambda = \lambda_{XY}$, see Table 2.1)
- with unequal number of repetitions: $n_X = 4$, $n_Y = 2$ ($\lambda_{XY} = 2\lambda$, see λ_{XY} in Table 2.1)

λ is assumed unknown and estimated with the replicated data. The results obtained by replicated data are displayed in Figure 2.3 for $n_X = n_Y = 2$, Figure 2.4 for $n_X = n_Y = 3$, Figure 2.5 for $n_X = 4, n_Y = 2$.
Dotted-lines correspond to the true parameters ($\beta = 1$ or $\alpha = 0$), λ_{XY}=1 and the theoretical bias for the OLS computed by formulae (2.2) and (2.3).

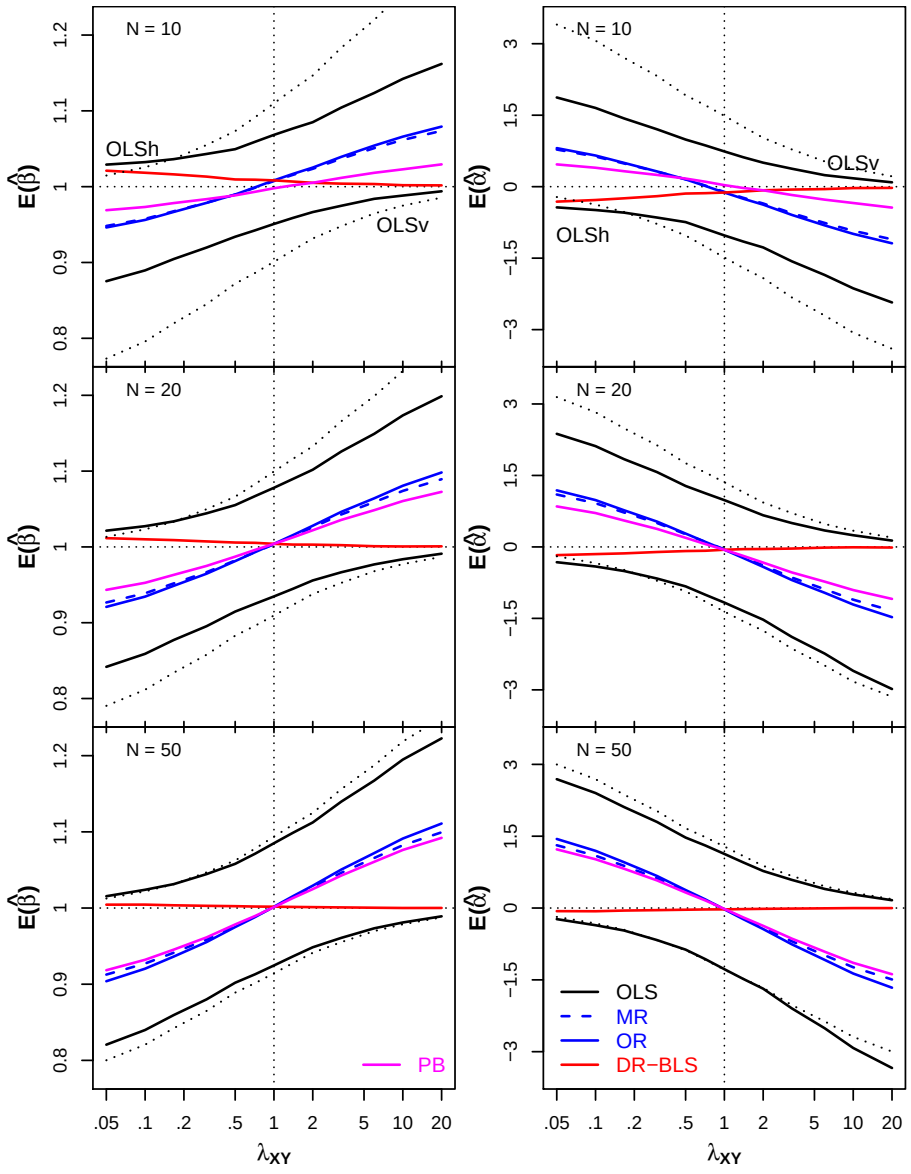

Figure 2.2: *Bias of the slope β (left) and intercept α (right) with respect to λ_{XY}, for N=10 (top), N=20 (middle) and N=50 (bottom), $n_X = n_Y = 1$*

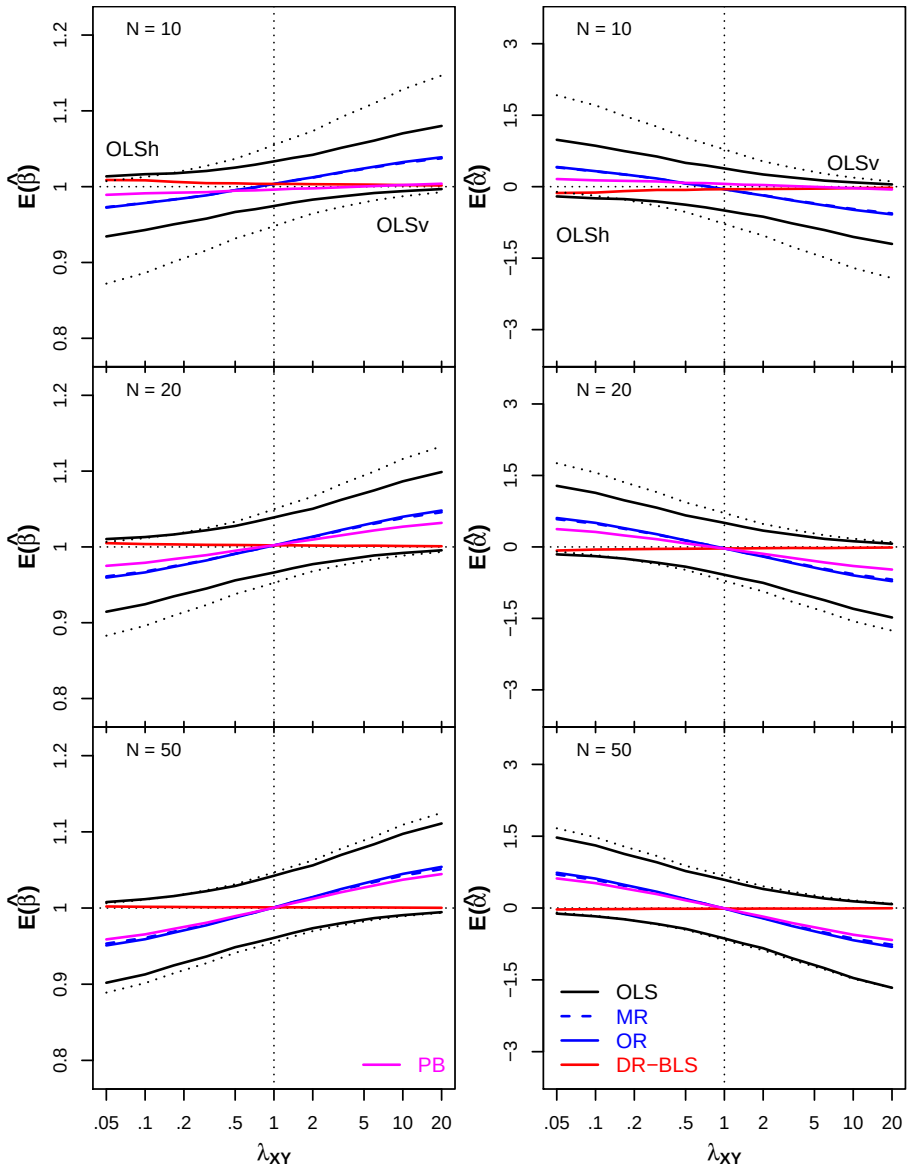

Figure 2.3: *Bias of the slope β (left) and intercept α (right) with respect to λ_{XY}, for N=10 (top), N=20 (middle) and N=50 (bottom), $n_X = n_Y = 2$*

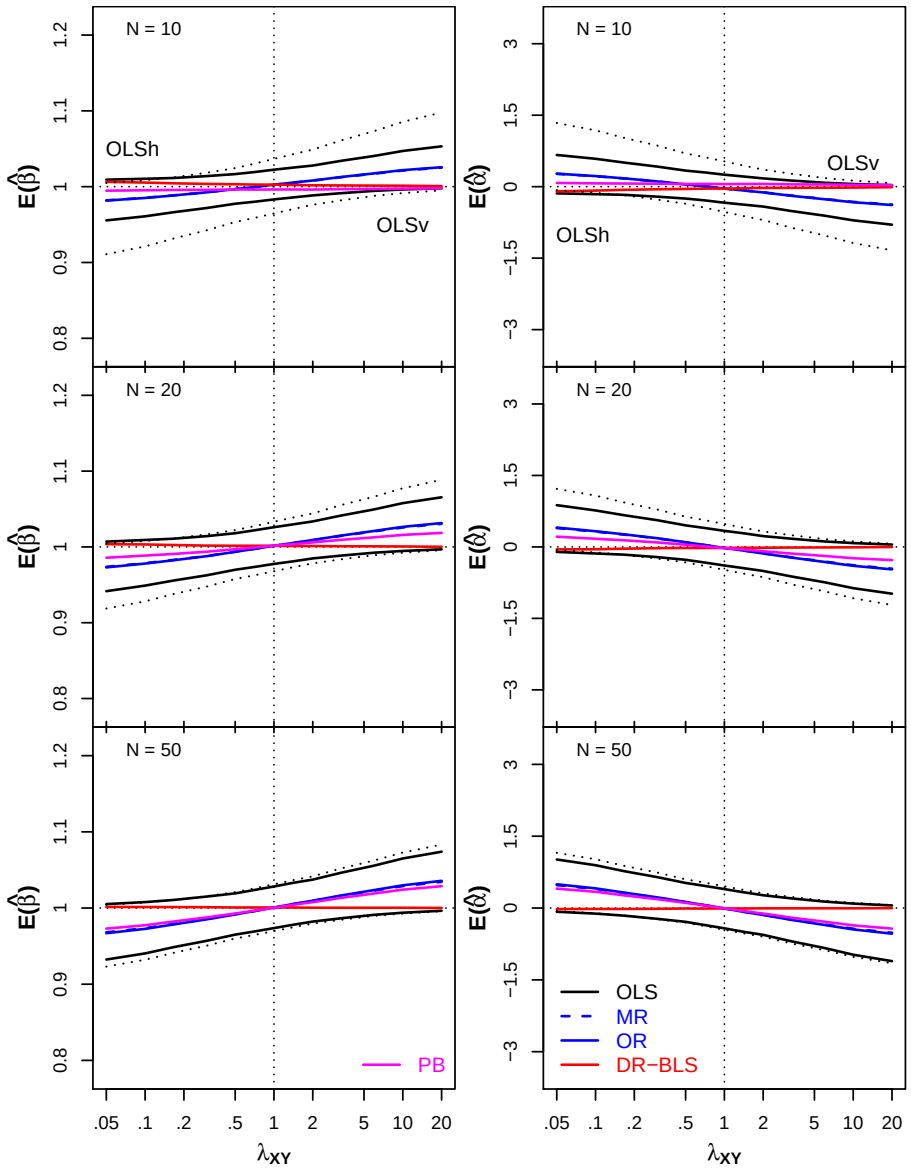

Figure 2.4: *Bias of the slope β (left) and intercept α (right) with respect to λ_{XY}, for N=10 (top), N=20 (middle) and N=50 (bottom), $n_X = n_Y = 3$*

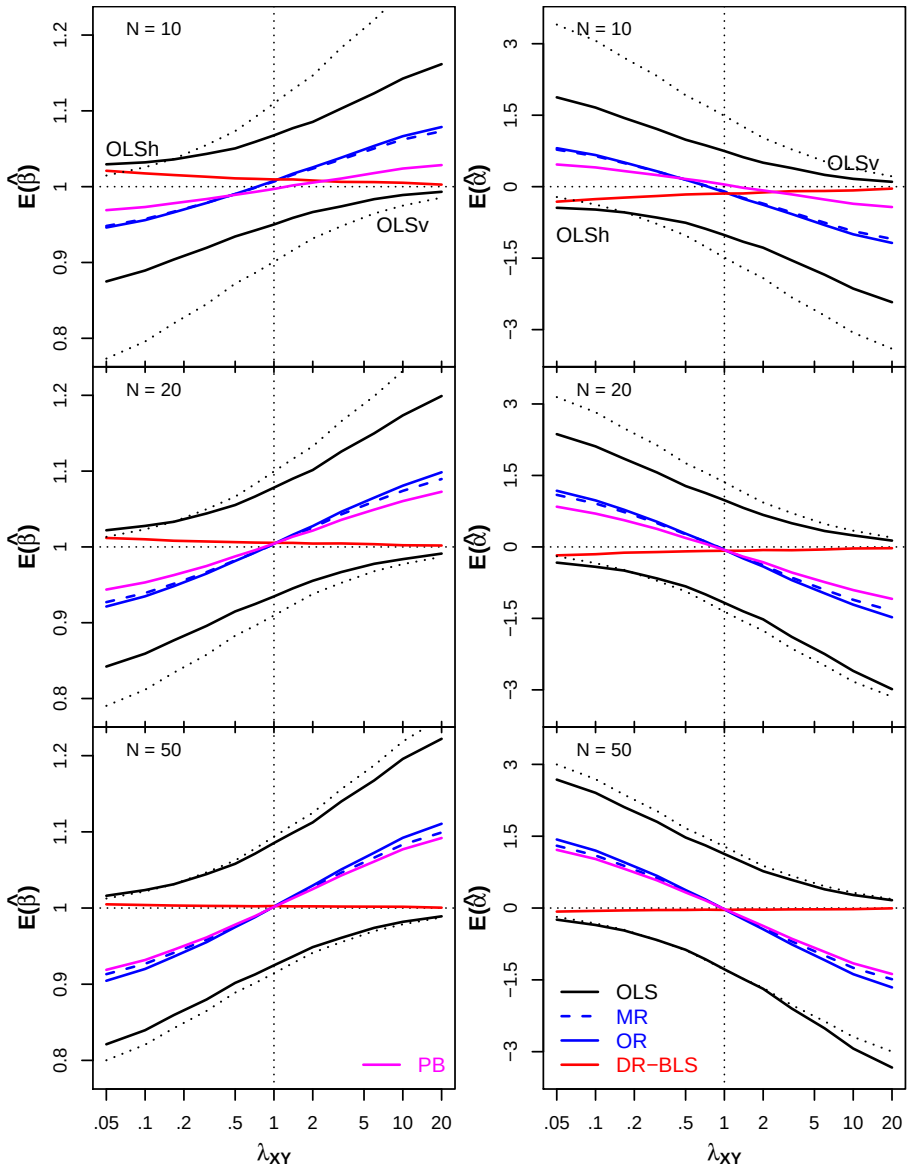

Figure 2.5: *Bias of the slope β (left) and intercept α (right) with respect to λ_{XY}, for N=10 (top), N=20 (middle) and N=50 (bottom), $n_X = 4, n_Y = 2$*

σ_ν^2	0.1	0.175	0.25	0.375	0.5	0.625		
σ_τ^2	2	1.75	1.5	1.25	1	0.875		
$\lambda = \lambda_{XY}$	0.05	0.1	0.167	0.3	0.5	0.714		
	0.75	0.875	1	1.25	1.5	1.75	2	σ_ν^2
	0.75	0.625	0.5	0.375	0.25	0.175	0.1	σ_τ^2
	1	1.4	2	3.333	6	10	20	$\lambda = \lambda_{XY}$

Table 2.1: *Values of σ_ν^2 and σ_τ^2 for the simulations with $n_X = n_Y$ and the corresponding values of λ and λ_{XY}*

We can observe in Figure 2.2 that:

- OLSv's curves move closer to the theoretical horizontal lines $\beta = 1$ or $\alpha = 0$ (the true slope or the true intercept) when λ_{XY} increases. Indeed, the bias obviously decreases for OLSv estimators when $\lambda_{XY} \to \infty$ because $\sigma_\tau^2 \to 0$ and therefore, σ_τ^2 moves closer to the assumption $\sigma_\tau^2 = 0$ of OLSv. When the sample size increases, the bias increases when the OLSv assumption is violated (the lowest values of λ_{XY}), otherwise the bias decreases when its assumption is slightly violated (the highest values of λ_{XY}). For OLSh, by analogy, the same interpretation can be provided with a symmetric effect: when λ_{XY} decreases, the OLSh assumption, $\sigma_\nu^2 = 0$, is violated and the bias decreases. The bias for the OLSv (OLSh) becomes negligible for λ_{XY} higher than 20 (lower than 1/20). The observed bias for the OLSv (and OLSh) move closer to the theoretical bias given by formulae (2.2) and (2.3) when N increases.
- the biases of OR and MR regressions are close to each other and are not biased for $\lambda_{XY} = 1$, as well as PB regression. Otherwise their biases are lower than those of OLS regressions. That's the reason why the OR (or MR) are sometimes used as a compromise instead of OLS.
- the DR and BLS estimators are identical and asymptotically unbiased (these estimators are consistent) whatever λ_{XY} (the red lines move closer to the theoretical lines when N increases) but the biases are lower for $\lambda_{XY} > 1$. That's the reason why we recommend to assign methods to both axes in such a way that λ_{XY} is higher than 1.
- when λ_{XY} is unknown and inestimable (unreplicated data), MR should be preferred to OR because it is less biased. Otherwise, DR or BLS regressions are the most suitable regressions.

With replicated data (Figure 2.3, Figure 2.4 and 2.5), all these findings are still valid: the shapes of the lines and their respective positions are identical but the value of λ_{XY} must be adapted. Moreover, when the data are replicated, as we regress the mean measures, R^2 is closer to 1 and so all the bias are reduced (the

range of the Y-axis is narrower). In the case of replicated data with $n_X \neq n_Y$, the values of σ_ν^2 and σ_τ^2 were selected in such a way that the values of λ_{XY} are identical to those given in Table 2.1, that's the reason why Figure 2.5 is analogous to Figure 2.2.

2.4 Confidence intervals for the regression parameters

This section proposes, for each regression method presented in Section 2.1, separated and joint confidence intervals for the regression parameters α and β. These confidence intervals may be used separately or jointly to test the equivalence hypotheses (1.12) ($H_0^\alpha : \alpha = 0$ and $H_0^\beta : \beta = 1$). As for parameter's estimators, most intervals proposed here are exact or derived as faithfully as possible from existing literature. Their coverage probabilities in equivalence testing will depend of measurement errors of the two measuring devices.

2.4.1 Confidence intervals for the slope β

Confidence intervals for β are well developed in the literature and we give the formulas for the different regressions in the following sections. These CI are useful to test the null hypothesis $H_0^\beta : \beta = 1$ to check whether there is a proportional bias or not between the two devices.

CI for β in OLSv

The classical $100(1-\gamma)\%$ CI for β_{OLSv} is symmetric around $\hat{\beta}_{OLSv}$ and given by [44]:

$$CI(\beta_{OLSv}) : \hat{\beta}_{OLSv} \pm t_{1-\frac{\gamma}{2},N-2} S_{\hat{\beta}_{OLSv}} \tag{2.15}$$

$$\text{with } S_{\hat{\beta}_{OLSv}} = \sqrt{\frac{S_{OLSv}^2}{S_{xx}}} \text{ and } S_{OLSv}^2 = \frac{1}{N-2} \sum_{i=1}^{N} \left(Y_i - \hat{\alpha}_{OLSv} - \hat{\beta}_{OLSv} X_i \right)^2$$

where $t_{1-\gamma/2,N-2}$ is the $100(1-\gamma/2)\%$ percentile of a t-distribution with $N-2$ df. This CI is exact under the OLSv assumptions, i.e. model (1.11) with $\sigma_\tau^2 = 0$.

CI for β in OLSh

The CI for β_{OLSh} is asymmetric around $\hat{\beta}_{OLSh}$ and can be computed from the symmetrical OLSv CI around $\hat{\beta}^*_{OLSv}$:

$$CI(\beta_{OLSh}) : 1/\left(\hat{\beta}^*_{OLSv} \mp t_{1-\frac{\gamma}{2},N-2} S_{\hat{\beta}^*_{OLSv}} \right) \tag{2.16}$$

with $S_{\hat{\beta}^*_{OLSv}} = \sqrt{\frac{S^{2*}_{OLSv}}{S_{yy}}}$ and $S^{2*}_{OLSv} = \frac{1}{N-2}\sum_{i=1}^{N}\left(X_i - \hat{\alpha}^*_{OLSv} - \hat{\beta}^*_{OLSv}Y_i\right)^2$

This CI is exact under the OLSh assumptions, i.e. model (1.11) with $\sigma^2_\nu = 0$.

CI for β in MR

An approximate CI for β_{MR} is proposed in Dagnelie [44] and is asymmetric around β_{MR}:

$$CI(\beta_{MR}) : \hat{\beta}_{MR}\sqrt{1 + 2k_{MR} \pm \sqrt{(1 + 2k_{MR})^2 - 1}} \tag{2.17}$$

where

$$k_{MR} = t^2_{1-\gamma/2,N-2}\frac{1 - R^2}{N - 2}$$

Note that the estimated standard error of the slope $S_{\hat{\beta}_{MR}}$ does not appear in this formula but can be computed approximately, if needed, as: $S_{\hat{\beta}_{MR}} \approx S_{\hat{\beta}_{OLSv}}$ [66]. Moreover, it can be noticed that this CI collapses to $\hat{\beta}_{MR}$ when $R^2 = 1$ (like for other regressions).

CI for β in OR

An approximate CI for β_{OR} is asymmetric around $\hat{\beta}_{OR}$ and provided by Sokal and Rohlf [66]:

$$CI(\beta_{OR}) : \frac{\hat{\beta}_{OR} \pm A}{1 \mp \hat{\beta}_{OR}A} \text{ where } A = \sqrt{H/(1 - H)},\ H = \frac{F_{1-\gamma,1,N-2}}{\left(\frac{\lambda_1}{\lambda_2} + \frac{\lambda_2}{\lambda_1} - 2\right)(N - 2)} \tag{2.18}$$

with

$$\lambda_1 = \frac{S_{xx} + S_{yy} + D}{2},\ \lambda_2 = \frac{S_{xx} + S_{yy} - D}{2} \text{ and } D = \sqrt{(S_{yy} - S_{xx})^2 + 4S^2_{xy}}$$

where $F_{1-\gamma,1,N-2}$ is the $100(1-\gamma)\%$ percentile of the F-distribution with 1 and $N-2$ df. This CI can sometimes not be computable, especially for small sample sizes (H can be higher than 1 and A not computable). More details about this problem and the length of the CI can be found in Tan and Iglewicz [28]. The exact distribution of $\hat{\beta}_{OR}$ and the corresponding CI are also proposed in the literature [28,67,68] but will not give it in this section as it is a particular case of the results given in the next section.

CI for β in DR

The CI for β_{DR} can always be computed approximately with a symmetric CI or it can be computed exactly with an asymmetric CI. An estimator for the variance of $\hat{\beta}_{DR}$ can be derived by the moment method and is proposed by Gillard and Iles [69]:

$$S^2_{\hat{\beta}_{DR}} = \frac{S_{xx}S_{yy} - S^2_{xy}}{N\left(\frac{S_{xy}}{\hat{\beta}_{DR}}\right)^2} \tag{2.19}$$

Gillard and Iles do not provide a CI for β_{DR} but an approximate and symmetric CI for β_{DR} can then simply be defined by associating a t-distribution to the standard error of $\hat{\beta}_{DR}$ as the estimators provided by the maximum likelihood are asymptotically normally distributed (replacing the normal distribution by a t-distribution prevents to get CI too narrow for small sample sizes):

$$\text{Approximate-}CI(\beta_{DR}) : \hat{\beta}_{DR} \pm t_{1-\gamma/2,N-2}S_{\hat{\beta}_{DR}} \tag{2.20}$$

The exact and asymmetric CI for β_{DR} can be computed from a generalization of the exact CI for β_{OR} as explained by Tan and Iglewicz [28]:

$$\text{Exact-}CI(\beta_{DR}) : \sqrt{\lambda_{XY}}\tan(\hat{\theta}_S \pm \phi) \text{ with } \hat{\theta}_S = \arctan(\hat{\beta}_{DR}/\sqrt{\lambda_{XY}}) \tag{2.21}$$

$$\text{where } \phi = \frac{1}{2}\arcsin\left(t_{1-\frac{\gamma}{2},N-2}\frac{2}{\sqrt{N-2}}\sqrt{\frac{\lambda_{XY}(S_{xx}S_{yy} - S^2_{xy})}{(S_{yy} - \lambda_{XY}S_{xx})^2 + 4\lambda_{XY}S^2_{xy}}}\right)$$

This exact CI is not always computable as the argument of the arcsine can, sometimes, be higher than 1 (see Tan and Iglewicz for more details [28]).

CI for β in BLS

The symmetric and approximate CI for β_{BLS} can be computed by using the variance-covariance matrix of the parameters proposed by Martinez et al. [8]: $\hat{V}(b) = S^2_{BLS}R^{-1}$ where the diagonal contains the estimated variances of the parameters. The CI for β is given by the following formula:

$$CI(\beta_{BLS}) : \hat{\beta}_{BLS} \pm t_{1-\gamma/2,N-2}S_{\hat{\beta}_{BLS}} \tag{2.22}$$

CI for β in PB

The $100(1-\gamma)\%$ CI for β_{PB} proposed by Passing and Bablok is computed from the ranked sequence of the slopes, β^{PB}_{ij}, and is given by the following formulae:

$$CI(\beta_{PB}) : \beta^{PB}_{M_1+K} \leq \beta \leq \beta^{PB}_{M_2+K} \tag{2.23}$$

with

$$M_1 = (N^{PB} - C_\gamma)/2,\ M_2 = N^{PB} - M_1 + 1$$

and

$$C_\gamma = z_{1-\gamma/2}\sqrt{\frac{N(N-1)(2N+5)}{18}}$$

where $z_{1-\gamma/2}$ is the $100(1-\gamma/2)\%$ percentile of the standardized normal distribution.

2.4.2 Confidence intervals for the intercept α

The confidence interval for the intercept α is often neglected in the literature. We then only consider, in this section, the regressions OLS, OR, DR, BLS and PB. The CI on the intercept is useful to test the null hypothesis $H_0^\alpha : \alpha = 0$ to check whether there is a constant bias or not between the two devices.

CI for α in OLSv

The classical CI for α_{OLSv} is symmetric around $\hat{\alpha}_{OLSv}$ and defined as [44]:

$$CI(\alpha_{OLSv}) : \hat{\alpha}_{OLSv} \pm t_{1-\gamma/2,N-2} S_{\hat{\alpha}_{OLSv}} \tag{2.24}$$

where

$$S_{\hat{\alpha}_{OLSv}} = \sqrt{S^2_{OLSv}\left(\frac{1}{N} + \frac{\overline{X}^2}{S_{xx}}\right)}$$

This CI is exact under the OLSv assumptions, i.e. model (1.11) with $\sigma_\tau^2 = 0$.

CI for α in OLSh

The CI for α is asymmetric around $\hat{\alpha}_{OLSh}$ and can be computed from the symmetrical CI around $\hat{\alpha}^*_{OLSv}$:

$$CI(\alpha_{OLSh}) : \frac{-(\hat{\alpha}^*_{OLSv} \pm t_{1-\gamma/2,N-2} S_{\hat{\alpha}^*_{OLSv}})}{\hat{\beta}^*_{OLSv}} \tag{2.25}$$

$$\text{with } S_{\hat{\alpha}^*_{OLSv}} = \sqrt{S^{2*}_{OLSv}\left(\frac{1}{N} + \frac{\overline{Y}^2}{S_{yy}}\right)}$$

This CI is exact under the OLSh assumptions, i.e. model (1.11) with $\sigma_\nu^2 = 0$.

CI for α in OR

The CI for α_{OR} proposed in this paper is based on the Mandel's procedure [56]. His procedure, developed in the context of inter-laboratories studies, can take into account the correlation between the error terms. Under assumption

(1.3), Mandel's regression is equivalent to DR. Mandel's procedure consists in transforming the (X_i, Y_i) data into $(U_i = X_i + kY_i, V_i = Y_i - \hat{\beta}_{Mandel}X_i)$ data such that U has a very small error. The OLSv regression is then applied to the (U, V) data and transformed back into the (X, Y) axis to finally get a regression line which takes into account errors in both axis. By using $\lambda_{XY} = 1$ instead of λ in the Mandel's procedure (with uncorrelated errors), the formulas are:

$$\hat{\beta}_{Mandel} = \frac{S_{xy} + kS_{yy}}{S_{xx} + kS_{xy}} \text{ and } \hat{\alpha}_{Mandel} = \overline{Y} - \hat{\beta}_{Mandel}\overline{X} \tag{2.26}$$

$$\text{with } k = \frac{\hat{\beta}_{Mandel}}{\lambda_{XY}} = \hat{\beta}_{Mandel} = \hat{\beta}_{OR}$$

By analogy, these notations are also applied:

$$\overline{U} = \frac{1}{N}\sum_{i=1}^{N} U_i \text{ and } \overline{V} = \frac{1}{N}\sum_{i=1}^{N} V_i$$

$$S_{uu} = \sum_{i=1}^{N}(U_i - \overline{U})^2,\ S_{vv} = \sum_{i=1}^{N}(V_i - \overline{V})^2 \text{ and } S_{uv} = \sum_{i=1}^{N}(U_i - \overline{U})(V_i - \overline{V}) = 0$$

The regression line in the (U, V) axis can be estimated with the OLSv's technique:

$$\hat{\beta}^{UV}_{OLSv} = \frac{S_{uv}}{S_{uu}} \text{ and } \hat{\alpha}^{UV}_{OLSv} = \overline{V} - \hat{\beta}^{UV}_{OLSv}\overline{U}$$

Then, it can be proved that $\hat{\beta}^{UV}_{OLSv} = 0$ and $\hat{\alpha}^{UV}_{OLSv} = \hat{\alpha}_{Mandel} = \hat{\alpha}_{OR}$. The approximate CI for α_{OR} can then be computed symmetrically around $\hat{\alpha}_{OR}$ with the following formulas:

$$CI(\alpha_{OR}) : \hat{\alpha}_{OR} \pm t_{1-\gamma/2,N-2}S_{\hat{\alpha}_{OR}} \tag{2.27}$$

$$\text{with } S_{\hat{\alpha}_{OR}} = S^{UV}_{\hat{\alpha}_{OLSv}} = \sqrt{S^{2UV}_{OLSv}\left(\frac{1}{N} + \frac{\overline{U}^2}{S_{uu}}\right)} \text{ where } S^{2UV}_{OLSv} = \frac{S_{vv}}{N-2}$$

CI for α in DR

The CI for α_{DR} could be computed with the Mandel's procedure as explained in the previous section but we propose to compute an approximate and symmetrical CI based on the ML property of the estimators. The variance of $\hat{\alpha}_{DR}$ can be derived by the method of moments [69] and by taking into account the replicated data as:

$$S^2_{\hat{\alpha}_{DR}} = \overline{X}^2 S^2_{\hat{\beta}_{DR}} + \frac{\hat{\beta}^2_{DR}\sigma^2_\tau/n_X + \sigma^2_\nu/n_Y}{N} \tag{2.28}$$

By analogy to Section 2.4.1, the following approximate CI is then proposed:

$$\text{Approximate-}CI(\alpha_{DR}) : \hat{\alpha}_{DR} \pm t_{1-\gamma/2,N-2} S_{\hat{\alpha}_{DR}} \tag{2.29}$$

If not available, the variance(s) σ_τ^2 and/or σ_ν^2 can be replaced by S_τ^2 and/or S_ν^2.

CI for α in BLS

By analogy to Section 2.4.1, the CI for α_{BLS} is given by the following formula [8]:

$$CI(\alpha_{BLS}) : \hat{\alpha}_{BLS} \pm t_{1-\gamma/2,N-2} S_{\hat{\alpha}_{BLS}} \tag{2.30}$$

At the opposite of the approximate CI for α given by the DR, this one is related to λ_{XY} and not directly to σ_τ^2 and σ_ν^2 individually.

CI for α in PB

The CI for α_{PB} proposed by Passing and Bablok is computed from the CI of the slopes and is computed as follows [24]:

$$CI(\alpha_{PB}) : \text{median}\{Y_i - \beta^{PB}_{M_2+K} X_i\} \leq \alpha \leq \text{median}\{Y_i - \beta^{PB}_{M_1+K} X_i\} \tag{2.31}$$

2.4.3 Joint Confidence Interval for the intercept α and the slope β

The covariance matrix $\hat{\Sigma}$ of the estimators $\hat{\theta} = (\hat{\alpha}, \hat{\beta})'$ of the regression coefficients $\theta = (\alpha, \beta)'$ is often neglected in the literature. This section considers it for the OLS, OR, DR and BLS regressions in order to compute confidence region for θ. Passing and Bablok do not propose a joint CI based on a covariance matrix. They do not reject the null hypothesis $H_0 : \theta = (0, 1)'$ whether the two separated hypotheses H_0^α and H_0^β are both not rejected. This suggestion is actually based on a confidence rectangle and it will not be considered in this thesis as confidence rectangles are not suitable in simultaneous hypotheses tests (especially when the parameters are dependent). The confidence region is useful to test the null hypothesis (1.13) ($H_0 : \theta = (0, 1)'$) to check jointly whether there is a bias (constant and/or proportional) or not between the two devices. This hypothesis is rejected if the point $(\alpha = 0, \beta = 1)$ lies outside the confidence region. The confidence regions proposed in this section are all confidence ellipses (before possible transformation) centered on $\hat{\theta}$ and can be computed with the following formula:

$$(\hat{\theta} - \theta)' \hat{\Sigma}^{-1} (\hat{\theta} - \theta) < 2F_{1-\gamma,2,N-2} \text{ where } \hat{\Sigma} = \begin{pmatrix} S^2_{\hat{\alpha}} & S_{\hat{\alpha}\hat{\beta}} \\ S_{\hat{\alpha}\hat{\beta}} & S^2_{\hat{\beta}} \end{pmatrix} \tag{2.32}$$

This common formula, derived from OLSv literature, is not exact in most situations but provides a practical solution with good coverage probabilities

as explained in further sections. $S^2_{\hat{\alpha}}$ and $S^2_{\hat{\beta}}$ are provided in Sections 2.4.1 and 2.4.2 for each method. This section provides only $S_{\hat{\alpha}\hat{\beta}}$.

Joint-CI for θ in OLSv

The covariance $S_{\hat{\alpha}\hat{\beta}_{OLSv}}$ is given by the following formula [44]:

$$S_{\hat{\alpha}\hat{\beta}_{OLSv}} = -\overline{X}\frac{S^2_{OLSv}}{S_{xx}} = -\overline{X}S^2_{\hat{\beta}_{OLSv}} \tag{2.33}$$

The corresponding confidence ellipse is exact under the OLSv assumptions, i.e. model (1.11) with $\sigma^2_\tau = 0$.

Joint-CI for θ in OLSh

Based on the inverse model (2.4), the covariance $S_{\hat{\alpha}^*\hat{\beta}^*_{OLSv}}$ is given by the following formula:

$$S_{\hat{\alpha}^*\hat{\beta}^*_{OLSv}} = -\overline{Y}\frac{S^{2*}_{OLSv}}{S_{yy}} = -\overline{Y}S^2_{\hat{\beta}^*_{OLSv}} \tag{2.34}$$

In the (Y, X) coordinate system, the null hypothesis $H_0 : \theta = (0, 1)'$ becomes $H^*_0 : \theta^* = (0, 1)'$ and the confidence ellipse becomes $(\hat{\theta}^* - \theta^*)'\hat{\Sigma}^{-1}_*(\hat{\theta}^* - \theta^*) < 2F_{1-\gamma,2,N-2}$. This confidence region is exact under the OLSh assumptions, i.e. model (1.11) with $\sigma^2_\nu = 0$.

Joint-CI for θ in OR

The confidence region is based on the Mandel's procedure [56] by applying the OLSv's technique in the (U, V) axes. The confidence region built with the (U, V) axes can be moved back to the (X, Y) axes [41] but this section tests the joint hypothesis in the (U, V) axes. The null hypothesis $H_0 : \theta = (0, 1)'$ becomes $H^{UV}_0 : \theta^{UV} = (0, (1 - \hat{\beta}_{OR})/(1 + \hat{\beta}_{OR}))'$ in the (U, V) space and the confidence ellipse becomes $(\hat{\theta}^{UV} - \theta^{UV})'\hat{\Sigma}^{-1}_{UV}(\hat{\theta}^{UV} - \theta^{UV}) < 2F_{1-\gamma,2,N-2}$. The variances of the parameters can be computed with the OLSv's technique, see Section 2.4.2, and additionally:

$$S^{2UV}_{\hat{\beta}_{OLSv}} = \frac{S^{2UV}_{OLSv}}{S_{uu}} \text{ and } S^{2UV}_{\hat{\alpha}\hat{\beta}_{OLSv}} = -\overline{U}S^{2UV}_{\hat{\beta}_{OLSv}} \tag{2.35}$$

Joint-CI for θ in DR

The covariance $S_{\hat{\alpha}\hat{\beta}_{DR}}$ can be derived by the method of moments [69] and is given by the following formula [69]:

$$S_{\hat{\alpha}\hat{\beta}_{DR}} = -\overline{X}S^2_{\hat{\beta}_{DR}} \tag{2.36}$$

This corresponding confidence region is approximate.

Joint-CI for θ in BLS

By analogy to Sections 2.4.1 and 2.4.2, the covariance $S_{\hat{\alpha}\hat{\beta}_{BLS}}$ can be extracted from the covariance matrix $\hat{\Sigma}_{BLS} = \hat{V}(b) = S^2_{BLS}R^{-1}$ [8] in order to get the confidence ellipse around $\hat{\theta}$. Martinez et al. [60] propose to display the confidence ellipse around the hypothesized values ($\alpha = 0, \beta = 1$) to slightly improve the coverage probabilities. Galea-Rojas et al. [17] discuss the disadvantages of this approximate ellipse and propose a better alternative in the case of heteroscedasticity that will be detailed in a further section.

2.4.4 Estimation and CI for λ and λ_{XY}

A CI for λ is useful to test the null hypothesis that both devices have the same precision (see the Introduction):

$$H_0 : \sigma^2_\nu = \sigma^2_\tau \equiv H_0^\lambda : \lambda = 1 \tag{2.37}$$

A CI on λ_{XY} is also useful to test the null hypothesis that the variances on the regression variables X and Y axes are equal or not:

$$H_0 : \frac{\sigma^2_\nu}{n_Y} = \frac{\sigma^2_\tau}{n_X} \equiv H_0^{\lambda_{XY}} : \lambda_{XY} = 1 \tag{2.38}$$

It is also important to assess the impact of the uncertainty of $\hat{\lambda}_{XY}$ on $\hat{\beta}_{DR}$ since λ_{XY} must be known in the DR and is, in practice, replaced by $\hat{\lambda}_{XY}$ if needed (Section 2.1.5). Moreover, one must then evaluate the impact of supposing $\lambda_{XY} = 1$ or known in choosing a given regression approach.
This section focuses on the two-sided CI on λ but the formulas can be easily adapted for λ_{XY} and for one-sided CI. λ is unknown when at least one of the variances σ^2_ν or σ^2_τ is unknown.

Estimation and CI for λ when σ^2_τ is known and σ^2_ν unknown

When σ^2_τ is known (for instance when measures of a gold-standard with known precision are provided in the X-axis) and σ^2_ν estimated by S^2_ν (see (1.7)), λ can be estimated by:

$$\hat{\lambda}_{S\sigma} = \frac{S^2_\nu}{\sigma^2_\tau} \tag{2.39}$$

As S^2_ν is an unbiased estimator of σ^2_ν (whatever the sample size) and appears on the numerator of the ratio, $\hat{\lambda}_{S\sigma}$ is also an unbiased estimator of λ. By adapting the well-known exact CI for a simple variance, the CI for σ^2_ν and for λ are:

$$CI(\sigma^2_\nu) : \left[\frac{N(n_Y - 1)S^2_\nu}{\chi^2_{N(n_Y-1),1-\gamma/2}}, \frac{N(n_Y - 1)S^2_\nu}{\chi^2_{N(n_Y-1),\gamma/2}} \right]$$

and

$$CI(\lambda) : \left[\frac{N(n_Y - 1)S_\nu^2}{\sigma_\tau^2 \chi^2_{N(n_Y-1),1-\gamma/2}}, \frac{N(n_Y - 1)S_\nu^2}{\sigma_\tau^2 \chi^2_{N(n_Y-1),\gamma/2}} \right] \tag{2.40}$$

where $\chi^2_{N(n_Y-1),1-\gamma/2}$ is the $100(1-\gamma/2)\%$ percentile of a χ^2 distribution with $N(n_Y - 1)$ df.

Estimation and CI for λ when σ_τ^2 and σ_ν^2 are unknown

When both variances σ_τ^2 and σ_ν^2 are unknown (for instance with two new measurement methods) and estimated by S_τ^2 and S_ν^2, the estimator of λ is usually given by the ratio of the estimates S_τ^2 and S_ν^2 [42]:

$$\hat{\lambda} = \frac{S_\nu^2}{S_\tau^2} \tag{2.41}$$

Unfortunately, this estimator is biased (but consistent) like the simple ratio of two independent variances [70,71]. Indeed, the expectation of $\hat{\lambda}$ can easily be computed (with $n_{X_i} = n_X$ and $n_{Y_i} = n_Y \forall i$):

$$E\left(\frac{S_\nu^2}{S_\tau^2}\right) = E\left(\frac{\sum_{i=1}^N S_{\nu_i}^2/N}{\sum_{i=1}^N S_{\tau_i}^2/N}\right) = E\left(\frac{\sum_{i=1}^N \chi^2_{N(n_Y-1)}\sigma_\nu^2/(N(n_Y-1))}{\sum_{i=1}^N \chi^2_{N(n_X-1)}\sigma_\tau^2/(N(n_X-1))}\right)$$
$$= \frac{\sigma_\nu^2}{\sigma_\tau^2} E\left(\frac{\chi^2_{N(n_Y-1)}/(N(n_Y-1))}{\chi^2_{N(n_X-1)}/(N(n_X-1))}\right) = \frac{\sigma_\nu^2}{\sigma_\tau^2} E\left(F_{N(n_Y-1),N(n_X-1)}\right) = \frac{\sigma_\nu^2}{\sigma_\tau^2}\frac{N(n_X-1)}{N(n_X-1)-2}$$

So, we propose the following unbiased estimator:

$$\hat{\lambda}_{Unb} = \frac{S_\nu^2}{S_\tau^2} k_X \text{ with } k_X = \frac{N(n_X - 1) - 2}{N(n_X - 1)} \tag{2.42}$$

where $k_X \to 1$ when N and/or n_X increase(s) as shown in Figure 2.6. From the CI of a ratio of two variances [72], the exact CI for λ is:

$$CI(\lambda) : \left[\hat{\lambda}/F_{1-\gamma/2,N(n_Y-1),N(n_X-1)},\ \hat{\lambda}/F_{\gamma/2,N(n_Y-1),N(n_X-1)}\right] \tag{2.43}$$

As mentioned in Section 2.1.5, DR supposes that λ_{XY} is known and, in practice, it is replaced by its estimate if needed. In order to compare the impact of estimators $\hat{\lambda}$ (Formula (2.41)) or $\hat{\lambda}_{Unb}$ (Formula (2.42)), on the estimator of $\hat{\beta}_{DR}$ (and $\hat{\alpha}_{DR}$) (Formula (2.9)), we simulated 100000 samples under equivalence with $n_X = n_Y = 2$ as described in Section 2.3. For each simulated data set, $\hat{\beta}_{DR}$ and $\hat{\alpha}_{DR}$ were computing according to DR formulas (2.9) with λ_{XY} known and also by replacing λ_{XY} by $\hat{\lambda}$ or $\hat{\lambda}_{Unb}$. The obtained averages of estimated slopes $\hat{\beta}_{DR}$ and estimated intercepts $\hat{\alpha}_{DR}$ are displayed in Figure 2.7.
We can notice that replacing λ_{XY} by $\hat{\lambda}$ provides a better estimation of $\hat{\beta}_{DR}$ and $\hat{\alpha}_{DR}$ as they are closer to the true parameters. When N increases, obviously,

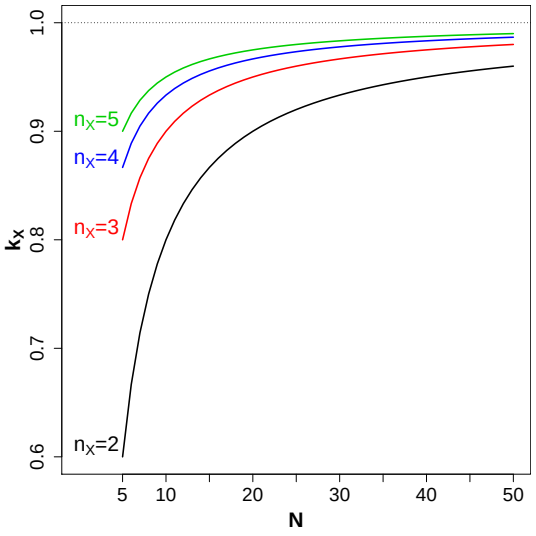

Figure 2.6: *Effect of sample size N and number of replicates n_X on k_X*

$\hat{\lambda}_{Unb} \to \hat{\lambda}$ and when they are plugged into formulas (2.9), they then provide nearly equal estimates of $\hat{\beta}_{DR}$ and $\hat{\alpha}_{DR}$ (the two curves move closer to each other when N increases). Moreover, remind that estimators (2.9) are consistent and, therefore, move closer to β_{DR} and α_{DR}.
However, $\hat{\lambda}_{Unb}$ remains a better estimator than $\hat{\lambda}$ in order to simply estimate the precision ratio of two devices.

Estimation and CI for λ when σ_ν^2 is known and σ_τ^2 unknown

When σ_ν^2 is known and σ_τ^2 estimated by S_τ^2 (Section 1.2), we propose to estimate λ with the following unbiased estimator (the proof is analogous to the one of previous section, with the expectation of an Inverse Chi-square distribution):

$$\hat{\lambda}_{\sigma S} = \frac{\sigma_\nu^2}{S_\tau^2} k_X \tag{2.44}$$

The exact CI for σ_τ^2 and λ are:

$$CI(\sigma_\tau^2) : \left[\frac{N(n_X - 1)S_\tau^2}{\chi^2_{N(n_X-1),1-\gamma/2}}, \frac{N(n_X - 1)S_\tau^2}{\chi^2_{N(n_X-1),\gamma/2}} \right]$$

and

$$CI(\lambda) : \left[\frac{\sigma_\nu^2 \chi^2_{N(n_X-1),\gamma/2}}{N(n_X - 1)S_\tau^2}, \frac{\sigma_\nu^2 \chi^2_{N(n_X-1),1-\gamma/2}}{N(n_X - 1)S_\tau^2} \right] \tag{2.45}$$

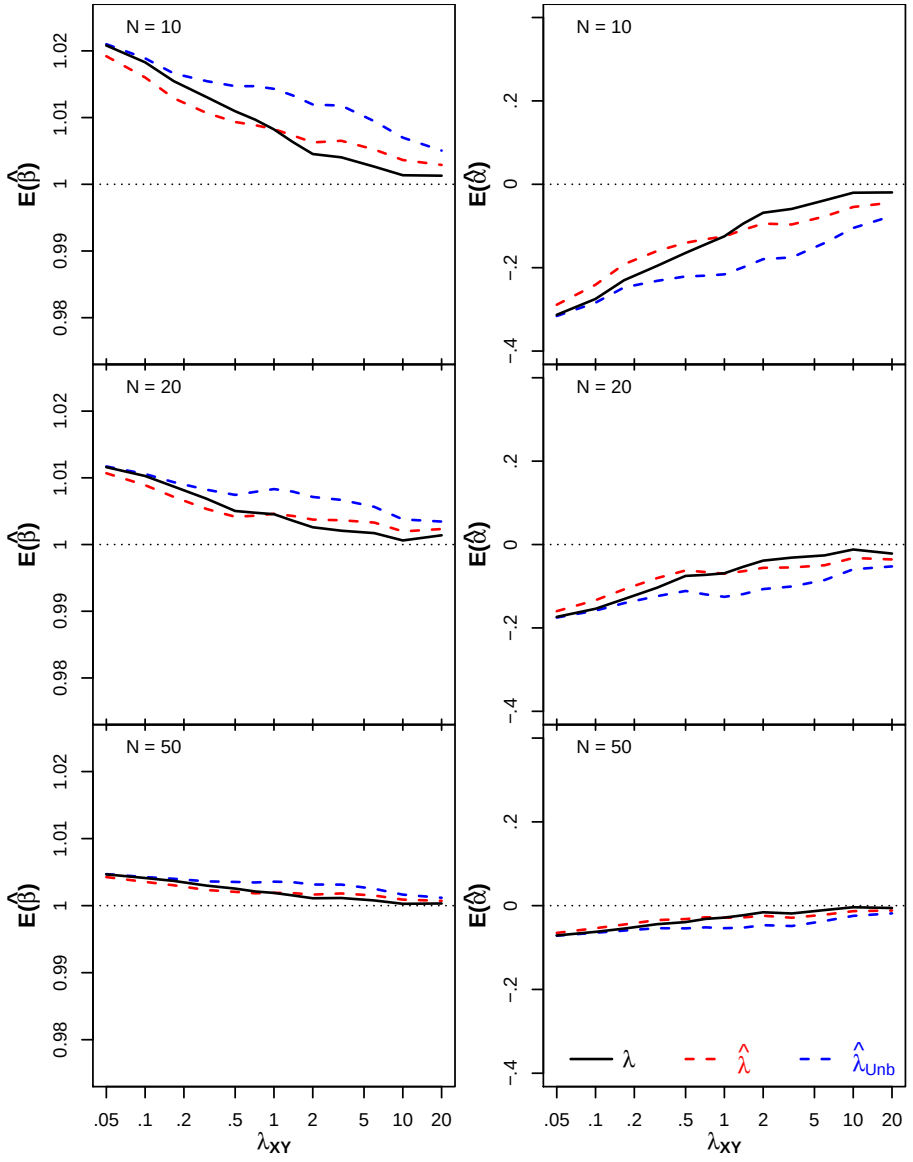

Figure 2.7: *Bias of the slope β_{DR} (left) and intercept α_{DR} (right) with respect to λ_{XY}, for N=10 (top), N=20 (middle) and N=50 (bottom), $n_X = n_Y = 2$, with λ_{XY} known or when replacing λ_{XY} by $\hat{\lambda}$ or $\hat{\lambda}_{Unb}$*

λ_{XY} unknown and inestimable because of unreplicated data

We explained in Section 2.3 that MR should be applied when λ_{XY} is unknown and inestimable but, unfortunately, we did not find in the literature a confidence interval for α or a confidence region. To tackle this problem and find these CI for the MR, we can use the property that the BLS is the most general regression (Section 2.2) and includes DR. Therefore, the estimated slope given by the MR $\hat{\beta}_{MR}$ can be plugged into formulas (2.9), the slope given by DR, to find the corresponding value of λ_{XY} called, here, λ_{MR}. In other words, λ_{MR} is the value of λ_{XY} such that $\hat{\beta}_{MR} = \hat{\beta}_{DR}$ (and $\hat{\alpha}_{MR} = \hat{\alpha}_{DR}$) and is estimated by $\hat{\lambda}_{MR}$:

$$\hat{\lambda}_{MR} = \frac{\hat{\beta}_{MR}\left(S_{yy} - \hat{\beta}_{MR}S_{xy}\right)}{\hat{\beta}_{MR}S_{xx} - S_{xy}} \tag{2.46}$$

Then, $\hat{\lambda}_{MR}$ can be plugged into formula (2.11), the BLS estimators (with, for instance, $\sigma^2_{\nu_{MR}} = \hat{\lambda}_{MR}$ and $\sigma^2_{\tau_{MR}} = 1$) to compute the CI according to the BLS formulas.
However, in Section 2.6, we will give some guidelines and describe charts, which are more useful to assess equivalence with λ_{XY} unknown and inestimable.

2.5 Comparison of the coverage probabilities of the different CI

2.5.1 Coverage probabilities of the different CI under equivalence, H_0

In order to compare the separated and joint confidence intervals for α and β provided by the different regressions, we simulated 100000 samples under equivalence with unreplicated data (and λ_{XY} supposed known) as described in section 2.3. However, the N values of η_i are, here, selected randomly in a Uniform distribution $U(10, 20)$ for each simulated sample. The coverage probabilities were computed (at a nominal level = 95%) per value of λ_{XY} and displayed in Figure 2.8 for $n_X = n_Y = 1$.
The coverage probabilities obtained by replicated data (with λ_{XY} unknown) are displayed in Figure 2.9 for $n_X = n_Y = 2$, Figure 2.10 for $n_X = n_Y = 4$ and Figure 2.11 for $n_X = 4, n_Y = 2$ (where the values of σ^2_τ and σ^2_ν were chosen in such a way that the values of λ_{XY} are identical to those of the unreplicated case).

We can notice in Figure 2.8 that the coverage's probabilities (for α, β and their confidence region) for the OLSv move closer to 95% when λ_{XY} increases which is obvious because we move closer to the OLSv assumption. For OLSh, it

is exactly the same with a symmetric effect: the coverage's probabilities move closer to 95% when λ_{XY} decreases.
For the slope, we can observe that the coverage's probabilities for OR and MR are close to each other and close to 95% when λ_{XY} is close to 1, otherwise the coverage's probabilities collapse, as well for PB. Obviously, the exact CI for β provided by the DR is excellent whatever λ_{XY}. Unfortunately, we did not find in the literature exact formulas for the CI for α and the joint-CI. The approximate CI given for DR have nearly constant coverage probabilities whatever λ_{XY} is and outperforms BLS when N increases ($N = 50$). Otherwise, BLS is better for small N ($N = 10$) or middle N ($N = 20$) with $\lambda_{XY} > 1$. The coverage's probabilities of BLS (α, β and their joint-CI) are close to 93% when $\lambda_{XY} < 1$ and close to 95% when $\lambda_{XY} > 1$.
In the case of replicated data, all these findings are still valid but the value of λ_{XY} must be adapted.

2.5.2 Coverage probabilities of the different CI under non-equivalence, H_1

In order to assess the power of the hypotheses tests, based on confidence intervals for α and β provided by the consistent regressions DR and BLS, we simulated 10000 samples under non-equivalence with unreplicated or replicated data with $n_X = n_Y = 2$, as described in Section 2.5.1. However, λ_{XY} was set to 1 (see $\lambda_{XY} = 1$ in Table 2.1) by default. To simulate the non-equivalence, firstly, α was set to 0 (no constant bias) but β ranged from 0.5 to 1.5 (by 0.1) and secondly β was set to 1 (no proportional bias) but α ranged from -5 to 5 (by 1). Figure 2.12 displays the powers of the hypotheses tests H_0^β and H_0^α for $N = 10, 20, 50$ and $n_X = n_Y = 1$ with respect to, respectively, β and α for the consistent regressions DR and BLS. The powers obtained by replicated data, $n_X = n_Y = 2$, are displayed in Figure 2.13. The power of the joint confidence interval will be assessed in detail in Chapter 3.

We can notice in Figure 2.12 that the powers are very close to each other between the DR (with or without exact confidence for the slope) and BLS regressions. Obviously, the powers increase when we move away from the equivalence, H_0, and when the sample size, N, increases. With replicated data, these findings are similar (Figure 2.13).

2.6 Practical recommendations

In this section, we provide some guidelines for the practitioners.
To estimate the regression line, when the variances σ_τ^2 and σ_ν^2 are known or can be estimated by (1.7), we recommend to put on the X-axis the measurement

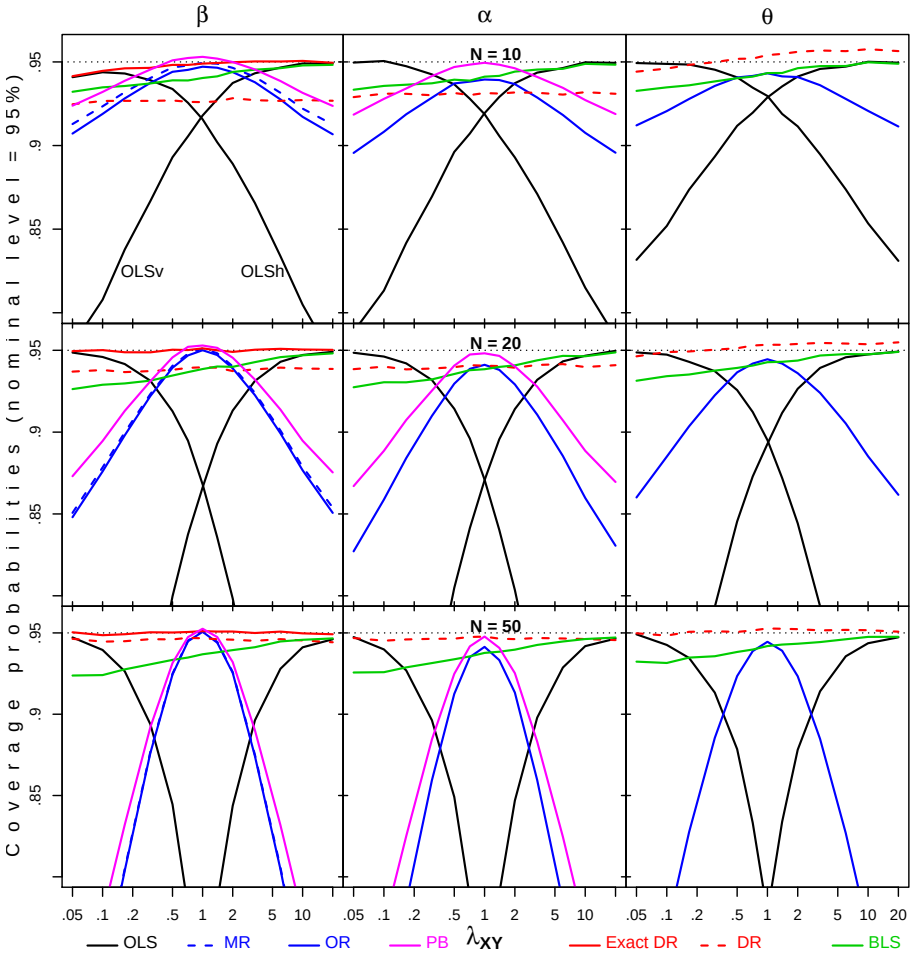

Figure 2.8: *Coverage probabilities of the CI for β (left), α (middle) and the joint-CI (right) related to λ_{XY} in a logarithmic scale, for N=10 (top), N=20 (middle) and N=50 (bottom), $n_X = n_Y = 1$*

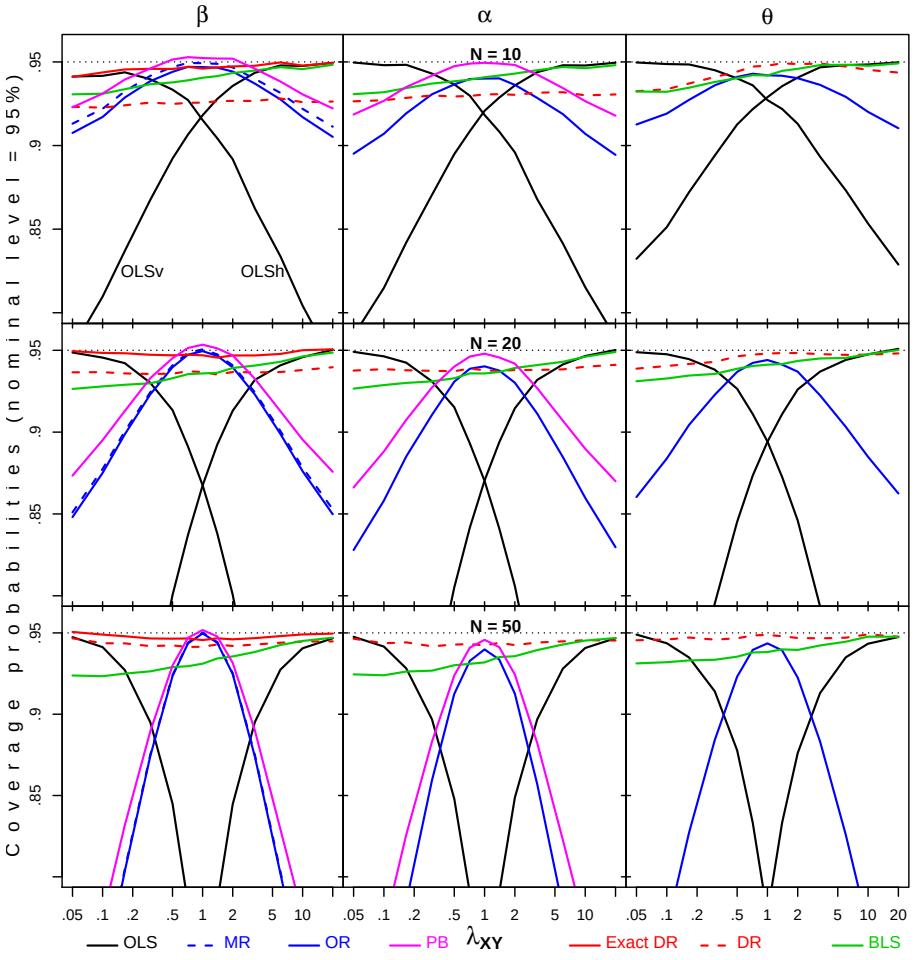

Figure 2.9: *Coverage probabilities of the CI for β (left), α (middle) and the joint-CI (right) related to λ_{XY} in a logarithmic scale, for N=10 (top), N=20 (middle) and N=50 (bottom), $n_X = n_Y = 2$*

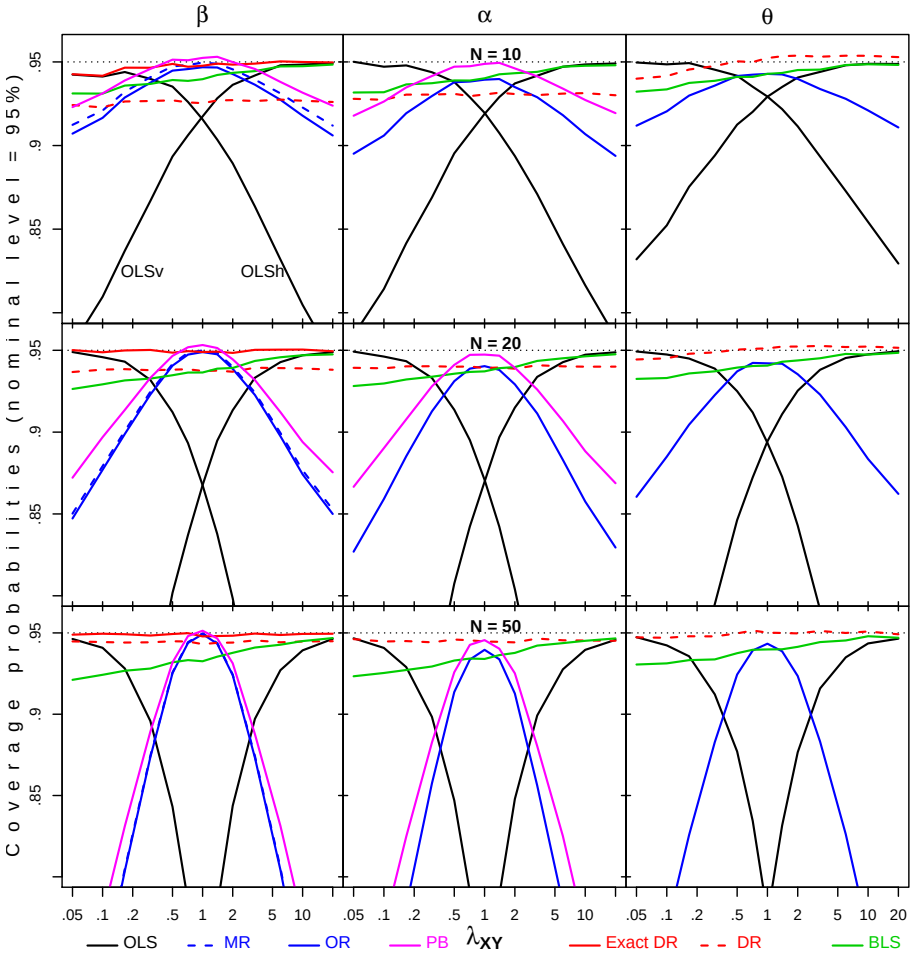

Figure 2.10: *Coverage probabilities of the CI for β (left), α (middle) and the joint-CI (right) related to λ_{XY} in a logarithmic scale, for N=10 (top), N=20 (middle) and N=50 (bottom), $n_X = n_Y = 4$*

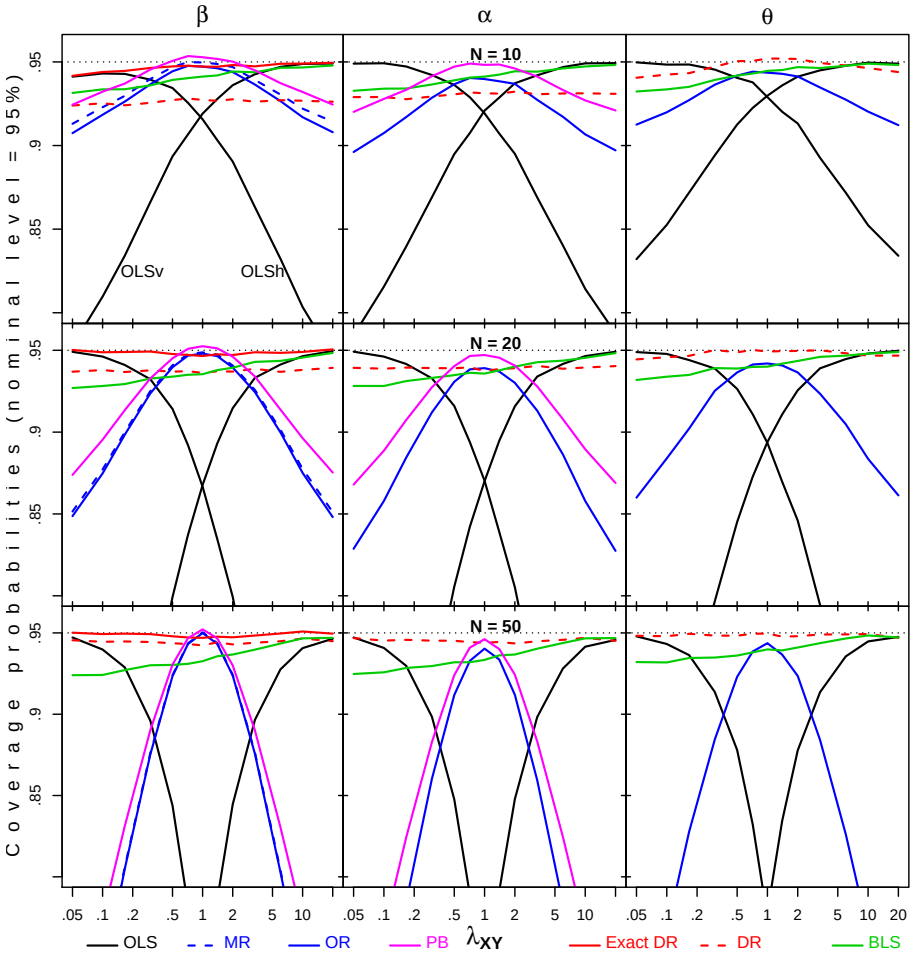

Figure 2.11: *Coverage probabilities of the CI for β (left), α (middle) and the joint-CI (right) related to λ_{XY} in a logarithmic scale, for N=10 (top), N=20 (middle) and N=50 (bottom), $n_X = 4, n_Y = 2$*

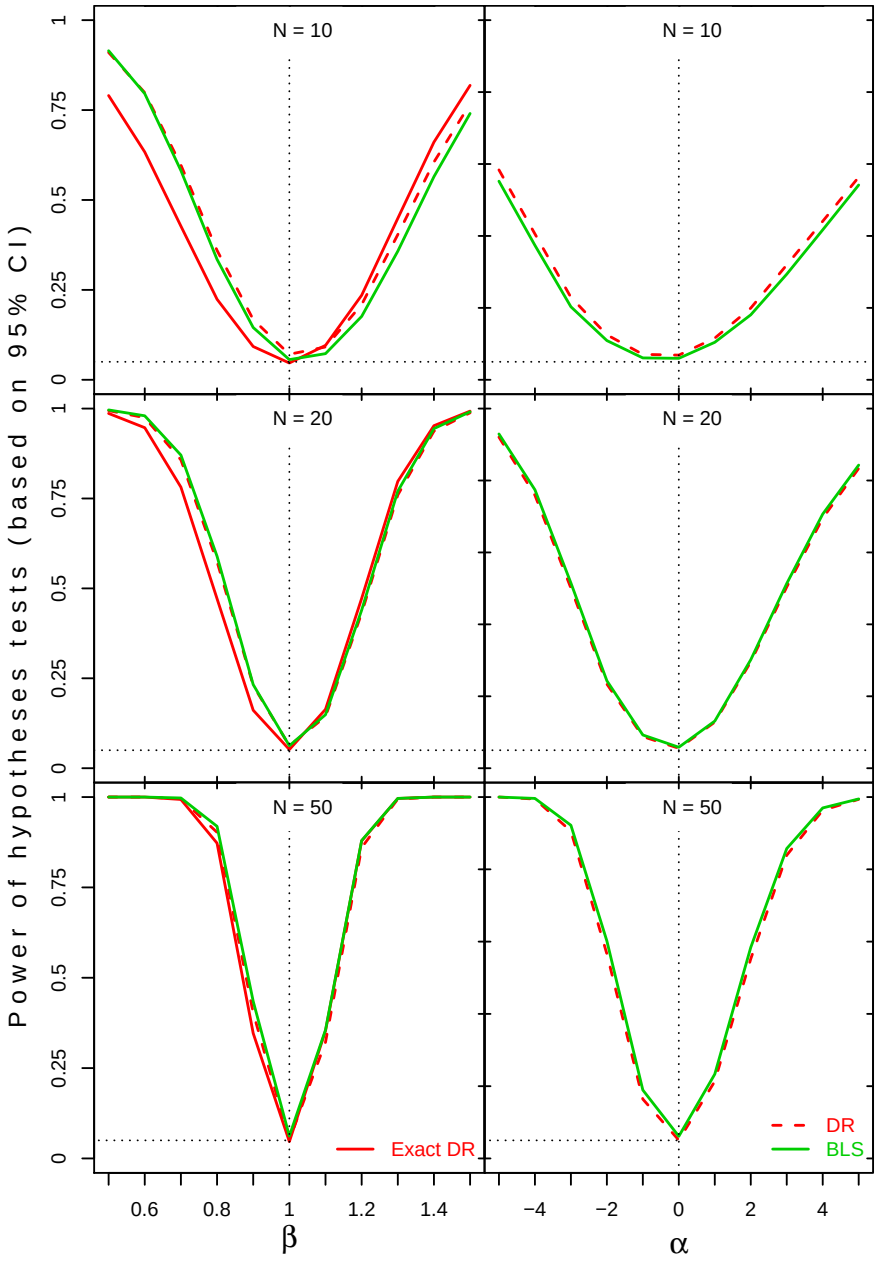

Figure 2.12: *Power of the hypotheses tests (based on CI) for β (left) and α (right), for N=10 (top), N=20 (middle) and N=50 (bottom), $n_X = n_Y = 1$*

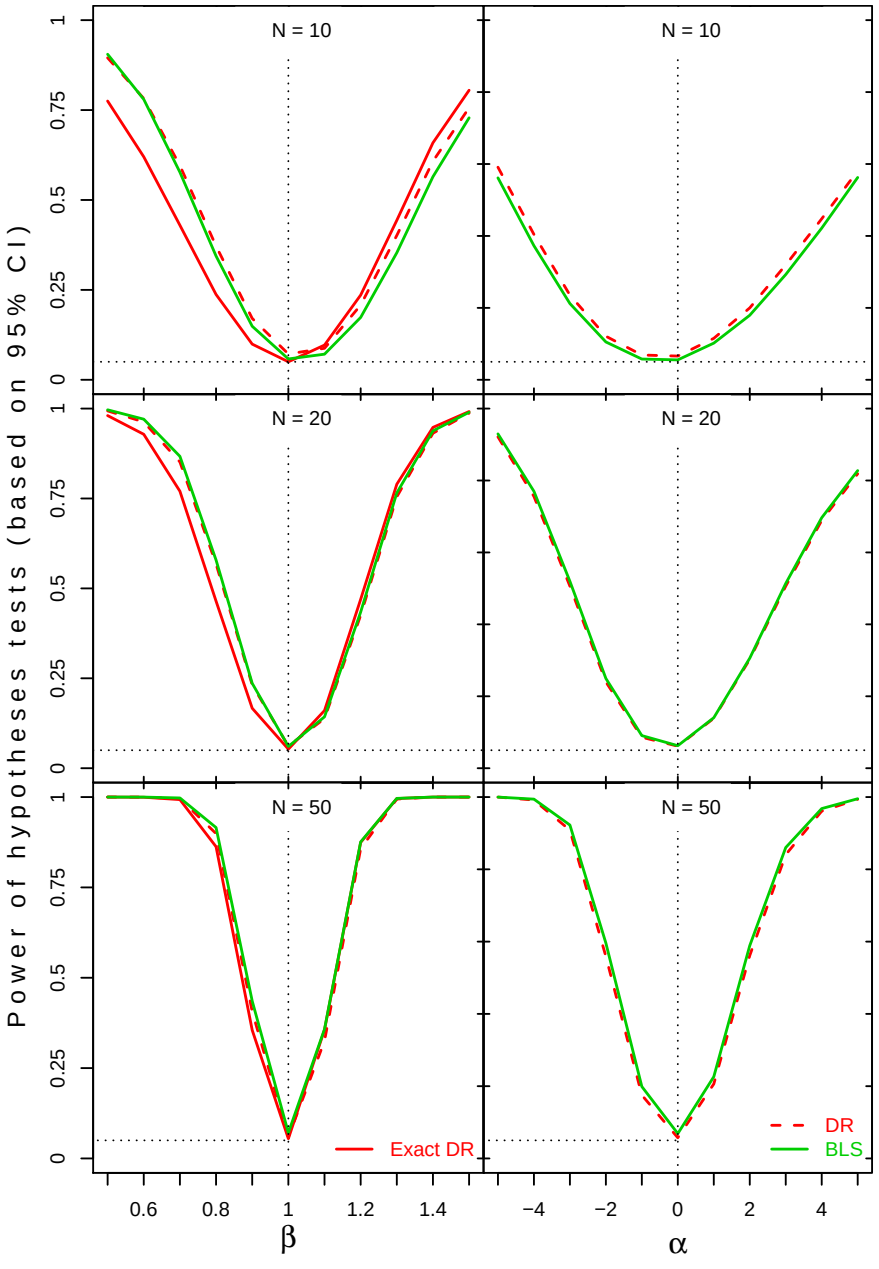

Figure 2.13: *Power of the hypotheses tests (based on CI) for β (left) and α (right), for N=10 (top), N=20 (middle) and N=50 (bottom), $n_X = n_Y = 2$*

method with the lower variance. Indeed, the biases of DR or BLS regressions (DR and BLS provide identical regression lines under homoscedasticity) are lower for $\lambda_{XY} > 1$ and the coverage probabilities better (Sections 2.3 and 2.5.1). If needed, λ can be estimated by $\hat{\lambda}$ (Formula (2.41)). To compare the precision of both measurement methods, we recommend to estimate λ by $\hat{\lambda}_{Unb}$ (estimator (2.42)). To assess the equivalence, the confidence intervals for α, β and θ can then be computed by DR or BLS (remind that DR provides an exact CI for β). The equivalence is then assessed by:

- checking whether 0 lies into the CI for the intercept α or not (no constant bias or constant bias),

- and checking whether 1 lies into the CI for the slope β or not (no proportional bias or proportional bias),

- or, ideally, checking whether (0,1) lies into the confidence region for the parameters θ or not (no bias or bias).

When λ_{XY} is unknown and inestimable, we recommend to display on a chart, for a given nominal level, the confidence intervals for α or β according to λ_{XY} in the X-axis (logarithmic scale) **from OLSh to OLSv** with the intermediate values of λ_{XY}, see example in Figure 2.14 (bottom left and top right). In practice, the CI provided by the OLSh can be displayed at $\lambda_{XY} = 0.02$ (instead of $\lambda_{XY} = 0$) and the one provided by the OLSv can be displayed at $\lambda_{XY} = 50$ (instead of $\lambda_{XY} = \infty$). The confidence regions can also be displayed on a chart from OLSh to OLSv with intermediate values of λ_{XY}. The lower and upper bounds of the intermediate separated confidence intervals can be displayed with smoothed lines and the intermediate confidence region (ellipses) can be displayed in dotted-lines or grey lines like in Figure 2.14-bottom right.
Then, such charts provide the full information about the CI and sometimes the equivalence hypothesis can be rejected (or not rejected) whatever λ_{XY}. This is very useful when λ_{XY} is unknown and inestimable because of no replicated data. When λ_{XY} is known or estimated, such charts can also be useful to provide additional information but a 'full' chart is not compulsory, it may be restricted from $\lambda_{XY} = 1$ to $\lambda_{XY} = \infty$ (50 in practice). Even in the case of λ_{XY} unknown, the chart can be restricted, for instance, from $\lambda_{XY} = 1/3$ to $\lambda_{XY} = 3$ if the practitioner knows that a device is not three times more precise than the other. To sum-up, these charts, given and explained in the next section with real data, are useful to generalize the conclusions whatever the measurement errors.

2.7 Applications

2.7.1 The systolic blood pressure - replicated data

In the systolic blood pressure data [11], simultaneous measurements were made by each of two observers (denoted J and R) using a sphygmomanometer and by a semi-automatic blood pressure monitor (denoted S). Measurements made by observer J with the manual device and those made by the semi-automatic monitor S will only be considered. The systolic blood pressure was, so, measured three times per patient (85 patients) by the semi-automatic device S and three times per patient by the observer J with the manual device. If we decide to put on the Y-axis the mean measures given by S and on the X-axis those given by J (the reason will be explained few lines below), we have: $N = 85$ ($i = 1, ..., 85$), $n_{X_i} = n_X = n_{Y_i} = n_Y = 3\ \forall i$, $\lambda = \lambda_{XY}$. The variances $\sigma^2_{\tau_i}$ and $\sigma^2_{\nu_i}$ are unknown but can be estimated with $S^2_{\tau_i}$ and $S^2_{\nu_i}$: $S^2_{\tau_1} = 14.333, ..., S^2_{\tau_{85}} = 33.333$ and $S^2_{\nu_1} = 9.333, ..., S^2_{\nu_{85}} = 13$.
Under homoscedasticity, the 'global' estimates S^2_τ and S^2_ν can be derived: $S^2_\tau = 37.408$ and $S^2_\nu = 83.141$ and λ (=λ_{XY} as $n_X = n_Y$) is estimated by $\hat{\lambda} = 2.223$ or $\hat{\lambda}_{Unb} = 2.196$. λ is higher than one, so we put on the Y-axis the mean measures given by S and on the X-axis those given by J. We have also: $\overline{X} = 127.408$, $\overline{Y} = 143.027$, $S_{xx} = 79598.750$, $S_{yy} = 84916.269$, $S_{xy} = 67200.826$ and $R^2 = 0.67$.
The estimated coefficients for the different regression's lines are given in Table 2.2.

The different regression's lines are displayed in Figure 2.14-top left with the equivalence line $Y = X$. Some points are outliers but they are not removed in didactic purposes to better distinguish the different regression lines (obviously, the most the R^2 is close to 1, the most the regression lines are close to each other). Moreover, it is not recommended to remove outliers [24] without clear and justified reasons. We can observe that the two OLS regression lines are the 'extremes' or 'outside' lines (for parametric regressions) because they assume errors in only one axis. All the parametric regression lines pass through the mean point $(\overline{X}, \overline{Y})$ and PB's line is very close to this point. All the parametric errors-in-variable regressions are 'inside' the two OLS lines. The OR and MR are very close to each other and the BLS or the DR (confounded under homoscedasticity) are the best estimated lines because the errors in both axes are taken into account the best with $\lambda_{XY} \neq 1$. PB's regression line is very close to DR or BLS lines.
In Figure 2.14-top right, the CI for β of the two OLS regressions are exact (under OLS assumptions) and the one given by the OLSh is asymmetric. We can observe that the CI given by the DR for β would coincide to the CI given by the two OLS if we extend the DR's curve on the chart until the CI of the OLS.

	$\hat{\beta}$	$\hat{\alpha}$
OLSv	0.844	35.464
OLSh	1.264	-17.968
MR	1.033	11.433
OR	1.040	10.479
DR	0.956	21.230
BLS		
PB	0.994	14.631

Table 2.2: *SBP data ($n_X = n_Y = 3$), values of $\hat{\beta}$ and $\hat{\alpha}$ for the seven regressions*

This is because the CI given by DR is also exact and includes those of the OLS. The CI given by BLS is symmetric and approximate but this approximation is close to the exact CI given by the DR when $\lambda_{XY} > 1$ while it is less and less reliable for small λ_{XY}. We also display on the chart the horizontal line $\beta = 1$ corresponding to the null hypothesis $H_0^\beta : \beta = 1$ and a vertical line at the estimated value $\hat{\lambda}$ of λ_{XY} with the corresponding 95% CI in wheat color: $CI(\lambda) = CI(\lambda_{XY})$ (as $n_X = n_Y$) : $[2.223F_{0.025,170,170}, 2.223F_{0.975,170,170}] = [1.644, 3.005]$.

Note that the CI for λ is, in practice, more useful than the one for λ_{XY} but both CI are equivalent when $n_X = n_Y$. When $n_X \neq n_Y$, the CI for λ cannot be displayed on such chart as the X-axis refers to λ_{XY}.

As the lines $\beta = 1$ and $\hat{\lambda} = 2.196$ cross inside the curves of DR or BLS confidence intervals, the null hypothesis of no proportional bias is not rejected meaning that the slope $\hat{\beta}_{DR} = \hat{\beta}_{BLS} = 0.956$ is not significantly different of 1. This is still valid from the lower to the upper bounds of the CI for λ_{XY}.

The null hypothesis that both devices have the same accuracy is rejected because 1 is not included inside the CI for λ.

Similarly, on Figure 2.14-bottom left, the null hypothesis of no constant bias, $H_0^\alpha : \alpha = 0$, is rejected at $\hat{\lambda}_{XY}$ but is borderline on the lower bound of its CI.

Figure 2.14-bottom right shows the confidence region from the OLSv to OLSh. The confidence ellipse is exact for OLSv. The other ellipses are computed approximately by BLS for different values of λ_{XY} (we could also display the ellipses computed approximately by DR). Obviously, the approximation is better when we move closer to the OLSv, so when $\lambda_{XY} > 1$. As the 'equivalence point' $(\alpha = 0, \beta = 1)$ is outside the approximate red ellipse given by BLS at $\hat{\lambda}_{XY}$, the null hypothesis (1.13) (no bias between the two devices) is rejected. In other words, the manual and the semi-automatic devices are not equivalent according to the joint hypothesis given in Section 1.3.3.

Moreover, the advantages of such charts is that the conclusion can be generalized whatever the value of λ_{XY}. Indeed, in this case, the equivalence point is outside all the ellipses from OLSv to OLSh meaning that the null hypothesis is rejected whatever the measurement errors.
Remind that BLS includes the other regressions as explained in Section 2.2 to estimate the regression parameters of model (1.11) which encourages to choose it in priority. The confidence intervals are, otherwise, computed differently even if the confidence region presented in Section 2.4.3 are all ellipses. In particular, the exact confidence region given by OLSh (under OLSh assumptions, i.e. model (1.11) with $\sigma_\nu^2 = 0$) is an ellipse based on the inverse model (2.4). But, when this confidence ellipse is moved back into the initial (X, Y) coordinate system, his shape is slightly distorted while the BLS confidence region is always a confidence ellipse.

Figure 2.15-right is a zoom of Figure 2.14-bottom right where the exact OLSh confidence region with a distorted shape is superposed to the approximate confidence ellipse computed by BLS under OLSh assumption, $\sigma_\nu^2 = 0$ ($\sigma_\nu^2 = 0.000001$ for instance in practice to avoid that BLS formulas are not computable). Although the exact distorted ellipse of OLSh and the approximate confidence ellipse of BLS are not, obviously, identical, we can notice that they are very similar.
Remind that the coverage probabilities provided by BLS are better when $\lambda_{XY} > 1$ but even for $\lambda_{XY} < 1$ they are still good by being close to 93% as mentioned in Section 2.5.1. Figure 2.15-left is a zoom of Figure 2.14-bottom right where the exact OLSv confidence ellipse is superposed to the confidence ellipse computed by BLS under OLSv assumption ($\sigma_\tau^2 = 0$). We can notice, obviously, that the confidence ellipse provided by BLS under OLSv assumption superposes perfectly well the exact confidence ellipse given by OLSv.
To sum-up, the confidence region is exactly an ellipse under OLSv assumptions, while this confidence ellipse slightly distorts with errors in both axes to get a 'maximal distortion' under OLSh assumptions. But, the approximate confidence ellipse given by BLS still coincides very well with the exact distorted ellipse given by OLSh.

2.7.2 The systolic blood pressure - unreplicated data with λ unknown

To deal with unreplicated data, the systolic blood pressure data is used for didactic purposes by considering only the two first measurements given by the sphygmomanometer (in order to get an example without outlier). In other words, the systolic blood pressure is supposed to be measured two times per patient with the sphygmomanometer (observer J) and the equivalence between the first measurements and the second measurements from the same device is

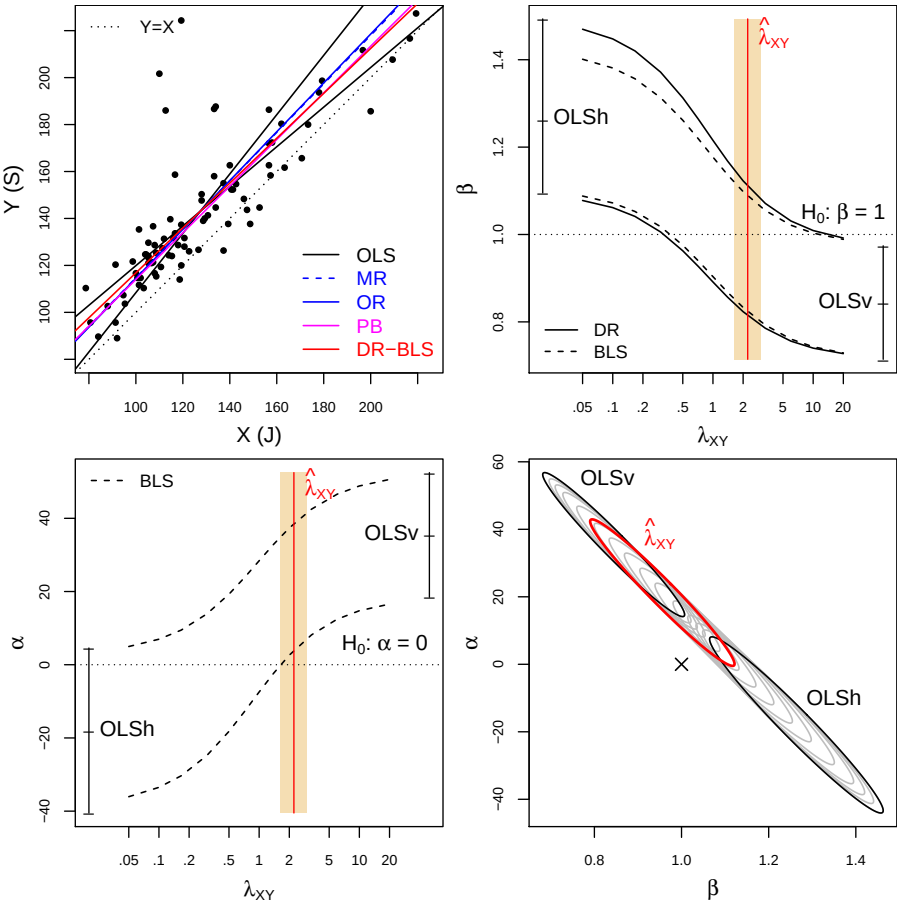

Figure 2.14: *SBP data ($n_X = n_Y = 3$), the different regressions lines (Top-left) and their CI for the parameters (Top-right for β, Bottom-left for α and Bottom-right for the joint-CI)*

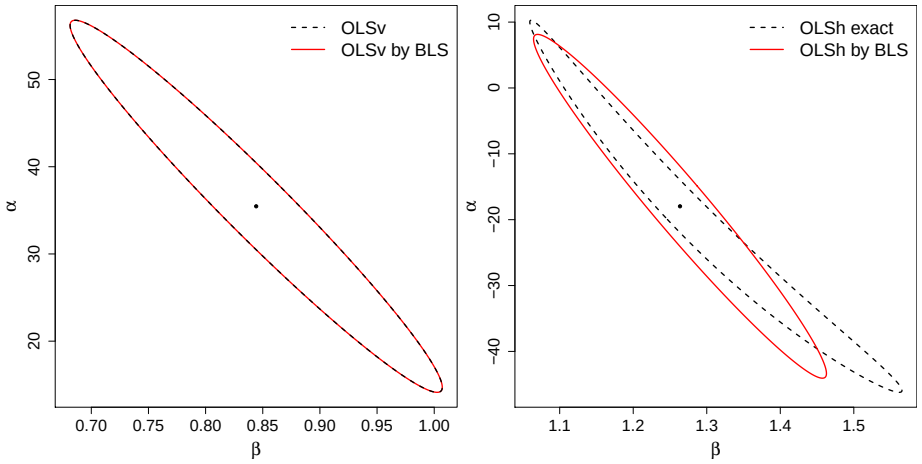

Figure 2.15: *The OLS confidence ellipses superimposed to the BLS ellipses under OLSv assumption on the left and OLSh assumption on the right*

tested. If we decide to put on the X-axis the first measures and on the Y-axis the second measures, we have: $N = 85$ $(i = 1, ..., 85)$, $n_{X_i} = n_X = n_{Y_i} = n_Y = 1$ $\forall i$. Since measures are taken with the same device, we expect of course not to reject the equivalence.
The variances $\sigma^2_{\tau_i}$ and $\sigma^2_{\nu_i}$ are unknown and cannot be estimated because of no replicated data (but we still suppose the homoscedasticity), so λ_{XY} is also unknown and inestimable. We have: $\overline{X} = 128.541$, $\overline{Y} = 127.294$, $S_{xx} = 83179.11$, $S_{yy} = 83557.65$, $S_{xy} = 80428.47$ and $R^2 = 0.93$.
The estimated coefficients for the different regression's lines are given in Table 2.3. Because of rounded numbers (and unreplicated data, no average is computed), many slopes β^{PB}_{ij} have the same value and are equal, in this example, to 1 in the middle of their sorted sequence.

The regressions DR and BLS cannot be applied because σ^2_τ, σ^2_ν and λ_{XY} are unknown and inestimable. But, we can compute the value of λ_{XY} such that $\beta_{MR} = \beta_{DR}$ and $\alpha_{MR} = \alpha_{DR}$ (for $n_X = n_Y = 1$): $\hat{\lambda}_{MR} = 1.005$. Figure 2.16 is analogue to Figure 2.14. The OR and MR lines are very close to each other, such that it is not possible to distinguish them on the plot (the most the true slope is equal to one and the most the OR and MR are close to each other). The MR is the best estimated line by default (Sections 2.3 and 2.5.1).
The CI for α or β are displayed from OLSv to OLSh as explained in Section 2.7.1. The separated null hypotheses (1.12) $H^\alpha_0 : \alpha = 0$ and $H^\beta_0 : \beta = 1$ are not rejected whatever λ_{XY}, so whatever the applied regression. When these two hypotheses are tested jointly, it can be noticed that the 'equivalence point'

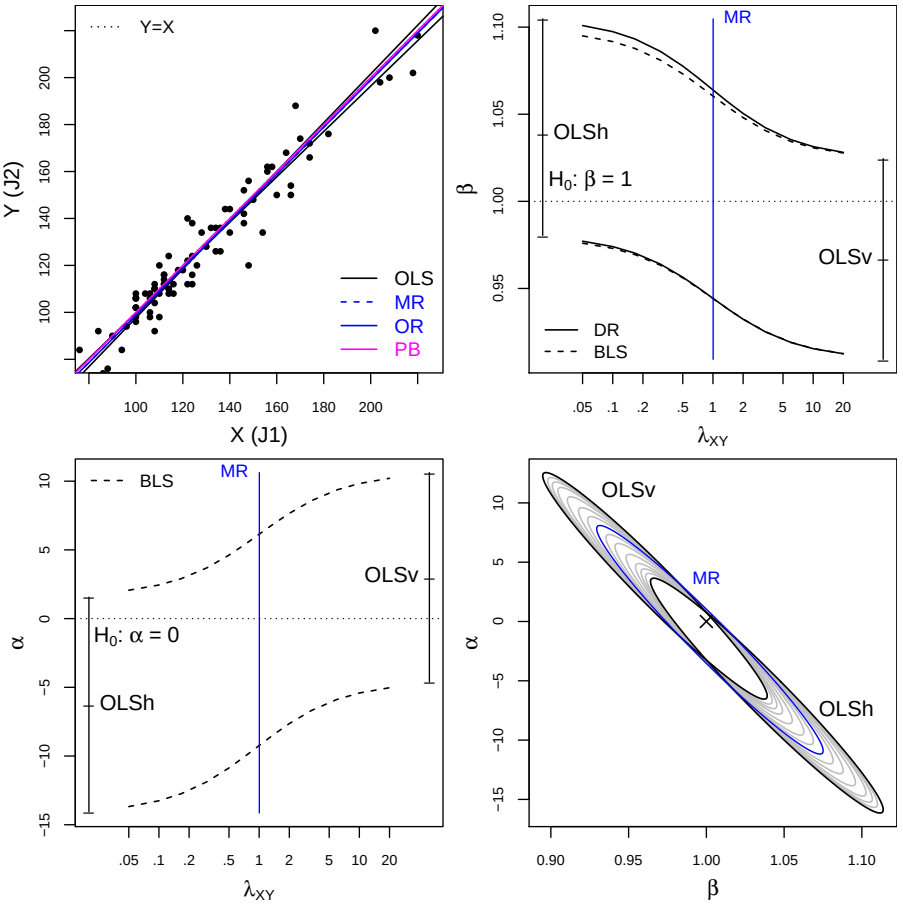

Figure 2.16: *SBP data ($n_X = n_Y = 1$), the different regressions lines (Top-left) and their CI for the parameters (Top-right for β, Bottom-left for α and Bottom-right for the joint-CI)*

	$\hat{\beta}$	$\hat{\alpha}$
OLSv	0.967	3.004
OLSh	1.039	-6.248
MR	1.002	-1.539
OR	1.002	-1.550
PB	1	0

Table 2.3: *SBP data ($n_X = n_Y = 1$), values of $\hat{\beta}$ and $\hat{\alpha}$ for the seven regressions*

$(\alpha = 0, \beta = 1)$ is inside all the ellipses from OLSv to OLSh, which means that the joint hypothesis (1.13) is also not rejected whatever λ_{XY}. In other words, we do not reject the hypothesis that the first measures are equivalent to the second (according to the null hypotheses given in Section 1.3.3).

Chapter 3

Focus on the joint confidence interval for errors-in-variables regression lines parameters

Contents

This chapter studies in detail the joint confidence interval of the parameters of errors-in-variables regression lines under homoscedasticity (with four methodologies) or heteroscedasticity (with two methodologies). The size and the shape of the confidence ellipses are compared and the corresponding hyperbolic confidence bands are provided. This comparison between confidence ellipses and confidence bands under homoscedasticity or heteroscedasticity has been published by Francq and Govaerts [10]. Additionally, a tip is given to improve the coverage probabilities of the joint confidence interval under heteroscedasticity with unknown variances by modeling the locally estimated variances. The robustness of these confidence ellipses or confidence bands are then assessed with the presence of outliers under homoscedasticity or heteroscedasticity.

3.1 Hyperbolic confidence bands of errors- in-variables regression lines

This section studies further the joint-CI presented in Section (2.4.3) where approximate confidence ellipses for regression parameters can be computed, for given values of λ_{XY}, for DR or BLS. We propose here two additional procedures to test the equivalence hypothesis (1.13) with a joint-CI. The four methodologies are then compared in detail with coverage probabilities and their properties and advantages discussed. Moreover, we propose to transform these confidence ellipses into hyperbolic confidence bands (CB) for the regression line $Y = \alpha + \beta X = x'\theta$ over $X \in (-\infty, \infty)$ where $x = (1, X)'$ which are equivalent to the confidence ellipses but easier to display and to interpret. The hypothesis (1.13) $H_0 : \theta = (0, 1)'$ is rejected if the line $Y = X$ intercepts the CB and not rejected if the line lies inside the CB.
The joint-CI and the confidence bands of the very well-known OLS regression will be first reviewed, explained and compared. Then, four different approaches to compute a joint-CI for errors-in-variables regression parameters will be presented. Afterwards, these ellipses (in the (β, α) space) will be transformed into confidence bands (in the (X, Y) space).
Non-parametric and bootstrap confidence bands for errors-in-variables regressions are available in the literature [73] but according to our knowledge, the (parametric) confidence bands given in this section are novel.
Confidence bands of other shapes [74,75] like two straight lines around the estimated line will not be considered, as the hyperbolic curve is the only correct way to fit the distribution of estimated lines.

3.1.1 The confidence bands for the OLSv line

This section reviews results on the classical estimation of a regression line by OLSv. It will be crucial in the development in further sections on errors-in-variables models. In particular, the not well known concept of confidence bands (CB) around a regression line is reviewed. It is shown that it is equivalent to the joint confidence interval for the regression parameters and that they can both be used for equivalence testing.

Recall that the equivalence between two measurements methods (under OLSv assumption) is rejected if:

$$\begin{pmatrix}\hat{\alpha}_{OLSv}-0 & \hat{\beta}_{OLSv}-1\end{pmatrix}\begin{pmatrix}S^2_{\hat{\alpha}_{OLSv}} & S_{\hat{\alpha}\hat{\beta}_{OLSv}}\\ S_{\hat{\alpha}\hat{\beta}_{OLSv}} & S^2_{\hat{\beta}_{OLSv}}\end{pmatrix}^{-1}\begin{pmatrix}\hat{\alpha}_{OLSv}-0\\ \hat{\beta}_{OLSv}-1\end{pmatrix} > 2F_{1-\gamma,2,N-2}$$

Note that when $|S_{\hat{\alpha}\hat{\beta}_{OLSv}}|$ increases, the ellipse becomes narrower and collapses to a line, otherwise when $|S_{\hat{\alpha}\hat{\beta}_{OLSv}}| \to 0$ the ellipse becomes wider until its major and minor axes become parallel to the β-axis and α-axis.

The confidence bands vs confidence intervals

The confidence bands for the OLSv line is a concept not very well-known by non-statisticians despite many papers already published. Indeed, the confidence band for a line is often not available in classical software. The confidence bands is the confidence interval for the regression line and should not be confused with the CI for the conditional mean $E(\hat{Y}_0|X_0)$ given by the following well-known formula [44]:

$$\hat{Y}_0 \pm t_{1-\frac{\gamma}{2},N-2}S_{\hat{Y}_0} \tag{3.1}$$

$$\text{where } \hat{Y}_0 = \hat{\alpha}_{OLSv} + \hat{\beta}_{OLSv}X_0$$

$$\text{and } S^2_{\hat{Y}_0} = S^2_{\hat{\alpha}_{OLSv}} + X_0^2 S^2_{\hat{\beta}_{OLSv}} + 2X_0 S_{\hat{\alpha}\hat{\beta}_{OLSv}} = S^2_{OLSv}\left(\frac{1}{N} + \frac{(X_0-\overline{X})^2}{S_{xx}}\right)$$

This formula is computed 'point by point' but can, of course, be displayed on a graph for all possible values of X_0 and has a hyperbolic shape. The CI for the line $Y = \alpha + \beta X$ over $X \in (-\infty, \infty)$ (and not for a given X_0) relies on the bivariate dimension of a line as we have to take into account jointly the uncertainties of both parameters (α and β). The confidence bands (CB) of the line $Y = \alpha + \beta X$ under OLS assumptions is given by the following formula [22]:

$$\left(\hat{\alpha}_{OLSv} + \hat{\beta}_{OLSv}X\right) \pm \sqrt{2F_{1-\gamma,2,N-2}}\sqrt{S^2_{OLSv}\left(\frac{1}{N} + \frac{(X-\overline{X})^2}{S_{xx}}\right)} \tag{3.2}$$

Both formulas (3.1 and 3.2) are similar but a F-distribution is used for the confidence bands based on the work of Working and Hotelling [76].

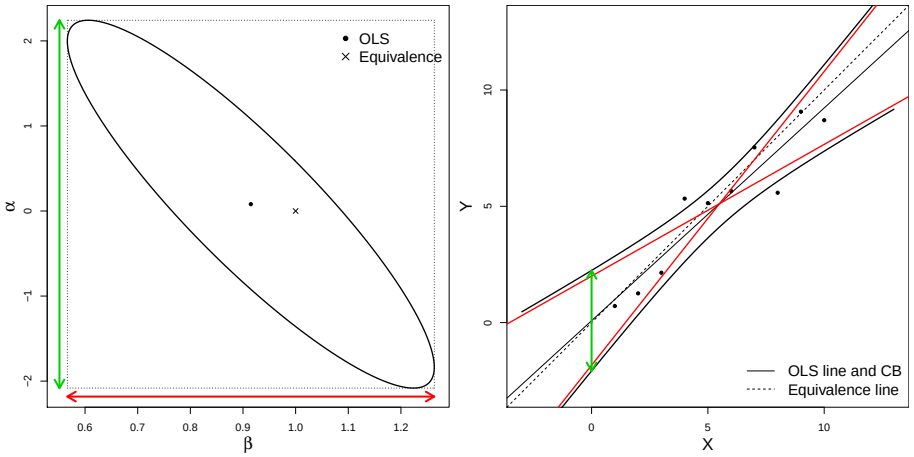

Figure 3.1: *Mathematical equivalence between the joint-CI on* (β, α) *(ellipse on the left) and the CB (hyperbolic band on the right) computed around the line*

Comparison of the joint-CI and the confidence bands

Actually, the joint-CI on the regression coefficients and the confidence bands of the regression line are mathematically equivalent. In other words, to test the equivalence between two measurement methods, we can check whether the point $(\beta = 1, \alpha = 0)$ lies inside the ellipse or check whether the identity line $(Y = X)$ lies inside the confidence bands. Figure 3.1 displays a simulated data set $(N = 10, X_{i=1,\dots,10}, Y_i = X_i + \epsilon_i$ and $\epsilon_i \sim N(0, 1))$ on the right with the OLSv line and the 95% CB. We can notice that the identity line (the dashed line) lies inside the CB which means that the equivalence is not rejected. On the left, the 95% joint-CI on (β, α) is displayed as an ellipse and the 'equivalence point' $(\beta = 1, \alpha = 0)$ lies inside the ellipse.
More generally, if a given point (β, α) lies inside the ellipse, the corresponding line $(Y = \alpha + \beta X)$ lies inside the CB and vice-versa. On the other hand, if a given point (β, α) lies outside the ellipse, the corresponding line intercepts the CB. For a given point (β, α) on the edge of the ellipse, the corresponding line is tangent to the CB.
The equivalence between the joint-CI (ellipse) and the confidence band can be justified as follows:
If we consider all the lines which lie inside the CB, it is easy to understand that the two extreme slopes correspond to the oblique asymptotes of the hyperbolic CB (red lines in Figure 3.1-right). These two extreme slopes correspond to the domain of the ellipse (red arrow on Figure 3.1-left) and it is easy to check that these slopes are:

$$\hat{\beta}_{OLSv} \pm \sqrt{2F_{1-\gamma,2,N-2}} S_{\hat{\beta}_{OLSv}} \tag{3.3}$$

From the CB, the slopes of the oblique asymptotes can be easily computed:

$$\lim_{X\to\infty} \frac{(\hat{\alpha}_{OLSv} + \hat{\beta}_{OLSv}X) \pm \sqrt{2F_{1-\gamma,2,N-2}}\sqrt{S^2_{OLSv}\left(\frac{1}{N} + \frac{(X-\overline{X})^2}{S_{xx}}\right)}}{X}$$

$$= \lim_{X\to\infty} \hat{\beta}_{OLSv} \pm \sqrt{2F_{1-\gamma,2,N-2}}\sqrt{\frac{S^2_{OLSv}\left(\frac{1}{N} + \frac{(X-\overline{X})^2}{S_{xx}}\right)}{X^2}}$$

$$= \hat{\beta}_{OLSv} \pm \sqrt{2F_{1-\gamma,2,N-2}}S_{\hat{\beta}_{OLSv}}$$

By analogy, it is easy to check that the image of the ellipse gives the two extreme intercepts (green arrow in Figure 3.1-left):

$$\hat{\alpha}_{OLSv} \pm \sqrt{2F_{1-\gamma,2,N-2}}S_{\hat{\alpha}_{OLSv}}, \tag{3.4}$$

and correspond to the two intercepts of the CB (when $X = 0$). Finally, for any slope β_0 in the domain of the ellipse (3.3), there exists two lines tangent to the hyperbolic CB with slope β_0 and intercepts α_0^1 and α_0^2 which correspond on the ellipse, to the two points α_0^1 and α_0^2 where the vertical line $\beta = \beta_0$ intercepts the ellipse (details not given).

3.1.2 Joint-CI and CB of errors-in-variables regressions under homoscedasticity

This section presents and compares the joint-CI ellipses and the corresponding confidence bands for the general model (1.11) $Y_i = \alpha + \beta X_i + \epsilon_i$ for four different methodologies. Their approximate covariance matrix $\hat{\Sigma}$ of estimators $\hat{\theta} = (\hat{\alpha}, \hat{\beta})'$ of θ are, then, compared.

We consider four methodologies available in the literature:

- DR already presented in sections 2.1 and 2.4
- BLS already presented in sections 2.1 and 2.4
- the Mandel's procedure presented in sections 2.4.2 and 2.4.3 under OR assumption ($\lambda_{XY} = 1$) but considered in this section in a more general context for any λ_{XY}
- the Galea-Rojas et al. procedure (GR) developed to tackle the heteroscedasticity and presented in this Section under homoscedasticity and heteroscedasticity

The formulas are written with the notations presented in Section 1. Under homoscedasticity, these four methodologies provide identical estimations of the regression line but the associated variance-covariance matrix of the parameters are different and lead to different CI.

Remind that the confidence ellipse for $\theta = (\alpha, \beta)'$ can be computed with the following formula:

$$(\hat{\theta} - \theta)'\hat{\Sigma}^{-1}(\hat{\theta} - \theta) < c \tag{3.5}$$

where the critical constant c is chosen suitably depending on whether $\chi^2_{1-\gamma,2}$ or $2F_{1-\gamma,2,N-2}$ is used as the approximate distribution. This confidence ellipse can be represented equivalently as confidence bands:

$$x'\theta \in x'\hat{\theta} \pm \sqrt{c}\sqrt{x'\hat{\Sigma}^{-1}x} \tag{3.6}$$

Estimator of θ, covariance matrix $\hat{\Sigma}$ and CB by DR

The DR estimator of θ were provided in Section 2.1, and the covariance matrix $\hat{\Sigma}$ and the joint-CI in Section 2.4 (we consider here the estimators of the variances and covariance given by the method of moments).
The hyperbolic CB for Deming regression line (DR) is given by the following formula:

$$(\hat{\alpha}_{DR} + \hat{\beta}_{DR}X) \pm \sqrt{2F_{1-\gamma,2,N-2}}\sqrt{S^2_{\hat{\beta}_{DR}}(X - \overline{X})^2 + \frac{\hat{\beta}^2_{DR}\sigma^2_\tau/n_X + \sigma^2_\nu/n_Y}{N}} \tag{3.7}$$

where it can be easily noticed that the slopes of the oblique asymptotes are $\hat{\beta}_{DR} \pm \sqrt{2F_{1-\gamma,2,N-2}}S_{\hat{\beta}_{DR}}$ and the intercepts are $\hat{\alpha}_{DR} \pm \sqrt{2F_{1-\gamma,2,N-2}}S_{\hat{\alpha}_{DR}}$.

Estimator of θ, covariance matrix $\hat{\Sigma}$ and CB by Galea-Rojas et al.

Galea-Rojas et al. [17] (GR) propose a regression model based on a paper previously published by Ripley and Thompson [16] where maximum likelihood is applied to take into account the errors and heteroscedasticity in both axes. GR and DR regressions are both maximum likelihood procedures, but, respectively, derived under heteroscedasticity or homoscedasticity. Then, it can be verified that $\hat{\beta}_{GR} = \hat{\beta}_{DR}$ and $\hat{\alpha}_{GR} = \hat{\alpha}_{DR}$ (as the GR procedure is more general than the DR). We present in this section the GR formulas in the case of homoscedasticity (and with replicated data). The estimators of the parameters are:

$$\hat{\beta}_{GR} = \frac{\sum_{i=1}^N W_{GR}\hat{x}_i(Y_i - \overline{Y})}{\sum_{i=1}^N W_{GR}\hat{x}_i(X_i - \overline{X})} = \frac{\sum_{i=1}^N \hat{x}_i(Y_i - \overline{Y})}{\sum_{i=1}^N \hat{x}_i(X_i - \overline{X})} \text{ and } \hat{\alpha}_{GR} = \overline{Y} - \hat{\beta}_{GR}\overline{X} \tag{3.8}$$

where

$$W_{GR} = \frac{1}{\frac{\sigma^2_\nu}{n_Y} + \hat{\beta}^2_{GR}\frac{\sigma^2_\tau}{n_X}}$$

and

$$\hat{x}_i = \frac{\frac{\sigma_\nu^2}{n_Y} X_i + \hat{\beta}_{GR} \frac{\sigma_\tau^2}{n_X} (Y_i - \hat{\alpha}_{GR})}{\frac{\sigma_\nu^2}{n_Y} + \hat{\beta}_{GR}^2 \frac{\sigma_\tau^2}{n_X}}$$

$\hat{x}_i$ is the abscissa of the projection of the i^{th} point to the line in the oblique direction previously defined in Section 2.1.5. In practice, if σ_τ^2 and σ_ν^2 are unknown, they can be estimated with replicated data and replaced by S_τ^2 and S_ν^2 respectively.

Galea-Rojas et al. [17] provide the asymptotic variance-covariance matrix of the parameters derived by ML. In the case of homoscedasticity and with replicated data, this asymptotic variance-covariance matrix of the parameters is given by:

$$\Sigma_{GR} = W_N^{-1} V_N W_N^{-1} / N, \tag{3.9}$$

where

$$W_N = \frac{W_{GR}}{N} \begin{pmatrix} N & \sum_{i=1}^N \xi_i \\ \sum_{i=1}^N \xi_i & \sum_{i=1}^N \xi_i^2 \end{pmatrix} \text{ and } V_N = W_N + \begin{pmatrix} 0 & 0 \\ 0 & k_{GR} \end{pmatrix}$$

with

$$k_{GR} = \frac{W_{GR}}{C} \text{ and } C = \frac{1}{\sigma_\tau^2 / n_X} + \frac{\beta^2}{\sigma_\nu^2 / n_Y}$$

In practice, to get a consistent estimator of the variance-covariance matrix, β^2 can be replaced by $\hat{\beta}_{GR}^2$, ξ_i by $\hat{x}_i$ and ξ_i^2 by $\hat{x}_i^2 - 1/\hat{C}$ [17]. $\hat{C}$ is not clearly defined by Galea-Rojas et al. [17]. In this thesis, $\hat{C} = 1/(\sigma_\tau^2/n_X) + \hat{\beta}_{GR}^2/(\sigma_\nu^2/n_Y)$ where σ_τ^2 and σ_ν^2 can be replaced, respectively, by S_τ^2 and S_ν^2 if needed. From the Wald statistic given in the literature [17], the critical threshold is $c = \chi_{1-\gamma,2}^2$. The variances of the parameters can also be computed with the following equivalent formulas [17]:

$$\sigma_{\hat{\beta}_{GR}}^2 = \frac{1}{SS_W} \left(1 + \frac{N k_{GR}}{SS_W}\right), \sigma_{\hat{\alpha}_{GR}}^2 = \frac{1}{N W_{GR}} + \overline{\xi}^2 \sigma_{\hat{\beta}_{GR}}^2 \text{ and } \sigma_{\hat{\alpha}\hat{\beta}_{GR}} = -\overline{\xi} \sigma_{\hat{\beta}_{GR}}^2 \tag{3.10}$$

with $SS_W = W_{GR} \sum_{i=1}^N (\xi_i - \overline{\xi})^2$ and $\overline{\xi} = \frac{\sum_{i=1}^N \xi_i}{N}$.

In practice, $\sigma_{\hat{\beta}_{GR}}^2$, $\sigma_{\hat{\alpha}_{GR}}^2$ and $\sigma_{\hat{\alpha}\hat{\beta}_{GR}}$ can be estimated by replacing $\overline{\xi}$ by $\overline{X}$, and SS_W by $W_{GR} \sum_{i=1}^N (\hat{x}_i^2 - \hat{C}^{-1} - 2\hat{x}_i \overline{X} + \overline{X}^2)$.

The hyperbolic CB of the Galea-Rojas et al. procedure can be calculated by the following formula:

$$(\hat{\alpha}_{GR} + \hat{\beta}_{GR} X) \pm \sqrt{\chi_{1-\gamma,2}^2} \sqrt{S_{\hat{\beta}_{GR}}^2 (X - \overline{X})^2 + \frac{1}{N W_{GR}}} \tag{3.11}$$

where the slopes of the oblique asymptotes are $\hat{\beta}_{GR} \pm \sqrt{\chi_{1-\gamma,2}^2} S_{\hat{\beta}_{GR}}$ and the intercepts are $\hat{\alpha}_{GR} \pm \sqrt{\chi_{1-\gamma,2}^2} S_{\hat{\alpha}_{GR}}$.

Estimator of θ, covariance matrix $\hat{\Sigma}$ and CB by BLS

The BLS estimator of θ were provided in Section 2.1, and the covariance matrix $\hat{\Sigma}$ and the joint-CI in Section 2.4.
The corresponding hyperbolic CB is given by the following formula:

$$(\hat{\alpha}_{BLS} + \hat{\beta}_{BLS} X) \pm$$

$$\sqrt{2F_{1-\gamma,2,N-2}} \sqrt{\frac{W_{BLS}}{N\sum_{i=1}^{N} X_i^2 - (\sum_{i=1}^{N} X_i)^2} (\sum_{i=1}^{N} X_i^2 - 2X \sum_{i=1}^{N} X_i + NX^2) s_{BLS}^2} \tag{3.12}$$

where the slopes of the oblique asymptotes are $\hat{\beta}_{BLS} \pm \sqrt{2F_{1-\gamma,2,N-2}} S_{\hat{\beta}_{BLS}}$ and the intercepts are $\hat{\alpha}_{BLS} \pm \sqrt{2F_{1-\gamma,2,N-2}} S_{\hat{\alpha}_{BLS}}$.

Estimator of θ, covariance matrix $\hat{\Sigma}$ and CB by Mandel

This section considers the Mandel's procedure presented in Sections 2.4.2 and 2.4.3 under OR assumption ($\lambda_{XY} = 1$), generalized here for any values of λ_{XY}. Recall that Mandel's procedure consists in transforming the (X, Y) data into (U, V) data where $U_i = X_i + kY_i$ and $V_i = Y_i - \hat{\beta}_{Mandel} X_i$. The OLSv is then applied to the (U, V) data and transformed back into the (X, Y) axis to finally get a regression line which takes into account errors in both axis and their correlation. By using λ_{XY} instead of λ in the Mandel's procedure (with uncorrelated errors), the formulas are:

$$\hat{\beta}_{Mandel} = \frac{S_{xy} + kS_{yy}}{S_{xx} + kS_{xy}} \text{ and } \hat{\alpha}_{Mandel} = \overline{Y} - \hat{\beta}_{Mandel} \overline{X} \tag{3.13}$$

with

$$k = \frac{\hat{\beta}_{Mandel}}{\lambda_{XY}}$$

$\hat{\beta}_{Mandel}$ can be computed by iterations or by solving a 2^{nd} degree equation and it is easy to check that $\hat{\beta}_{Mandel} = \hat{\beta}_{DR} = \hat{\beta}_{GR} = \hat{\beta}_{BLS}$ and $\hat{\alpha}_{Mandel} = \hat{\alpha}_{DR} = \hat{\alpha}_{GR} = \hat{\alpha}_{BLS}$.
The variances of the parameters $\hat{\beta}_{Mandel}$ and $\hat{\alpha}_{Mandel}$ are derived by the general formula for the propagation of errors from the reconversion to (X, Y) scales:

$$S_{\hat{\beta}_{Mandel}}^2 = \frac{(1 + k\hat{\beta}_{Mandel})^2}{S_{uu}} S_{e_{Mandel}}^2 \tag{3.14}$$

and

$$S_{\hat{\alpha}_{Mandel}}^2 = \left(\frac{1}{N} + \frac{\overline{X}^2 (1 + k\hat{\beta}_{Mandel})^2}{S_{uu}} \right) S_{e_{Mandel}}^2 \tag{3.15}$$

with

$$S^2_{e_{Mandel}} = \frac{S_{vv}}{N-2}$$

The covariance and the joint-CI are not provided by Mandel but the covariance can be also easily computed by the formula for the propagation of errors:

$$S_{\hat{\alpha}\hat{\beta}_{Mandel}} = -\overline{X} S^2_{\hat{\beta}_{Mandel}} \tag{3.16}$$

We propose to use a F-distribution with $c = 2F_{1-\gamma,2,N-2}$ as the Mandel's procedure is based on the OLS technique.
The hyperbolic CB of the Mandel's procedure is given by the following formula:

$$(\hat{\alpha}_{Mandel} + \hat{\beta}_{Mandel} X) \pm$$

$$\sqrt{2F_{1-\gamma,2,N-2}} \sqrt{S^2_{e_{Mandel}} \left(\frac{(1 + k\hat{\beta}_{Mandel})^2}{S_{uu}} (X - \overline{X})^2 + \frac{1}{N} \right)} \tag{3.17}$$

The CB could be also computed with the OLS technique in the (U, V) axes and transform back in (X, Y) axes but this approach is not considered in this thesis.

3.1.3 Comparison of the coverage probabilities of the joint CI or CB under homoscedasticity

Coverage probabilities of the joint CI or CB under equivalence, H_0

In order to compare the coverage probabilities of the joint-CI or CB provided by the four methodologies presented in the previous sections, we simulated 100 000 samples as described in Section 2.5.1:

- under equivalence ($\alpha = 0, \beta = 1, \eta_i = \xi_i$)
- for the values of λ_{XY} given in Table 2.1
- with unreplicated data ($n_X = n_Y = 1, \lambda = \lambda_{XY}$ known)
- with replicated data to allow the estimation of σ^2_ν and σ^2_τ:
  - ↬ with equal number of replicates ($n_X = n_Y = 2$ and $\lambda = \lambda_{XY}$ or $n_X = n_Y = 4$ and $\lambda = \lambda_{XY}$)
  - ↬ with unequal number of replicates ($n_X = 4, n_Y = 2$ and $\lambda_{XY} = 2\lambda$) (in such a way that the values of λ_{XY} are identical to those of the unreplicated case)

The joint-CI or CB were computed for each simulated sample and the coverage probabilities (at a nominal level = 95%) computed per value of λ_{XY}.
To study in more details the effect of the sample size, we also run simulations:

- with $\lambda_{XY} = 0.33, 1, 3$,

- with N from 10 to 100 (10, 12, 14, 16, 20, 30, 50, 75, 100)
- with or without replicated data as shown in Table 3.1

The coverage probabilities are displayed for λ_{XY} known in Figure 3.2 with respect to λ_{XY} (left, on a logarithmic scale) and to N (right). Figures 3.3, 3.4, 3.5 display the coverage probabilities corresponding to replicated data and estimated λ_{XY}.
When the variances are known, all the coverage probabilities are between 93% and 96% and closer to 95% when $\lambda_{XY} > 1$ for the GR, BLS and Mandel procedures. As already mentioned, the coordinate system should, indeed, be chosen such that $\lambda_{XY} > 1$. When N increases, the GR and DR are closer to 95% while the BLS and Mandel procedures provide slightly lower coverage probabilities.
When the variances are estimated with replicated data, the coverage probabilities provided by the DR are slightly lower but are the closest to 95%. On the other hand, the coverage probabilities provided by the GR's procedure drop because of the uncertainties on the estimated variances but these coverage probabilities increase with N. For instance, when $n_X = n_Y = 2$ with $\lambda_{XY} = 1$, the GR outperforms the BLS with $N > 30$ (approximately). The BLS and Mandel methodologies provide similar coverage probabilities with known or unknown variances. In practice, the accuracy of one measurement method is rarely three times higher than the other such that λ_{XY} often lies between 0.33 and 3. In this interval, the DR and GR are always slightly more suitable with known variances while the DR is always more suitable with unknown variances.

Coverage probabilities of the joint CI or CB under non-equivalence, H_1

In order to assess the power of the hypothesis test based on the confidence region for θ provided by the four methodologies presented in the previous sections, we simulated 10000 samples under non-equivalence with unreplicated or replicated data with $n_X = n_Y = 2$, as described in Section 2.5.1. However, λ_{XY} was set to 1 (see $\lambda_{XY} = 1$ in Table 2.1) by default.
It is well known that the higher is the correlation between the parameters estimates and the more powerful the joint confidence interval is, in comparison to the separated confidence intervals. Indeed, the powers reach fastly 100% with the values of α and β described in Section 2.5.2. The non-equivalence is, then, simulated closer to H_0:

- with β ranged from 0.9 to 1.1 (by 0.025)
- with the corresponding values of α:
  - ranged from -1 to 1 (by 0.25) in order to move away H_0 in the direction of the **minor axis** of the confidence ellipse under equivalence,
  - ranged from 1 to -1 (by 0.25) in order to move away H_0 in the direction of the **major axis** of the confidence ellipse under equivalence.

Figure 3.6 displays the powers of the hypothesis test $H_0 : \theta = (0, 1)'$ for $N = 10, 20, 50$ and $n_X = n_Y = 1$ with respect to β and the corresponding values of α. The powers obtained by replicated data, $n_X = n_Y = 2$, are displayed in Figure 3.7.
We can notice in Figure 3.6 that the powers are very close to each other between the four methodologies. Obviously, the powers increase when we move away from the equivalence, H_0, and when the sample size, N, increases. It can also be noticed that the powers are higher when we move away H_0 in the direction of the minor axis of the confidence ellipse (right) in comparison to the major axis (left). Indeed, $\hat{\beta}$ and $\hat{\alpha}$ are negatively correlated. With replicated data, these findings are similar (Figure 3.7).

	$n_X = n_Y = 1$			$n_X = n_Y = 2$			$n_X = 4, n_Y = 2$			$n_X = n_Y = 4$		
σ_ν^2	0.25	0.75	0.75	0.5	1.5	1.5	0.5	1.5	1.5	1	3	3
σ_τ^2	0.75	0.75	0.25	1.5	1.5	0.5	3	3	1	3	3	1
λ	0.33	1	3	0.33	1	3	0.17	0.5	1.5	0.33	1	3
λ_{XY}	0.33	1	3	0.33	1	3	0.33	1	3	0.33	1	3

Table 3.1: *Values of σ_ν^2 and σ_τ^2 for the simulations with or without replicated data and the corresponding values of λ and λ_{XY}*

3.1.4 Joint-CI and CB of errors-in-variables regressions under heteroscedasticity

This section presents the formulas of the regression estimators of model (1.10) when the (heteroscedastic) variances $\sigma_{\tau_i}^2$ and $\sigma_{\nu_i}^2$ are known or estimated with replicates.
Different types of heteroscedasticity profiles can be encountered in method comparison studies as shown in Figure 3.8. On top-left, measurement errors are equal on both axes but increase with respect to ξ or η (this is common in chemistry for instance). On top-right, measurement errors increase on the X-axis and are constant on the Y-axis while it is the opposite on bottom-left. On bottom-right, the measurement errors increase linearly on the X-axis while the errors increase on the extreme of the Y-axis (when η moves closer to 10 or 20) and is minimal on the middle range of the Y-axis.
The Mandel's regression and the DR are not considered anymore as they are not suitable in the case of heteroscedasticity. Note that a weighted Deming regression is available in the literature [77] but it will not be considered anymore in this thesis (as the BLS is more suitable). The GR procedure and the BLS can take into account the heteroscedasticity and are identical for estimating the regression line but the variance-covariance matrix of the parameters are

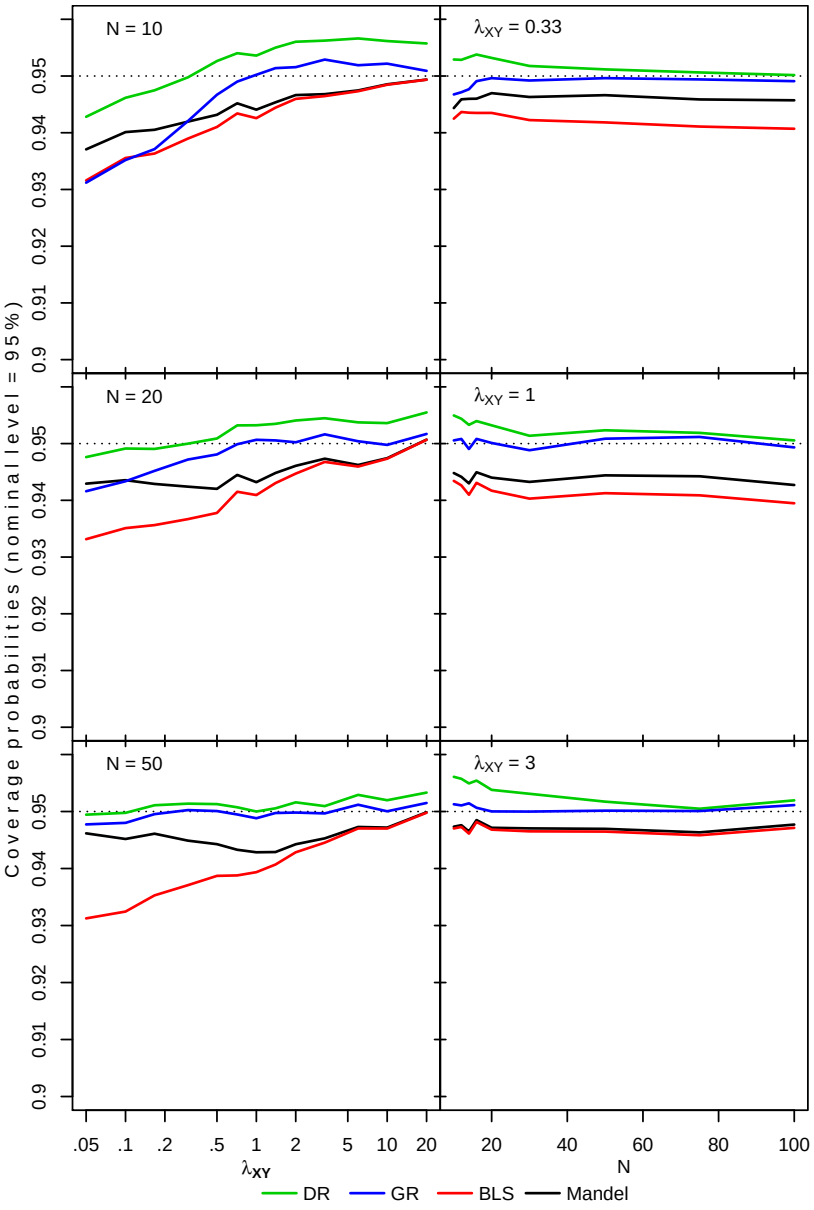

Figure 3.2: *Coverage probabilities of the joint-CI or CB related to* λ_{XY} *in a logarithmic scale with* $N = 10$*, 20, 50 (left) and related to* N *for* $\lambda_{XY} = 0.33$*, 1, 3 (right),* $n_X = n_Y = 1$

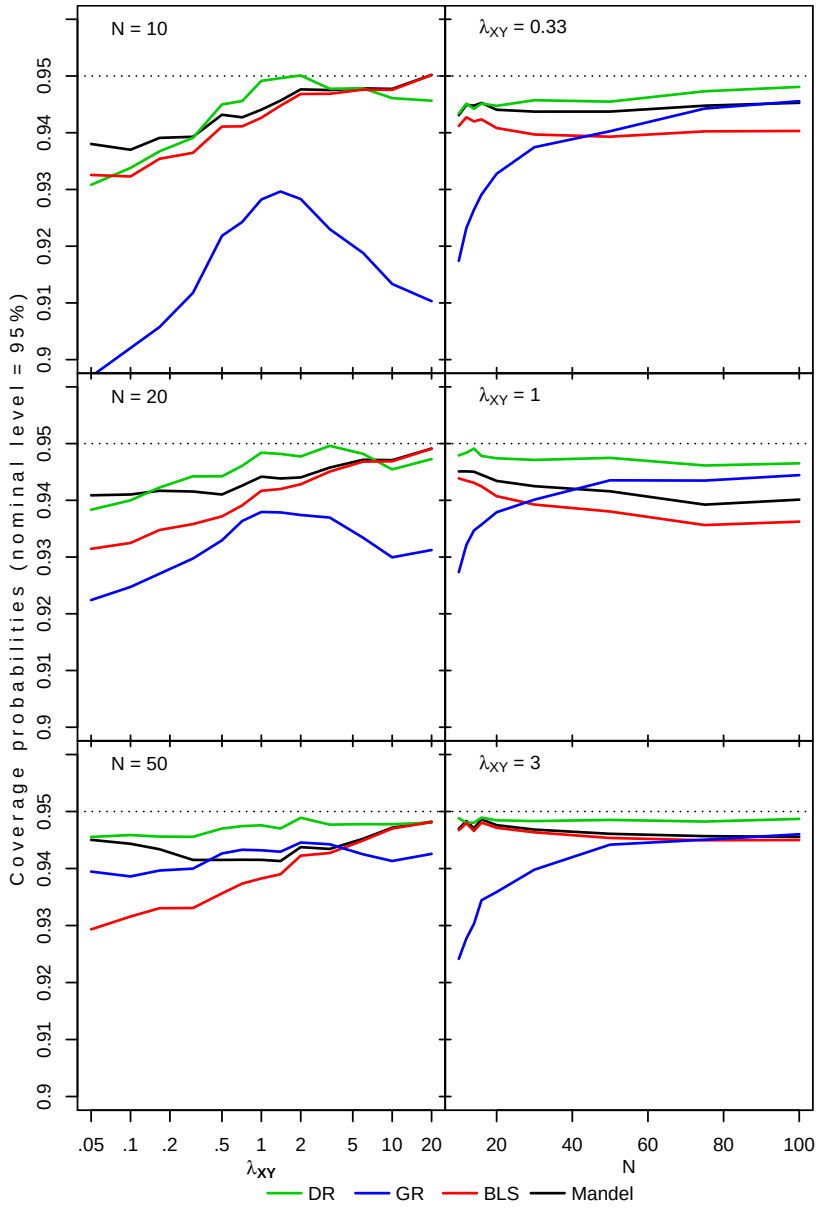

Figure 3.3: *Coverage probabilities of the joint-CI or CB related to* λ_{XY} *in a logarithmic scale with* $N = 10$*, 20, 50 (left) and related to* N *for* $\lambda_{XY} = 0.33$*, 1, 3 (right),* $n_X = n_Y = 2$

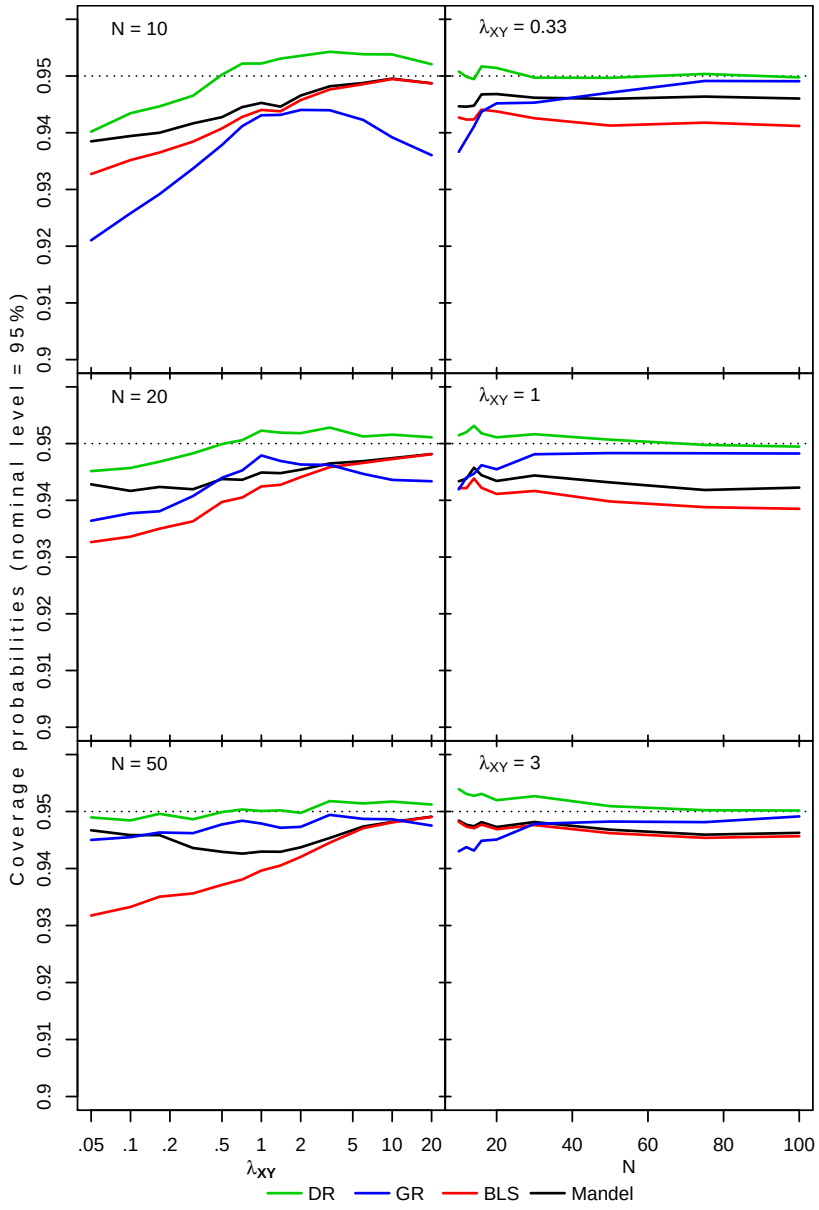

Figure 3.4: *Coverage probabilities of the joint-CI or CB related to* λ_{XY} *in a logarithmic scale with* $N = 10$*, 20, 50 (left) and related to* N *for* $\lambda_{XY} = 0.33$*, 1, 3 (right),* $n_X = n_Y = 4$

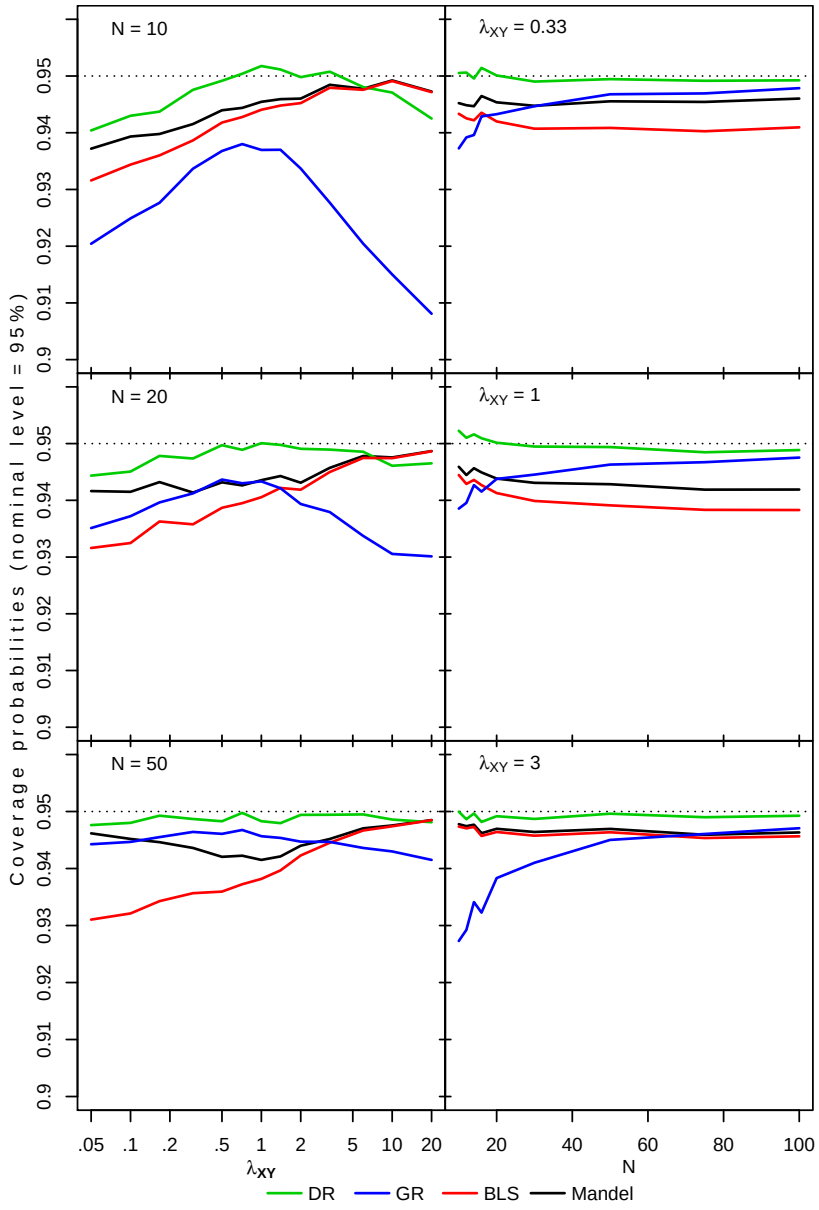

Figure 3.5: *Coverage probabilities of the joint-CI or CB related to* λ_{XY} *in a logarithmic scale with* $N = 10$*, 20, 50 (left) and related to* N *for* $\lambda_{XY} = 0.33$*, 1, 3 (right),* $n_X = 4, n_Y = 2$

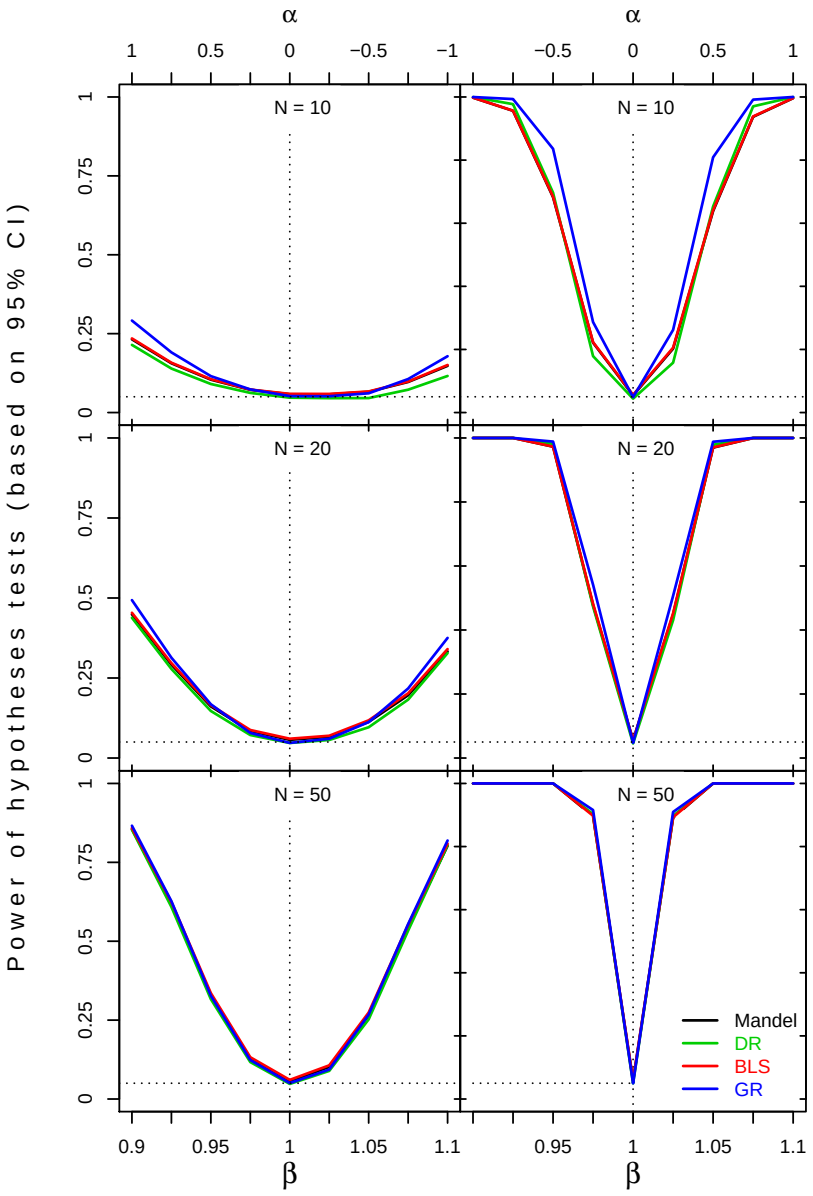

Figure 3.6: *Power of the hypothesis test (based on a joint-CI) for θ related to β from 0.9 to 1.1 and α from 1 to -1 (left) or -1 to 1 (right), for N=10 (top), N=20 (middle) and N=50 (bottom), $n_X = n_Y = 1$*

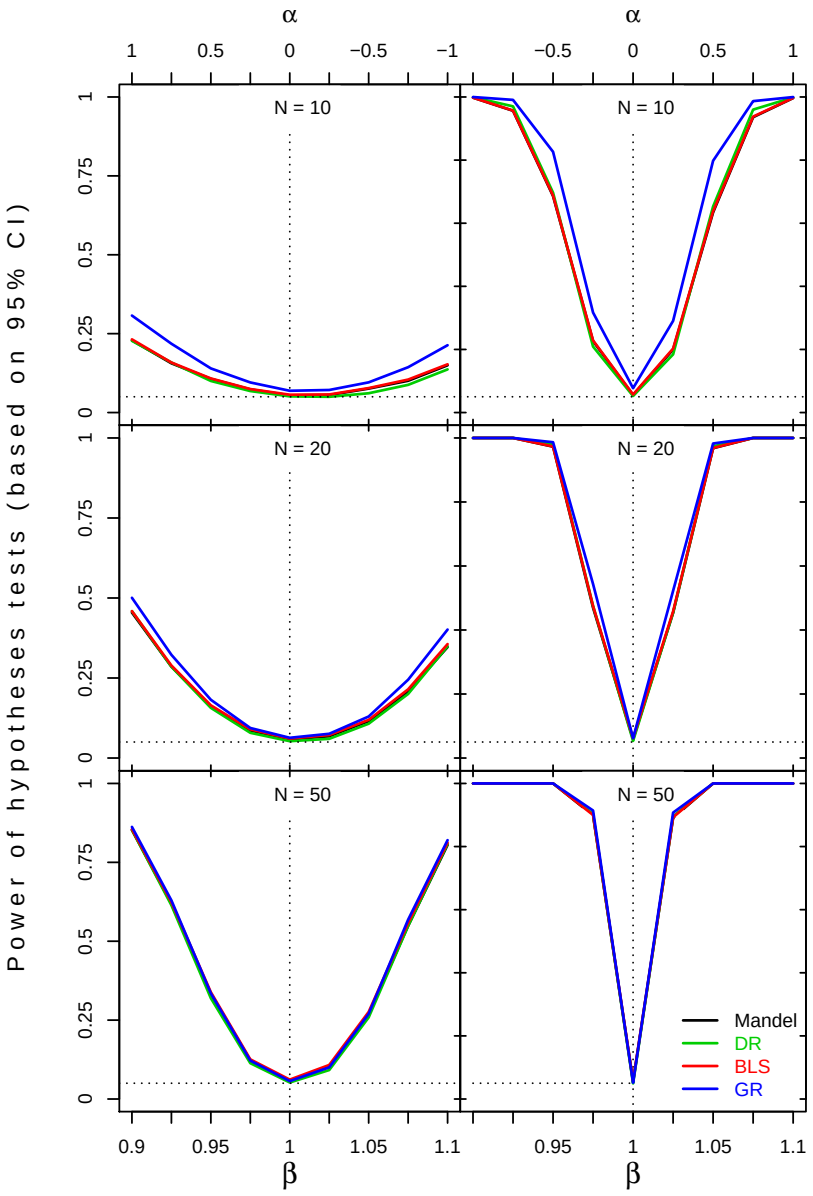

Figure 3.7: *Power of the hypothesis test (based on a joint-CI) for θ related to β from 0.9 to 1.1 and α from 1 to -1 (left) or -1 to 1 (right), for N=10 (top), N=20 (middle) and N=50 (bottom), $n_X = n_Y = 2$*

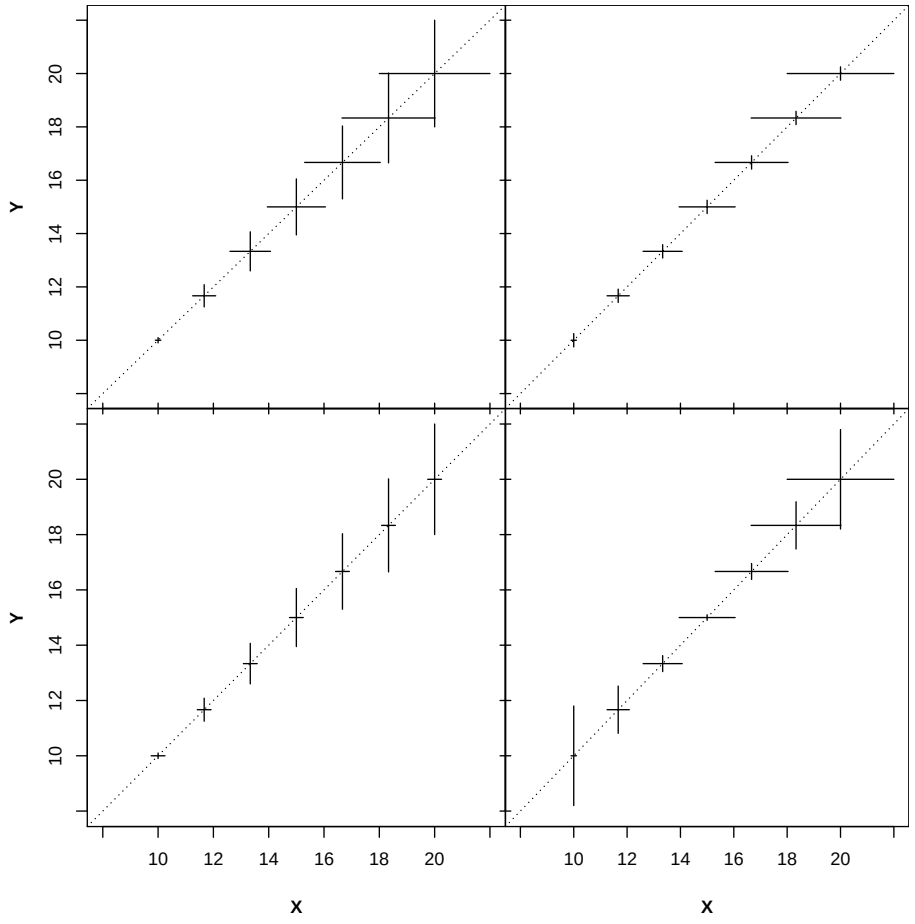

Figure 3.8: *Different profiles of heteroscedasticity encountered in method comparison studies*

computed differently. The covariance matrix $\hat{\Sigma}$ of the estimates $\hat{\theta} = (\hat{\alpha}, \hat{\beta})'$ given in this section can be plugged in $(\hat{\theta}-\theta)'\hat{\Sigma}^{-1}(\hat{\theta}-\theta) < c$ (formula (3.5)) or in $x'\theta \in x'\hat{\theta} \pm \sqrt{c}\sqrt{x'\hat{\Sigma}^{-1}x}$ (formula (3.6)) to get respectively the corresponding confidence ellipse or confidence bands (with the same critical constant c given in Section 3.1.2).

The Galea-Rojas et al. procedure

Based on a paper published previously by Ripley and Thompson [16], the ML parameters estimators are:

$$\hat{\beta}_{GR} = \frac{\sum_{i=1}^{N} W_{i_{GR}} \hat{x}_i (Y_i - \overline{Y})}{\sum_{i=1}^{N} W_{i_{GR}} \hat{x}_i (X_i - \overline{X})} \text{ and } \hat{\alpha}_{GR} = \overline{Y}_W - \hat{\beta}_{GR}\overline{X}_W, \tag{3.18}$$

where

$$W_{i_{GR}} = \frac{1}{\frac{\sigma^2_{\nu_i}}{n_{Y_i}} + \hat{\beta}^2_{GR}\frac{\sigma^2_{\tau_i}}{n_{X_i}}} \text{ and } \hat{x}_i = \frac{\frac{\sigma^2_{\nu_i}}{n_{Y_i}} X_i + \hat{\beta}_{GR}\frac{\sigma^2_{\tau_i}}{n_{X_i}}(Y_i - \hat{\alpha}_{GR})}{\frac{\sigma^2_{\nu_i}}{n_{Y_i}} + \hat{\beta}^2_{GR}\frac{\sigma^2_{\tau_i}}{n_{X_i}}},$$

$$\overline{X}_W = \frac{\sum_{i=1}^{N} W_{i_{GR}} X_i}{\sum_{i=1}^{N} W_{i_{GR}}} \text{ and } \overline{Y}_W = \frac{\sum_{i=1}^{N} W_{i_{GR}} Y_i}{\sum_{i=1}^{N} W_{i_{GR}}}$$

In practice, if $\sigma^2_{\tau_i}$ and $\sigma^2_{\nu_i}$ are unknown, they can be estimated with replicated data and replaced by $S^2_{\tau_i}$ and $S^2_{\nu_i}$. The asymptotic variance-covariance matrix of the parameters derived by ML is given by Galea-Rojas et al. [17]:

$$\Sigma_{GR} = W_N^{-1} V_N W_N^{-1}/N, \tag{3.19}$$

where

$$W_N = \frac{1}{N}\begin{pmatrix} \sum_{i=1}^{N} W_{i_{GR}} & \sum_{i=1}^{N} W_{i_{GR}}\xi_i \\ \sum_{i=1}^{N} W_{i_{GR}}\xi_i & \sum_{i=1}^{N} W_{i_{GR}}\xi_i^2 \end{pmatrix} \text{ and } V_N = W_N + \begin{pmatrix} 0 & 0 \\ 0 & k_{GR} \end{pmatrix}$$

with

$$k_{GR} = \frac{1}{N}\sum_{i=1}^{N} \frac{W_{i_{GR}}}{C_i} \text{ and } C_i = \frac{1}{\sigma^2_{\tau_i}/n_{X_i}} + \frac{\beta^2}{\sigma^2_{\nu_i}/n_{Y_i}}$$

In practice, to get a consistent estimator of the variance-covariance matrix, β^2 can be replaced by $\hat{\beta}^2_{GR}$, ξ_i by $\hat{x}_i$ and ξ_i^2 by ${\hat{x}_i}^2 - 1/\hat{C}_i$ [17]. $\hat{C}$ is not clearly defined by Galea-Rojas et al. [17]. In this thesis, $\hat{C} = 1/(\sigma^2_{\tau_i}/n_{X_i}) + \hat{\beta}^2_{GR}/(\sigma^2_{\nu_i}/n_{Y_i})$ where $\sigma^2_{\tau_i}$ and $\sigma^2_{\nu_i}$ can be replaced, respectively, by $S^2_{\tau_i}$ and $S^2_{\nu_i}$ if needed. The variances of the parameters can also be computed with the following equivalent formulas [17]:

$$\sigma^2_{\hat{\beta}_{GR}} = \frac{1}{SS_W}\left(1 + \frac{Nk_{GR}}{SS_W}\right), \sigma^2_{\hat{\alpha}_{GR}} = \frac{1}{\sum_{i=1}^{N} W_{i_{GR}}} + \overline{\xi}^2_W \sigma^2_{\hat{\beta}_{GR}} \tag{3.20}$$

and

$$\sigma_{\hat{\alpha}\hat{\beta}_{GR}} = -\overline{\xi}_W \sigma^2_{\hat{\beta}_{GR}} \tag{3.21}$$

with $SS_W = \sum_{i=1}^{N} W_{i_{GR}} (\xi_i - \overline{\xi}_W)^2$ and $\overline{\xi}_W = \frac{\sum_{i=1}^{N} W_{i_{GR}} \xi_i}{\sum_{i=1}^{N} W_{i_{GR}}}$.
In practice, $\sigma^2_{\hat{\beta}_{GR}}$, $\sigma^2_{\hat{\alpha}_{GR}}$ and $\sigma_{\hat{\alpha}\hat{\beta}_{GR}}$ can be estimated by replacing $\overline{\xi}_W$ by $\overline{X}_W$, and SS_W by $\sum_{i=1}^{N} W_{i_{GR}} (\hat{x}_i^2 - \hat{C}_i^{-1} - 2\hat{x}_i \overline{X}_W + \overline{X}^2_W)$.
The hyperbolic CB is given in practice by the following formula:

$$\alpha_{GR} + \beta_{GR} X \in$$

$$(\hat{\alpha}_{GR} + \hat{\beta}_{GR} X) \pm \sqrt{\chi^2_{1-\gamma,2}} \sqrt{\frac{1}{SS_W} \left(1 + \frac{N k_{GR}}{SS_W}\right) (X - \overline{X}_W)^2 + \frac{1}{\sum_{i=1}^{N} W_{i_{GR}}}} \tag{3.22}$$

The Bivariate Least Square regression: BLS

The BLS regression line minimizes the criterion C_{BLS} [58–60]:

$$C_{BLS} = \sum_{i=1}^{N} \frac{1}{W_{i_{BLS}}} (Y_i - \hat{\alpha} - \hat{\beta} X_i)^2 = (N-2) s^2_{BLS}$$

with

$$W_{i_{BLS}} = \sigma^2_{\epsilon_i} = \frac{\sigma^2_{\nu_i}}{n_{Y_i}} + \hat{\beta}^2 \frac{\sigma^2_{\tau_i}}{n_{X_i}}$$

The parameters are estimated by iterations with the following formula:

$$b = R^{-1} g \tag{3.23}$$

where

$$R = \begin{pmatrix} \sum_{i=1}^{N} \frac{1}{W_{i_{BLS}}} & \sum_{i=1}^{N} \frac{X_i}{W_{i_{BLS}}} \\ \sum_{i=1}^{N} \frac{X_i}{W_{i_{BLS}}} & \sum_{i=1}^{N} \frac{X_i^2}{W_{i_{BLS}}} \end{pmatrix}$$

and

$$g = \begin{pmatrix} \sum_{i=1}^{N} \frac{Y_i}{W_{i_{BLS}}} \\ \sum_{i=1}^{N} \left(\frac{X_i Y_i}{W_{i_{BLS}}} + \hat{\beta}_{BLS} \frac{\sigma^2_{\tau_i}}{n_{X_i}} \frac{(Y_i - \hat{\alpha}_{BLS} - \hat{\beta}_{BLS} X_i)^2}{W^2_{i_{BLS}}} \right) \end{pmatrix}$$

In practice, if $\sigma^2_{\tau_i}$ and $\sigma^2_{\nu_i}$ are unknown, they can be estimated with replicated data and replaced by $S^2_{\tau_i}$ and $S^2_{\nu_i}$. It can be proven (after some algebraic manipulations) that $\hat{\beta}_{BLS} = \hat{\beta}_{GR}$ and $\hat{\alpha}_{BLS} = \hat{\alpha}_{GR}$ under heteroscedasticity. The variance-covariance matrix of the parameters provided by Riu and Rius [59] is given by:

$$\hat{\Sigma}_{BLS} = s^2_{BLS} R^{-1} \tag{3.24}$$

or can be computed with the following formulas:

$$S^2_{\hat{\beta}_{BLS}} = \frac{s^2_{BLS}\sum_{i=1}^{N}\frac{1}{W_{i_{BLS}}}}{D},\ S^2_{\hat{\alpha}_{BLS}} = \frac{s^2_{BLS}\sum_{i=1}^{N}\frac{X_i^2}{W_{i_{BLS}}}}{D} \tag{3.25}$$

and

$$S_{\hat{\alpha}\hat{\beta}_{BLS}} = \frac{s^2_{BLS}\sum_{i=1}^{N}\frac{X_i}{W_{i_{BLS}}}}{D} \tag{3.26}$$

with $D = \sum_{i=1}^{N}\frac{1}{W_{i_{BLS}}}\sum_{i=1}^{N}\frac{X_i^2}{W_{i_{BLS}}} - \left(\sum_{i=1}^{N}\frac{X_i}{W_{i_{BLS}}}\right)^2$.

The hyperbolic CB is, then, given by the following formula:

$$\alpha_{BLS} + \beta_{BLS}X \in (\hat{\alpha}_{BLS} + \hat{\beta}_{BLS}X)\pm$$
$$\sqrt{2F_{1-\gamma,2,N-2}}\sqrt{\frac{s^2_{BLS}}{D}\left(\sum_{i=1}^{N}\frac{X_i^2}{W_{i_{BLS}}} - 2X\sum_{i=1}^{N}\frac{X_i}{W_{i_{BLS}}} + X^2\sum_{i=1}^{N}\frac{1}{W_{i_{BLS}}}\right)} \tag{3.27}$$

3.1.5 Coverage probabilities of the joint CI or CB under heteroscedasticity

The coverage probabilities of the joint-CI provided by GR and BLS have already been compared in the literature but only in the case of known variances [17]. Since in practice, the variances are never known but always estimated, we propose to study in more details the coverage probabilities with known and, chiefly, unknown variances. We simulated 10000 samples:

- with $N = 10$, 12, 15, 20, 30, 50, 75, 100
- with replicated data ($n_X = n_Y = 5$)
- under equivalence ($\alpha = 0, \beta = 1, \eta_i = \xi_i$)
- the ξ_i were drawn randomly from an Uniform distribution U(10,20) for each simulated sample

The variances are simulated:

- firslty, according to the design given in Figure 3.8 (top left) with the following functions:
$$\sigma^2_{\tau_i} = 0.45\xi_i - 4 \text{ and } \sigma^2_{\nu_i} = 0.45\xi_i - 4, \tag{3.28}$$
where the variances $\sigma^2_{\tau_i}$ and $\sigma^2_{\nu_i}$ increase with ξ_i and are equal for a given ξ_i (it is common in practice that the precision of a measurement method decreases when ξ_i increases) and such that $\sigma^2_{\tau_i} = \sigma^2_{\nu_i} = 0.5$ when $\xi_i = 10$ and $\sigma^2_{\tau_i} = \sigma^2_{\nu_i} = 5$ when $\xi_i = 20$,
- secondly, the variances were chosen independently and randomly ('random heteroscedasticity') from an Uniform distribution U(0.5,5).

The coverage probabilities were computed at a nominal level = 95%. To study in more details the effect of the replicates when the variances are estimated, simulations were also launched:

- with $N = 20$ and $n_X = n_Y = 2$, 4, 6, 8, 10, 12, 15, 20.

Figure 3.9 displays the coverage probabilities, which are close to 95% when the variances are known but closer to 95% for GR. The coverage probabilities collapse drastically when the variances have to be estimated (the number of replicates, $n_X = n_Y = 5$, is too low to estimate a variance) and drop more for GR and especially when N increases or with a random heteroscedasticity. Obviously, when the number of replicates increases, the coverage probabilities increase but BLS still outperforms for $n_X = n_Y = 20$. Actually, the uncertainties on the estimated variances are not taken into account with BLS or GR's methodologies and work is in progress to estimate such regression's lines by taking into account additionally the variances uncertainties.

3.2 Local variances versus variance functions

In Section 3.1.5, we highlighted that the coverage probabilities collapse under heteroscedasticity when the variances are estimated locally. Therefore, we propose two practical solutions to tackle this problem:

- a mathematical transformation can sometimes transform the heteroscedastic variances into homoscedastic variances
- the locally estimated variances can be replaced by predicted variances and plugged into the formulas of the regression line

3.2.1 Data transformation

The design of experiments is a very important step in measurement methods studies. Indeed, the quality of the statistical analysis is related to the design: the standard errors of the estimates, the power of the hypotheses tests,... Although this thesis does not deal in detail with the design of experiments in equivalence, we propose in Figure 3.10 a theoretical example where the design is chosen adequately to tackle the heteroscedasticity by a mathematical transformation of the data.

Figure 3.10-left shows fictive data with $N = 7$ and $X = Y = 1$, 2, 4, 8, 16, 32, 64 (such as successive dilutions in chemistry). The measurement error variances are equal in both axes but increase with respect to X or Y, following the function: $\sigma_\tau^2 = \sigma_\nu^2 = X^2/10$, such that the standard deviations increase

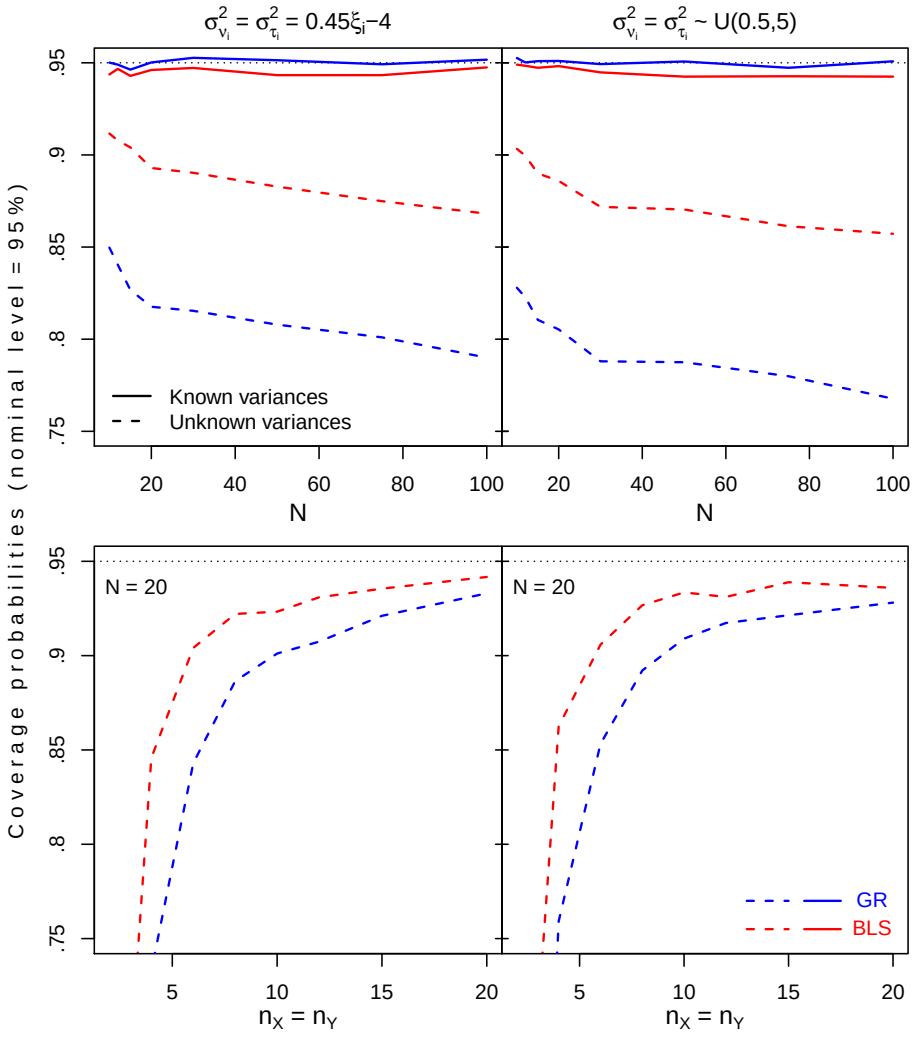

Figure 3.9: *Coverage probabilities of the joint-CI or CB with respect to N (top) or $n_X = n_Y$ (bottom) under heteroscedasticity and known (top) or unknown variances (top and bottom), simulated with a function (left) or randomly (right)*

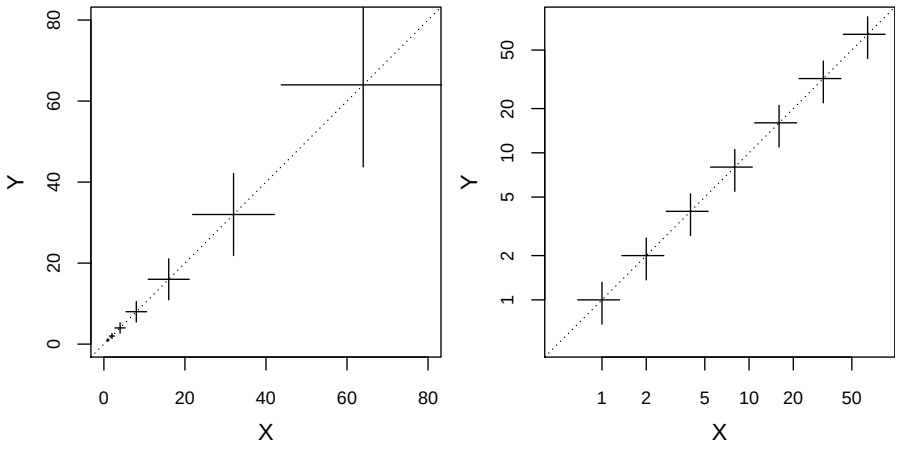

Figure 3.10: *Theoretical example of data set with heteroscedastic errors in a classical* (X, Y) *plot on the left and the corresponding data on the right with logarithmic scales where the errors become homoscedastic*

linearly with respect to X or Y. It is, then, obvious that the measurement error variances are heteroscedastic. In practice, since the variances are never known, they must be estimated locally with replicated data, point by point, as the variances are heteroscedastic. Unfortunately, as explained in Section 3.1.5, the coverage probabilities collapse because of the uncertainties on the locally estimated variances. But, when a logarithmic transformation is applied on both axes, the measurement error variances can become homoscedastic as shown in Figure 3.10-right. Then, a regression analysis can be performed under homoscedasticity. Other transformations can be considered in practice like the square root, inverse function,...

3.2.2 Predicted variances

The uncertainties of locally estimated variances lead to lower coverage probabilities as explained in Section 3.1.5. We propose then a practical solution to tackle this problem. The theoretical background of this practical solution is not rigorous but it will allow to better estimate the regression line under heteroscedasticity by improving the coverage probabilities of the joint-CI.
In practice, the measurement error variances in the X-axis (Y-axis) can sometimes be related to ξ (η). Since ξ_i and η_i are unobservable, we will consider that the variances $\sigma^2_{\tau_i}$ and $\sigma^2_{\nu_i}$ are function of, respectively, X_i and Y_i, like the

four examples in Figure 3.8:

$$\sigma^2_{\tau_i} = f(\xi_i, \theta) \approx f(X_i, \theta)$$

$$\sigma^2_{\nu_i} = f(\eta_i, \theta) \approx f(Y_i, \theta)$$

For instance, in Section 3.1.5, we simulated the heteroscedasticity under equivalence with the following functions: $\sigma^2_{\tau_i} = 0.45\xi_i - 4$ and $\sigma^2_{\nu_i} = 0.45\xi_i - 4$ or, more generally:

$$\sigma^2_{\tau_i} = \vartheta^\tau_0 + \vartheta^\tau_1 \xi_i \text{ and } \sigma^2_{\nu_i} = \vartheta^\nu_0 + \vartheta^\nu_1 \eta_i$$

Therefore, we propose to model the locally estimated variances $S^2_{\tau_i}$ and $S^2_{\nu_i}$ according, respectively, to the observed measures X_i and Y_i. Then, the unknown variances $\sigma^2_{\tau_i}$ and $\sigma^2_{\nu_i}$ can be replaced by the predicted variances $\hat{S}^2_{\tau_i}$ and $\hat{S}^2_{\nu_i}$ instead of locally estimated variances. Such variance functions are common in science, but we did not find any papers dealing with this in method comparison studies. By approximation and simplicity, we propose to model the variances on the X-axis and on the Y-axis, respectively, by the OLSv technique:

$$\hat{S}^2_{\tau_i} = \hat{\vartheta}^\tau_0 + \hat{\vartheta}^\tau_1 X_i \text{ and } \hat{S}^2_{\nu_i} = \hat{\vartheta}^\nu_0 + \hat{\vartheta}^\nu_1 Y_i, \tag{3.29}$$

where $\hat{\vartheta}^\tau_0$, $\hat{\vartheta}^\tau_1$, $\hat{\vartheta}^\nu_0$ and $\hat{\vartheta}^\nu_1$ are, respectively, the OLSv estimators of ϑ^τ_0, ϑ^τ_1, ϑ^ν_0 and ϑ^ν_1. Here are some remarks about variance modeling:

1. In practice, we recommend to plot the locally estimated variances on the X-axis (Y-axis) with respect to the X_i (Y_i) in order to known which model to apply to predict, at best, the variances.

2. As already mentioned, the OLSv is applied by simplicity to predict the variances. Indeed, the X_i and Y_i are observed with error and the response variables, $S^2_{\tau_i}$ and $S^2_{\nu_i}$, of such regressions (3.29) follow χ^2 distribution and not a normal distribution. Note that the standard deviation S_{τ_i} and S_{ν_i} can be modeled instead of the variances but they do not follow either normal distributions but χ distribution.

3. In the case of 'random heteroscedasticity', the locally estimated variances cannot be modeled by a regression. However, it can be supposed that this randomness is sometimes due to an unobserved variable (e.g. in the systolic blood pressure data, the heteroscedasticity could be due to the nervosity of the patients which is not measured).

To assess the improvement of replacing the unknown variances $\sigma^2_{\tau_i}$ and $\sigma^2_{\nu_i}$ by predicted variances instead of locally estimated variances, we simulated 10000 samples:

- under equivalence as described in Section 3.1.5 (with equally spaced values of ξ_i)

- under heteroscedasticity as defined by formulas (3.28) where $\vartheta_1^\tau = \vartheta_1^\nu = 0.45$ and $\vartheta_0^\tau = \vartheta_0^\nu = -4$

The regression line is estimated for each simulated data set by BLS and GR. $\sigma^2_{\tau_i}$ and $\sigma^2_{\nu_i}$ supposed unknown are replaced, respectively, by $S^2_{\tau_i}$ and $S^2_{\nu_i}$ or by $\hat{S}^2_{\tau_i}$ and $\hat{S}^2_{\nu_i}$. Note that sometimes, $\hat{S}^2_{\tau_i}$ or $\hat{S}^2_{\nu_i}$ can be lower than 0, in this case they can be set to 0 or 0.00001 in practice but by this way, the weight of the concerned point is infinity and the model does not reach the convergence. During the ten thousand simulations per value of N, this problem occurred only some dozens of times. It is obvious that this problem is rarer when $\sigma^2_{\tau_i}$ and $\sigma^2_{\nu_i}$ increase (as their probabilities to be predicted negative decrease) or when N, n_X and/or n_Y increase as the variances are then estimated more precisely locally.

Figure 3.11 shows the coverage probabilities given by BLS or GR with respect to N and $n_X = n_Y = 5$ (top-left) or with respect to $n_X = n_Y$ and $N = 20$ (top-right) when the variances are replaced by locally estimated variances or predicted variances. The expectations of the corresponding slopes are also displayed (bottom left and bottom right).
As already mentioned in Section 3.1.5, the coverage probabilities of the joint-CI or CB collapse drastically when the unknown variances are replaced by locally estimated variances $S^2_{\tau_i}$ and $S^2_{\nu_i}$ (especially for GR) but we can notice in Figure 3.11 that these coverage probabilities are close to the nominal level by plugging predicted variances $\hat{S}^2_{\tau_i}$ and $\hat{S}^2_{\nu_i}$ instead of $S^2_{\tau_i}$ and $S^2_{\nu_i}$ into the regression's estimator. However, we can notice in Figure 3.11-bottom that the expectations of the simulated slopes are nearly similar, except for small number of replicates with $N = 20$ (Figure 3.11-bottom right).

To sum-up, replacing the unknown variances by predicted variances instead of locally estimated variances improves the coverage probabilities while the estimated parameters are nearly similar. But, finding a suitable variance function is not always easy in practice.

3.2.3 Choice of variance functions

In this section, different variance functions are applied on the same simulated data set in order to assess the influence of a misspecified variance function on the coverage probabilities. We simulated 10000 samples as described in the previous section. However, the heteroscedasticity is, here, defined with a quadratic function or an exponential function, respectively, as follows:

$\hookrightarrow$ $\sigma^2_{\tau_i} = 0.18\xi_i^2 - 5.4\xi_i + 41$ and $\sigma^2_{\nu_i} = 0.18\eta_i^2 - 5.4\eta_i + 41$
$\hookrightarrow$ $\sigma^2_{\tau_i} = 0.02\ e^{0.3\xi_i}$ and $\sigma^2_{\nu_i} = 0.02\ e^{0.3\eta_i}$

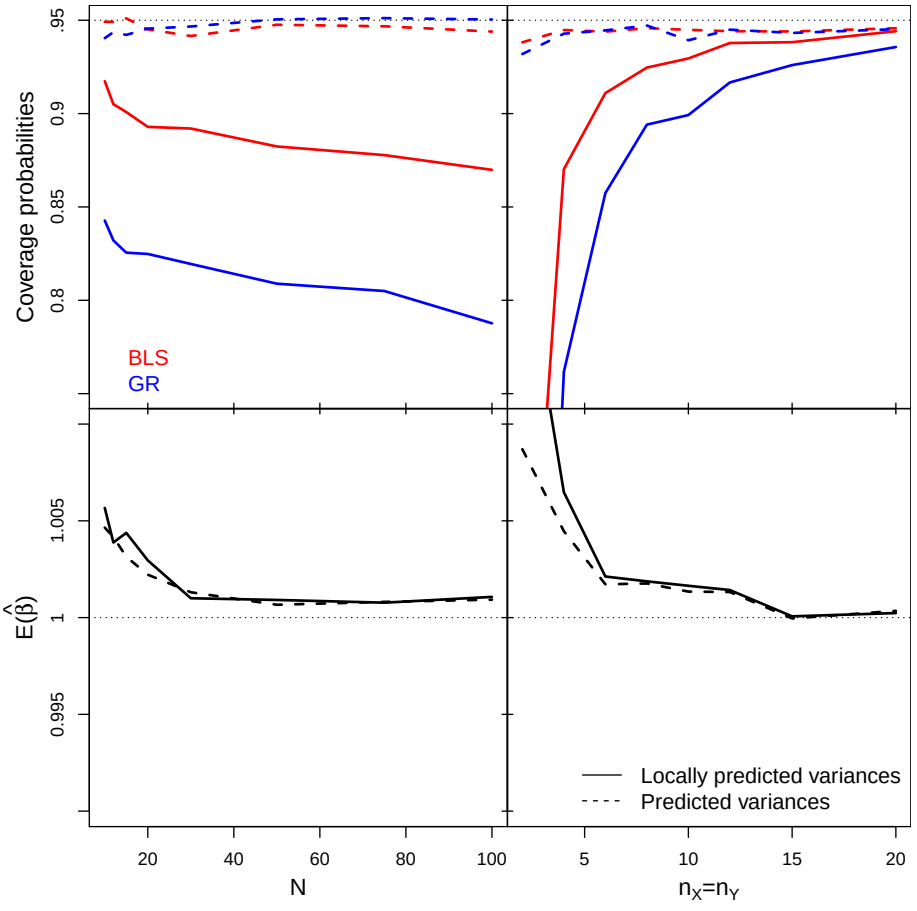

Figure 3.11: *Coverage probabilities (top) of the joint-CI or CB by BLS or GR, with respect to N (left) or $n_X = n_Y$ (right) with heteroscedasticity and locally estimated variances or predicted variances and the biais of the slope β (bottom)*

These variance functions are displayed in Figure 3.12. For each simulated data set, four different variance profiles are, then, applied:

1. none variance profile: the locally estimated variances are not modeled
2. a line where it can be noticed in Figure 3.12 that a line fits approximately well the exponential but not the parabola (on average, the estimated line is horizontal when applied to a parabola)
3. a quadratic function (parabola)
4. an exponential function

Therefore, for each data set, two variances profiles are (maybe) not suitable and one is the true. Figure 3.13 shows the coverage probabilities when the true variance function is quadratic (top) or exponential (bottom) with respect to N ($n_X = n_Y = 5$) (left) or $n_X = n_Y$ ($N = 20$) (right).

When the true variance function is quadratic (Figure 3.13 - top), better coverage probabilities are, evidently, provided when the locally estimated variances are modeled by a quadratic function. Variance functions such as the exponential or the line provide similar coverage probabilities for the BLS and slightly higher to those obtained without modeling the locally estimated variances, except when $n_X = n_Y$ increases (Figure 3.13 - top-right). On the other side, GR provides, curiously, lower coverage probabilities when an exponential function is applied instead of a quadratic, whatever N or $n_X = n_Y$.
When the true variance function is exponential (Figure 3.13 - bottom), better coverage probabilities are, evidently, provided when the locally estimated variances are modeled by an exponential function for the BLS. It can also be noticed that the quadratic and the exponential functions provide very similar results for the BLS. On the other side, GR provides, curiously, slightly better coverage probabilities when a quadratic function is applied instead of an exponential.

3.2.4 Simulated data set under heteroscedasticity

In order to illustrate the dissimilarities between ellipses or hyperbolic confidence bands computed with locally estimated variances or predicted variances, a data set was simulated under equivalence and heteroscedasticity as described in Section 3.2.2 with $N = 20$ and $n_X = n_Y = 5$. The locally estimated variances are modeled by OLSv as explained in Section 3.2.2. Figure 3.14 shows the scatterplot of the simulated data set with the regression line under heteroscedasticity (and locally estimated variances) and standard errors of the means. Figure 3.15 shows the confidence ellipses provided by BLS and GR with heteroscedasticity and the corresponding CB, computed with locally estimated variances or predicted variances. Note that the confidence ellipses provided by BLS or GR are very close to each other but the ones computed with predicted variances are

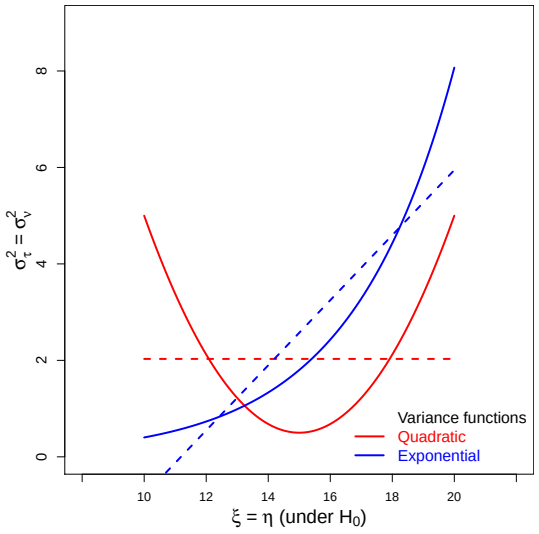

Figure 3.12: *Variance functions applied for the simulations: quadratic or exponential functions, where the dashed lines are the corresponding 'erroneous lines' when a line is estimating instead of the right function*

larger. Indeed, as explained in Section 3.2.2, with locally estimated variances, the coverage probabilities collapse which leads to too narrow joint-CI (these dissimilarities are not easy to visualize with the hyperbolic confidence bands).

3.3 Robustness of the joint-CI to outliers

Some papers, in the literature, deal with robust errors-in-variables regressions [78–80] but not, specifically, with the influence of outliers in method comparison studies. This section compares the bias, the width of confidence ellipses and coverage probabilities of the four different methodologies proposed in Section 3.1.2, in the presence of an outlier.

3.3.1 Outlier under homoscedasticity

There are four types of points in a simple regression [81]:

1. the regular observations
2. the vertical outliers: (X_i, Y_i) points which are not outlying regarding the X-axis but whose Y_i is outlying conditionally to X (the residual is outlying)

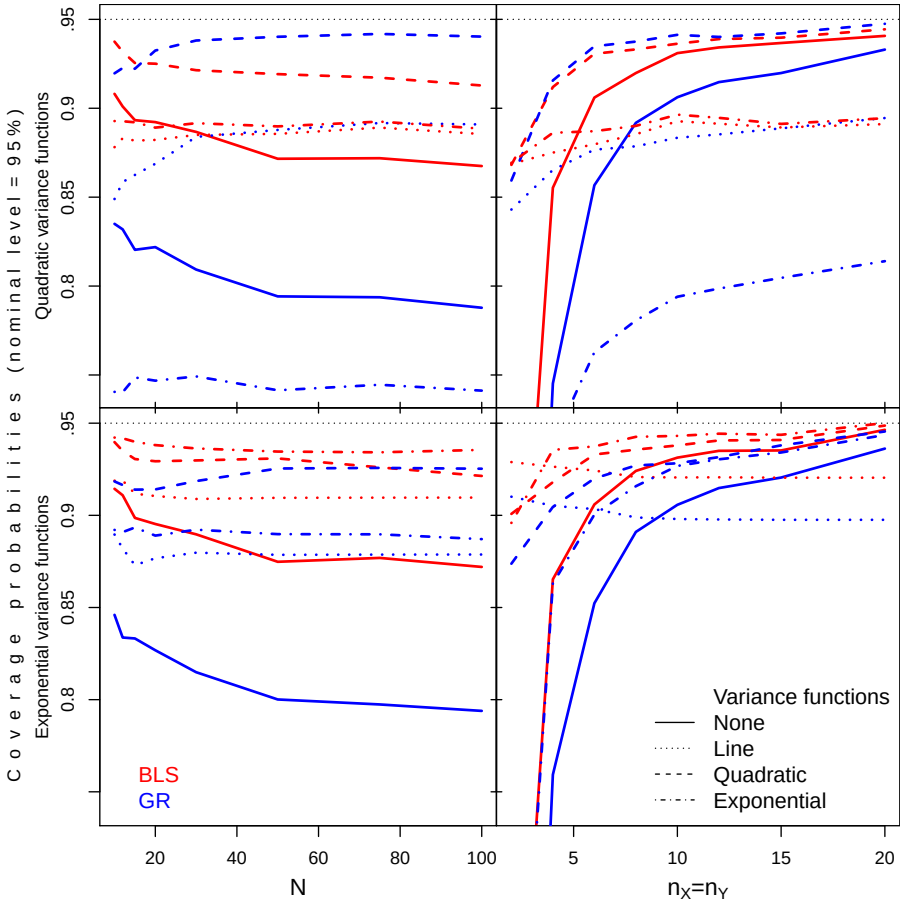

Figure 3.13: *Coverage probabilities of the joint-CI or CB by BLS or GR, with respect to N (left) or $n_X = n_Y$ (right) with heteroscedasticity and locally estimated variances (none variance function) or predicted variances with a line, a quadratic or an exponential function where the true variance function is quadratic (top) or exponential (bottom)*

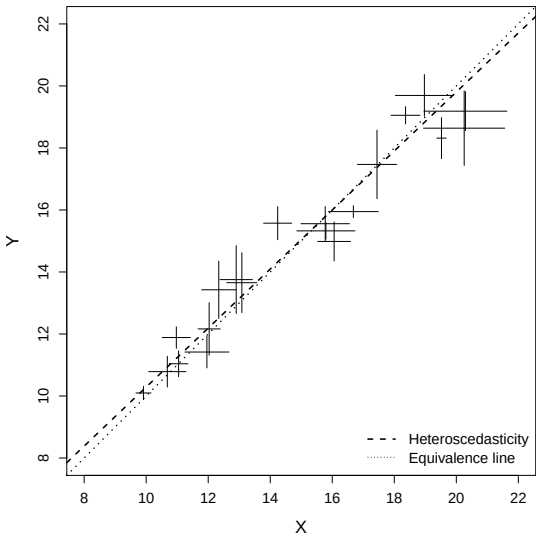

Figure 3.14: *Simulated data set, scatterplot with the regression line under heteroscedasticity (and locally estimated variances) and standard errors of the means*

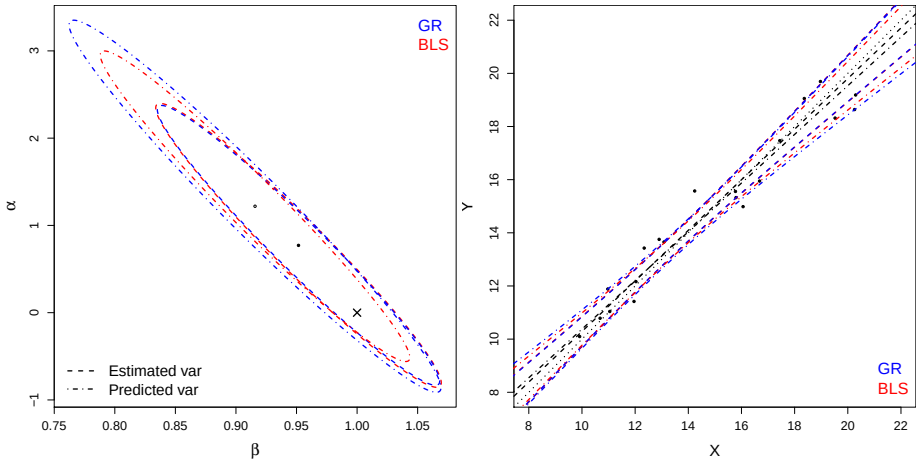

Figure 3.15: *Simulated data set (related to Figure 3.14), confidence ellipses provided by BLS and GR with heteroscedasticity and the corresponding CB, computed with locally estimated variances or predicted variances*

3. the good leverage points: (X_i, Y_i) points that follow the linear pattern but whose X_i are outlying regarding the X-axis
4. the bad leverage points: (X_i, Y_i) points that do not follow the linear pattern and whose X_i are outlying from the majority in X-axis

Vertical outliers and good leverage points will be considered in this section.

Vertical outlier

The ellipses previously presented in Section 3.1.2 are centered on the same point estimator but have different sizes. Moreover, in the presence of outliers, the dissimilarities between the variance-covariance matrices increase which can lead to very different ellipses or hyperbolic CB. In this section, the areas of the ellipses provided by the DR, GR, BLS and Mandel's procedure are compared with or without a vertical outlier. We simulated 10000 samples with $N = 10$, 20, 50 with unreplicated data ($n_X = n_Y = 1$, $\lambda = \lambda_{XY} = 1$ known) as described in the previous section. For each simulated sample, the variance-covariance matrix is computed by the 4 methodologies presented in section 3.1.2 and the two eigenvalues are also computed for each matrix:

$$\lambda_1^{DR}, \lambda_2^{DR}, \lambda_1^{GR}, \lambda_2^{GR}, \lambda_1^{BLS}, \lambda_2^{BLS}, \lambda_1^{Mandel}, \lambda_2^{Mandel}$$

where, for instance, λ_1^{DR} is the larger eigenvalue of the variance-covariance matrix given by the DR. For each simulated sample and for a given regression technique, the product of the square roots of the 2 eigenvalues is computed (as the area of the ellipse is related to this product). Finally, the ratios of these products are computed relatively to the GR's regression:

$$\frac{\lambda_1^{DR}\lambda_2^{DR}}{\lambda_1^{GR}\lambda_2^{GR}}, \frac{\lambda_1^{BLS}\lambda_2^{BLS}}{\lambda_1^{GR}\lambda_2^{GR}} \text{ and } \frac{\lambda_1^{Mandel}\lambda_2^{Mandel}}{\lambda_1^{GR}\lambda_2^{GR}}$$

To 'simulate' a vertical outlier, each simulated sample is modified by translating upwardly the first point (the point with the lowest abscissa) with a distance δ as shown in Figure 3.16. Figure 3.17-left shows the relative areas of the ellipses of the DR, BLS and Mandel's regressions in comparison to the GR with respect to δ from 0 (no outlier) to 10. Figure 3.17-right shows the coverage probabilities of the joint-CI or CB, under equivalence, computed by the 4 methodologies with ($\delta > 0$) or without ($\delta = 0$) a vertical outlier.

Without an outlier, the areas of the ellipses are similar as the ratios are nearly equal to 1 (Figure 3.17-left). Moreover, their coverage probabilities were very similar in Section 3.1.2 and confirmed, here, in Figure 3.17-right for $\delta = 0$.
However, with a vertical outlier, the areas of the ellipses differ increasingly with δ. When $\delta = 10$ and $N = 10$, the Mandel's procedure provides, on average, an ellipse nearly eight times larger than the GR, the BLS nearly six times larger

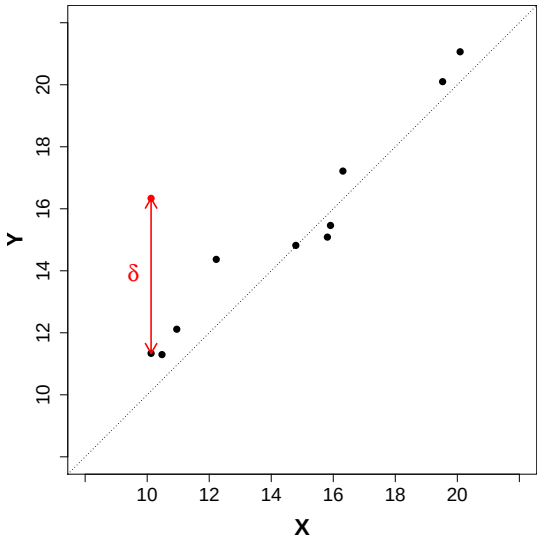

Figure 3.16: *Example of a vertical outlier*

and the DR nearly five times larger than the GR. Indeed, a vertical outlier influence the covariance, S_{xy}, by increasing its value (when the distance from the outlier to the line increases). The variance-covariance matrix given by the DR is, then, modified as it is related to S_{xy} (but it is also related to $\hat{\beta}_{DR}$ which is also modified as the line becomes biased with the outlier). The variance-covariance matrix provided by the BLS is related to the sum of the weighted residuals and obviously this sum increases when a point moves away from the line. On the contrary, the variance-covariance matrix provided by the GR is not related to S_{xy} or the residuals but it is related to W_{GR} and $\hat{x}_i$ which are not modified by the presence of outliers. This effect of one outlier is obviously mitigated when N increases. Figure 3.17-right shows the coverage probabilities under equivalence ($N = 20$ has been discarded for more clarity) where it can be noticed that they drop drastically for the GR's procedure. On the other side, the coverage probabilities provided by the BLS and Mandel's procedure seem to be not affected by the outlier as their joint-CI or CB expand to still cover the null hypothesis.

Good leverage point

Additionally to the previous section, we propose now to study the effect of a good leverage point on the different covariance matrices as previously described.

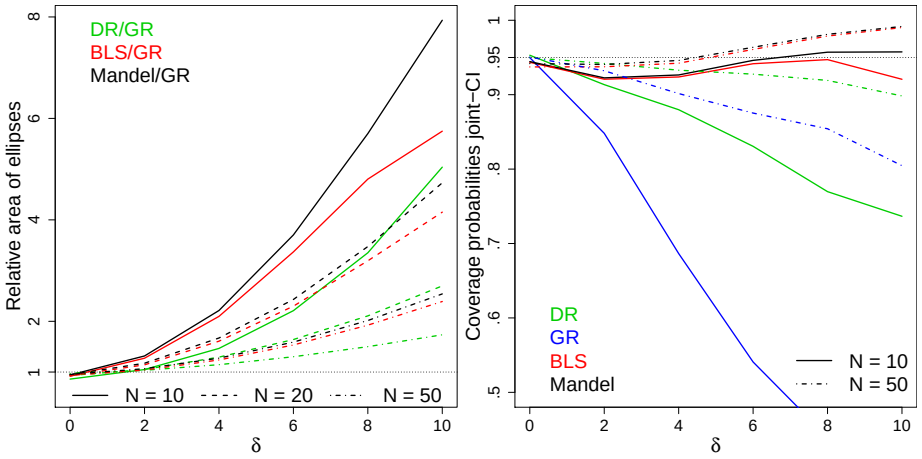

Figure 3.17: *The relative area of ellipses (left) with respect to δ with $N = 10$, 20, 50 and the corresponding coverage probabilities for $N = 10$ and 50 (right)*

To 'simulate' a good leverage point, each simulated sample is modified by translating the last point (the point with the highest abscissa) with a distance δ to the direction of the linear pattern (the direction of the equivalence line) as shown in Figure 3.18. Figure 3.19-left shows the relative areas of the ellipses of the DR, BLS and Mandel's regressions in comparison to the GR with respect to δ from 0 (no outlier) to 10. Figure 3.19-right shows the coverage probabilities of the joint-CI or CB, under equivalence, computed by the 4 methodologies with ($\delta > 0$) or without ($\delta = 0$) a good leverage point.

The DR provides the smallest confidence ellipse for small sample sizes ($N = 10$) whatever δ, so whatever the presence or absence of a good leverage point (Figure 3.19-left). Otherwise, the ellipses have nearly the same areas as their ratios are between 0.9 and 1 in comparison to the GR ellipses. This means that the ellipses are not 1.1 times larger than other ellipses, whatever δ.
The coverage probabilities are very close to each other and to 95% (Figure 3.19-right) whatever δ. But, these coverage probabilities decrease slightly for, approximately, $\delta > 5$ and reach 93% for $\delta = 10$ (the maximal value of δ in our simulations) with $N = 50$. For small sample sizes, i.e. $N = 10$, the coverage probabilities seem not to be affected by the outlier.

3.3.2 Outlier under heteroscedasticity

This section considers the presence of a vertical outlier under heteroscedasticity and its influence on the coverage probabilities and on the bias of the slope.

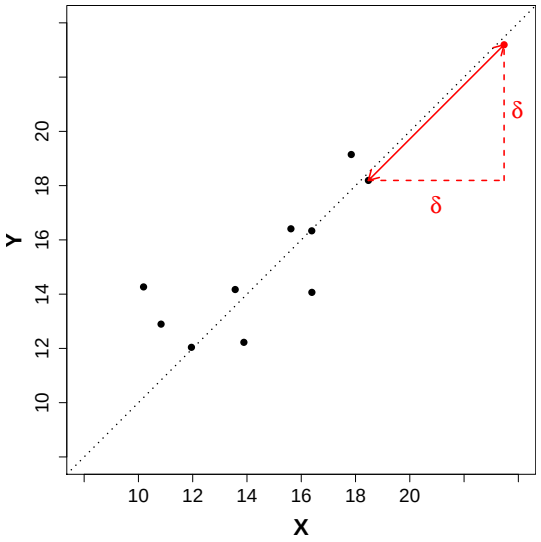

Figure 3.18: *Example of a good leverage point*

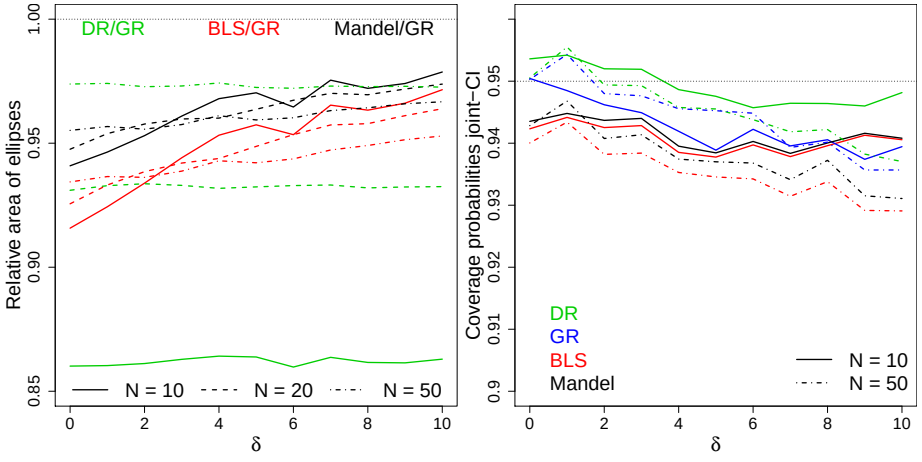

Figure 3.19: *The relative area of ellipses (left) with respect to* δ *with* $N = 10$*, 20, 50 and the corresponding coverage probabilities for* $N = 10$ *and 50 (right)*

Vertical outlier

In this section, the effect of a vertical outlier with model (1.10) is assessed. We simulated 10000 samples under equivalence with ξ_i and η_i values as described in Section 2.3, with $N = 10$, 20 and 50.

1. The variances are supposed known with $\sigma^2_{\nu_i} = \sigma^2_{\tau_i} = 1$ and $n_{X_i} = n_{Y_i} = 1\ \forall i$. The coverage probabilities of the joint-CI were, then, computed by BLS and GR regressions under homoscedasticity.
2. The first point, $X_{(1)}, Y_{(1)}$, of each simulated sample was moved upwardly with $\delta = 10$ as described in Section 3.3.1. The variance of the measurement error of this outlier was set then:
   - from 0.75 to 4.75 in both axes ($\sigma^2_{\nu_{(1)}} = \sigma^2_{\tau_{(1)}}$ from 0.75 to 4.75),
   - or one axis ($\sigma^2_{\nu_{(1)}}$ from 0.75 to 4.75 and $\sigma^2_{\tau_{(1)}} = 4.75$ or $\sigma^2_{\tau_{(1)}}$ from 0.75 to 4.75 and $\sigma^2_{\nu_{(1)}} = 4.75$) in order to simulate a precise outlier or imprecise outlier,

   and the coverage probabilities were computed under heteroscedasticity.

Figure 3.20-left shows the coverage probabilities provided by GR and BLS for $N = 10$ with or without a vertical outlier. Figure 3.20-right shows the expectation of the corresponding slopes $\hat{\beta}$ (recall that $\hat{\beta}_{GR} = \hat{\beta}_{BLS}$). In Figure 3.20-top, the precision of the outlier varies in both axes while in Figure 3.20-middle, the precision of the outlier varies only in the X-axis with $\sigma^2_{\nu_{(1)}} = 4.75$ constant in the Y-axis, and in Figure 3.20-bottom, the precision of the outlier varies only in the Y-axis with $\sigma^2_{\tau_{(1)}} = 4.75$ constant in the X-axis. Figure 3.21 and Figure 3.22 are similar, for respectively, $N = 20$ and $N = 50$.

The coverage probabilities are:

- without an outlier, close to 95% (Figure 3.20-left) as already mentioned in Section 3.1.5 with known variances.
- with a vertical outlier, nearly constant for BLS and higher than 95% whatever the outlier is precise in one or both axes. On the other hand, the coverage probabilities for GR drop drastically: firstly, when the outlier is precise in both axes (small variances in both axes, Figure 3.20-top left) and secondly, when the outlier is precise in the X-axis (small values of $\sigma^2_{\tau_{(1)}}$, Figure 3.20-middle left). However, the coverage probabilities for GR seem to be not influenced by an outlier precise only on the Y-axis (Figure 3.20-bottom left) by being closer to the nominal level.
- when the precision of the outlier decreases (the variance(s) increase(s)), move closer to the nominal level for GR as the weight of the outlier decreases.
- when the sample size N increases and with the presence of a vertical outlier, move closer to the nominal level as the weight of the outlier decreases in comparison the N (Figures 3.21-3.22).

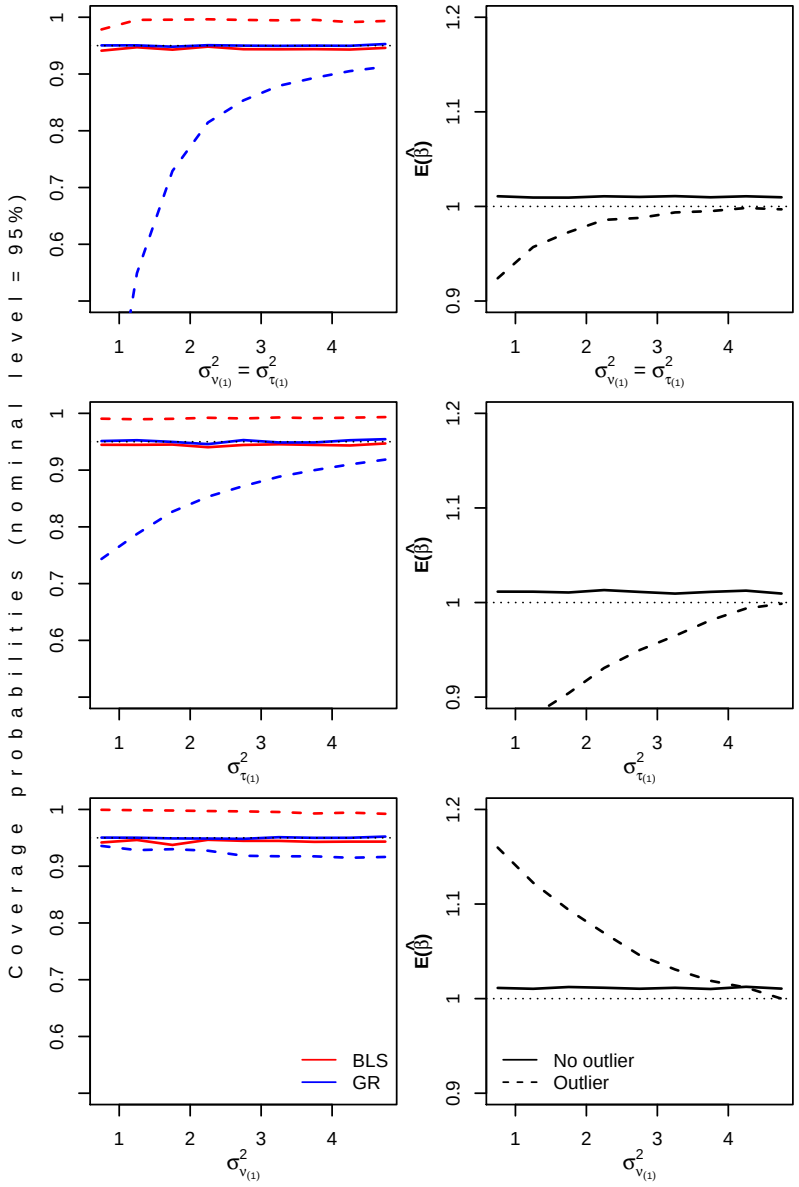

Figure 3.20: *Coverage probabilities for $N = 10$ provided by GR and BLS with or without an outlier (left) and the expectation of the corresponding slopes (right), when the precision of the outlier varies in both axes (top) or only in X-axis (middle) or only in Y-axis (bottom)*

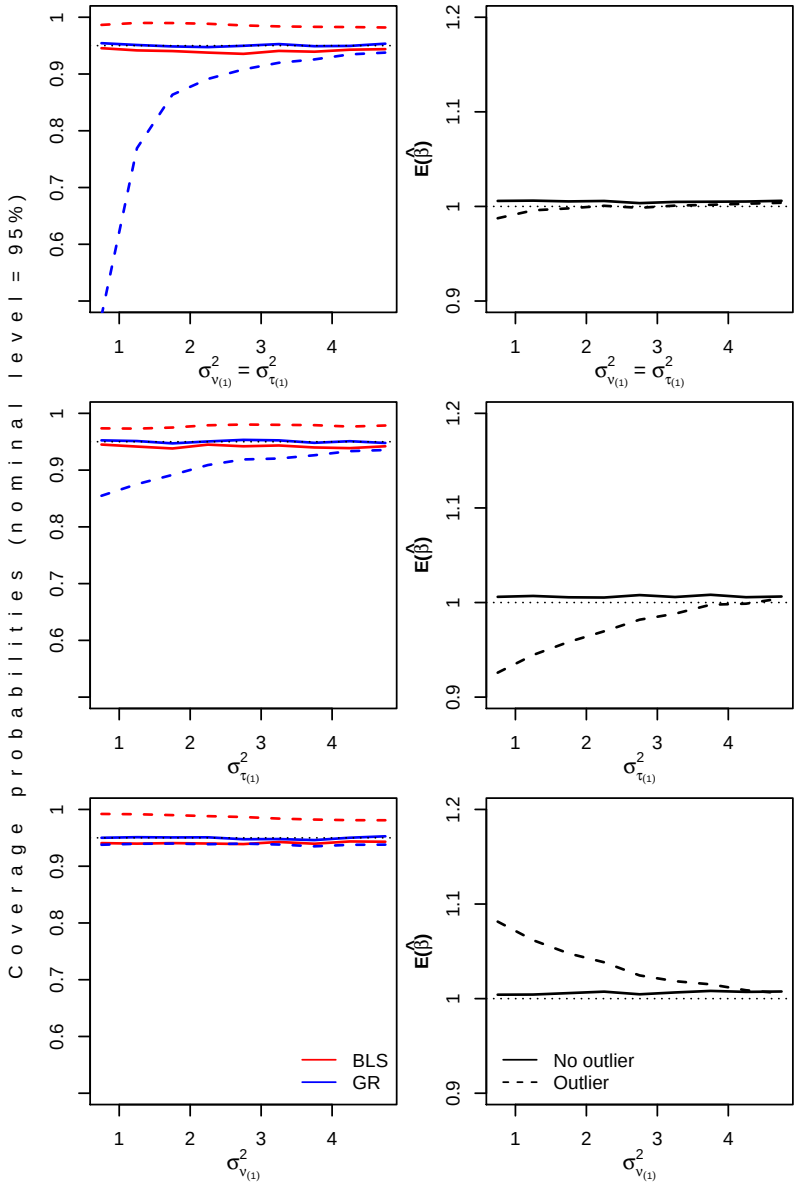

Figure 3.21: *Coverage probabilities for* $N = 20$ *provided by GR and BLS with or without an outlier (left) and the expectation of the corresponding slopes (right), when the precision of the outlier varies in both axes (top) or only in X-axis (middle) or only in Y-axis (bottom)*

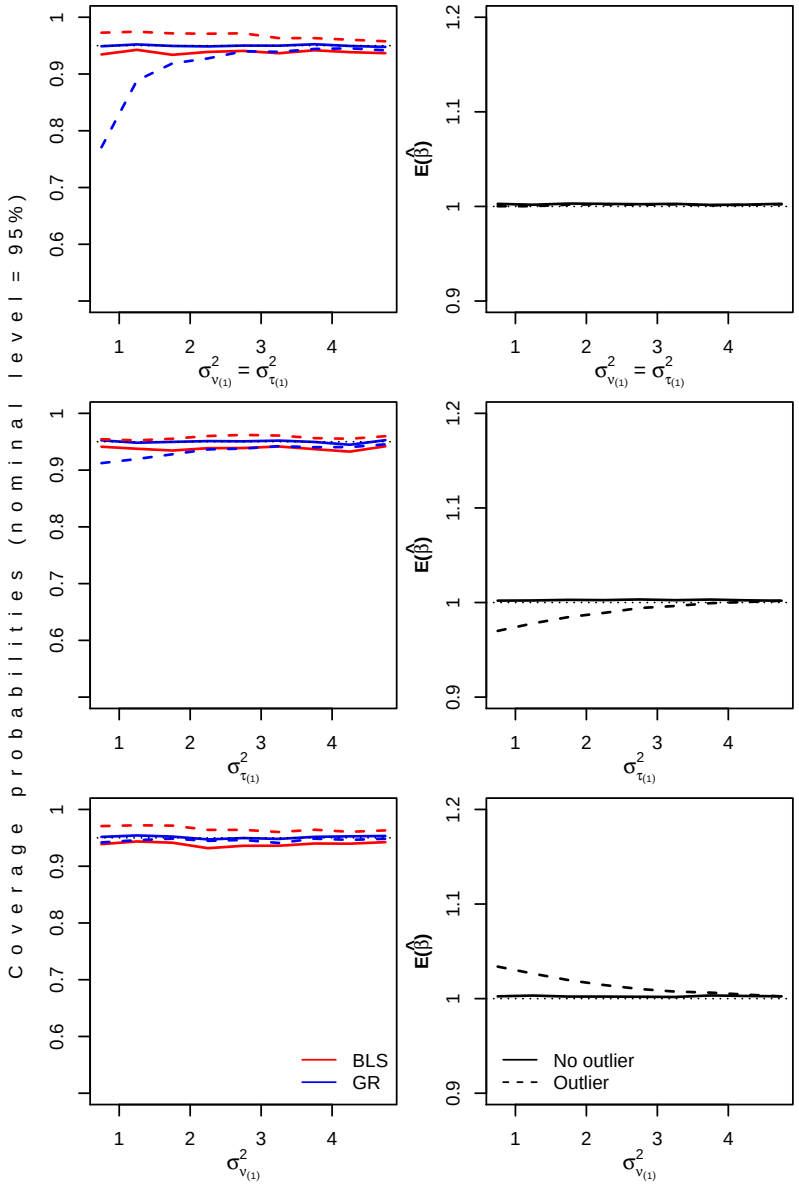

Figure 3.22: *Coverage probabilities for $N = 50$ provided by GR and BLS with or without an outlier (left) and the expectation of the corresponding slopes (right), when the precision of the outlier varies in both axes (top) or only in X-axis (middle) or only in Y-axis (bottom)*

The slopes of the regression lines are:

- biased with an outlier (Figure 3.20-right).
- obviously, underestimated with an outlier precise in both axes (small variances in both axes, Figure 3.20-top right). This bias increases with an outlier precise only in the X-axis (Figure 3.20-middle right).
- overestimated with an outlier precise in the Y-axis (Figure 3.20-bottom right).

These biases, of course, decrease when the sample size N increases (Figures 3.21-3.22).

In order to illustrate these findings, we simulated a data set under homoscedasticity and equivalence with $N = 10$, $n_X = n_Y = 1$ and known variances as previously described ($\sigma^2_{\nu_i} = \sigma^2_{\tau_i} = 1 \; \forall i$). Firstly, the regression line is estimated by BLS and GR under homoscedasticity. Secondly, the first point (the smallest abscissa), $(X_{(1)}, Y_{(1)})$, was moved upwardly (with $\delta = 10$) as previously described. The variance of the measurement error of this outlier was set then to 0.75 in both axes (precise outlier in both axes) or one axis (precise outlier in one axis, with the variance of the measurement error in the other axis set to 4.75) and the regression line was estimated under heteroscedasticity by BLS and GR.
Figure 3.23-left shows the simulated data set (with standard errors) and the estimated regression line under homoscedasticity without the outlier (continuous line). The dashed line is the regression line estimated under heteroscedasticity by taking into account the outlier (with standard errors in dashed lines):

- with a precise outlier in both axes (Figure 3.23-top left), the estimated line moves closer to the outlier by underestimating the slope β of the regression line (as explained in the previous section).
- with an outlier precise in X-axis (Figure 3.23-middle left), the estimated line moves also closer to the outlier by underestimating β but this observed bias is lower than the one with a precise outlier in both axes. Nevertheless, on average, we showed with the simulations that the regression line moves closer to the outlier by underestimating β much more with an outlier precise in X-axis rather than an outlier precise in both axes.
- with an outlier precise in Y-axis (Figure 3.23-bottom left), the estimated line moves closer to the outlier by overestimating β.

Figure 3.23-right shows the confidence ellipses provided by BLS and GR under heteroscedasticity by taking into account the outlier. The BLS ellipses are larger than the ones given by GR (we also explained in Section 3.3.1 that BLS provides larger ellipses with an outlier), firstly with an outlier precise in both axes, secondly with an outlier precise in X-axis and thirdly with an outlier precise in Y-axis. Finally, without the outlier, the regression line is estimated

more precisely and the ellipses are narrower than the ones estimated with an outlier (and superimposed on the plot).

3.4 Applications

The systolic blood pressure data have already been analyzed under homoscedasticity in Chapter 2. This section proposes to compare in more details the confidence ellipses and the corresponding confidence bands provided by the four methodologies presented in this chapter: DR, Mandel, BLS and GR procedure. Additionally, the systolic blood pressure data are here analyzed under heteroscedasticity together with the arsenate ion in natural water data.

3.4.1 The systolic blood pressure (replicated data) under homoscedasticity

Figure 3.24 displays the scatterplot of the data with standard errors of the mean ($S_\tau/\sqrt{n_X}$ and $S_\nu/\sqrt{n_Y}$). Figure 3.25-right displays the 95% hyperbolic confidence bands computed by the four methodologies (Mandel, DR, BLS, GR) and one can see that the equivalence line ($Y = X$) is not inside the confidence bands which means that the manual device and the semi-automatic one are not equivalent: the null hypothesis is rejected, the estimated line is significantly different from the equivalence line. The confidence bands provided by Mandel and BLS are very close to each other (nearly superimposed on the graph) and close to DR when we move away from $\bar{X}$ while the GR is the narrowest. In fact, there are many vertical outliers which leads to ellipses (or CB) with larger areas for the Mandel, BLS or DR in comparison to GR, as explained in Section 3.3.1. This can also be noticed in Figure 3.25-left where the equivalence point ($\beta = 1$ and $\alpha = 0$) is outside all the ellipses and the one provided by GR is the smallest. The minor axes of the GR and DR ellipses are equal and by analogy, the widths of their hyperbolic CB are equal at $(\overline{X}, \overline{Y})$. The Mandel and BLS's ellipses are nearly 9 times larger than the GR's ellipse and the DR's ellipse nearly 3 times larger than the GR's ellipse.

3.4.2 The systolic blood pressure (replicated data) under heteroscedasticity

If we suppose heteroscedasticity in the systolic blood pressure data, the variances $\sigma^2_{\tau_i}$ and $\sigma^2_{\nu_i}$ are unknown but can be estimated with $S^2_{\tau_i}$ and $S^2_{\nu_i}$ from the replicates for each patient and both devices ($n_{X_i} = n_X = n_{Y_i} = n_Y = 3$ $\forall i$). The estimates GR and BLS regression lines are: $\hat{\beta}_{GR} = \hat{\beta}_{BLS} = 0.960$ and $\hat{\alpha}_{GR} = \hat{\alpha}_{BLS} = 18.913$.
Figure 3.26 displays the scatterplot of the data with local standard errors of the

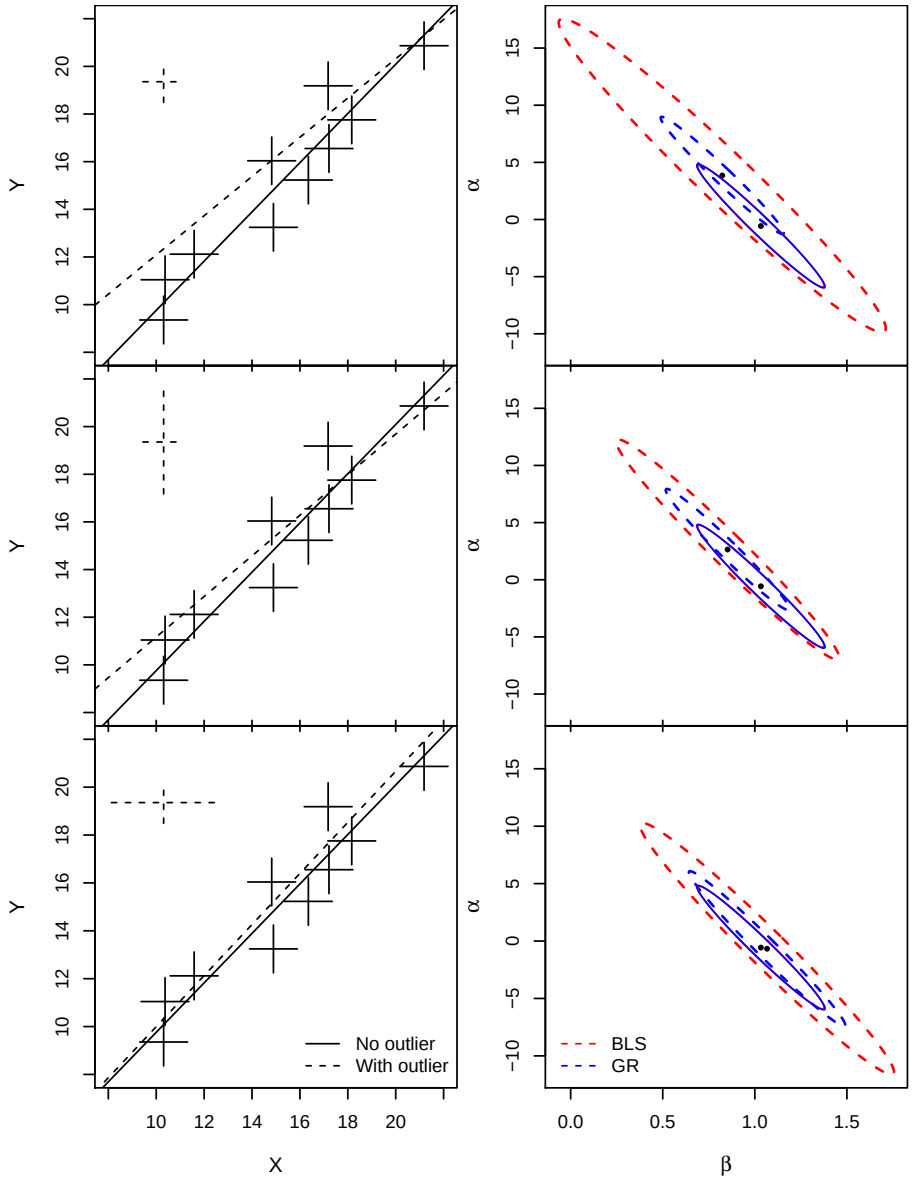

Figure 3.23: *Simulated data set ($N = 10$) under equivalence and homoscedasticity without outlier (continuous line) or with an outlier (dashed lines): precise outlier in both axes (top), in X-axis (middle) or in Y-axis (bottom). (X, Y) data and regression lines on the left and the corresponding confidence ellipses given by BLS and GR on the right under heteroscedasticity by taking into account the outlier or under homoscedasticity (without outlier)*

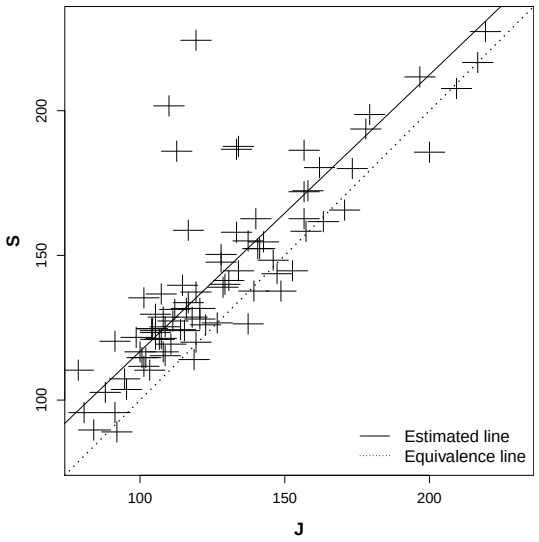

Figure 3.24: *SBP data, scatterplot with the regression line under homoscedasticity and standard errors of the mean*

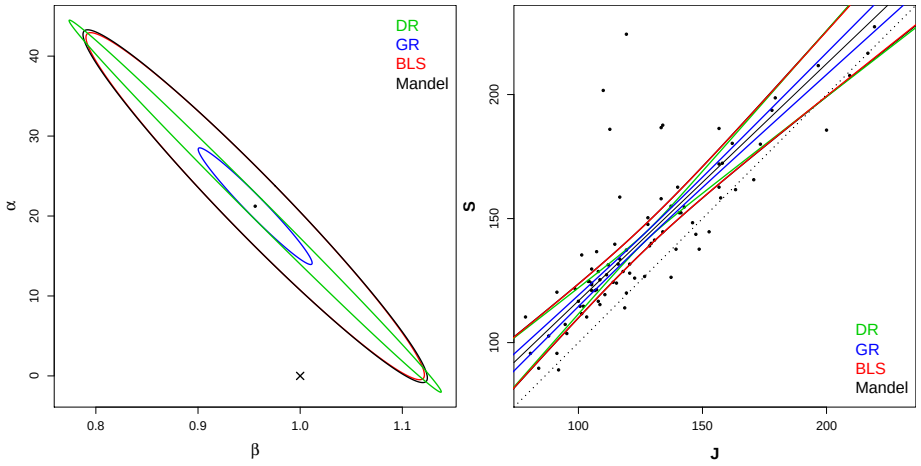

Figure 3.25: *SBP data, confidence ellipses (left) computing by the four methodologies and the corresponding CB (right)*

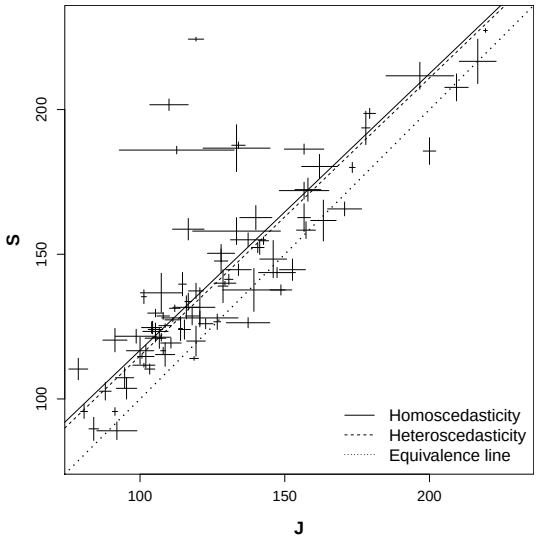

Figure 3.26: *SBP data, scatterplot with the regression's lines under heteroscedasticity and standard errors of the mean*

means. The estimated line is close to the one estimated under homoscedasticity assumption (the parameters are very close to each other). The 95% hyperbolic CB provided by the BLS and GR under heteroscedasticity are displayed in Figure 3.27-right with the ones under homoscedasticity (for comparative purpose). It can be noticed that the CB are narrower under heteroscedasticity. Actually, some of the outliers have lower weights and are less taken into account in the heteroscedasticity analysis which leads to lower variances of the parameters as explained in Section 3.3.2. Like in the previous section, the equivalence line ($Y = X$) is not inside the CB: the equivalence between both devices is rejected. This is confirmed in Figure 3.27-left where the equivalence point ($\beta = 1$ and $\alpha = 0$) is outside all the ellipses and the ones provided by the GR are the narrowest (under homoscedasticity and heteroscedasticity).

3.4.3 Arsenate ion in natural river water under heteroscedasticity

In the arsenate ion in natural river water data [16], 30 pairs of measures are provided by 2 methods: firstly, a continuous selective reduction and atomic absorption spectrometry and secondly, a non-selective reduction, cold trapping and atomic emission spectrometry. The mean measures (X_i, Y_i) with their standard errors of the mean are given and analyzed in the literature [16,17].

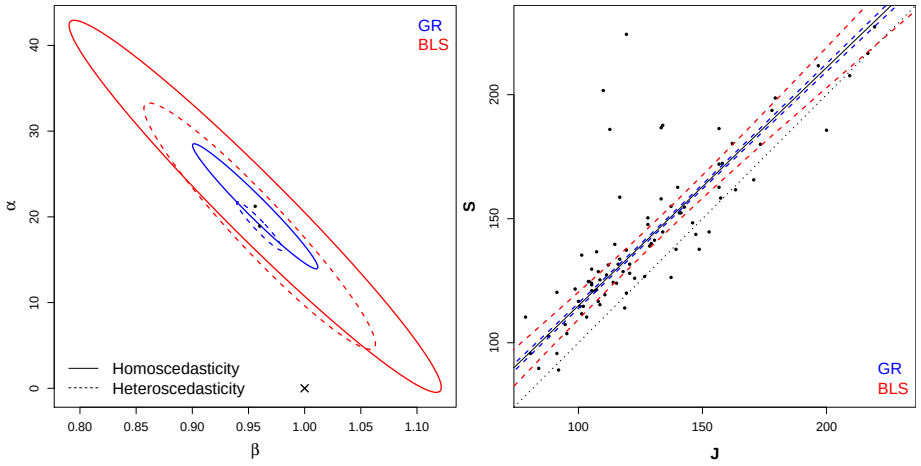

Figure 3.27: *SBP data, confidence ellipses provided by BLS and GR with or without heteroscedasticity (left) and the corresponding CB under heteroscedasticity (right)*

Figure 3.28 shows the scatterplot of the data. We can observe that lower concentrations are frequent while higher concentrations are rarer (logarithmic scales could be tested). The errors increase with concentration for both devices (like our simulations in Section 3.1.5). Figure 3.29 shows the scatterplot of the locally estimated standard errors of the means X_i (Y_i) with respect to X_i (Y_i) and their estimated regression OLSv line. We can notice that the OLSv regression line fits very well the data as the points are very close to it with $R^2 = 0.9999872$ ($R^2 = 0.9999874$). This means that the predicted variances by the OLSv technique are very close to the locally estimated variances.
Figure 3.30 shows the confidence ellipses for the parameters (left) and the corresponding confidence bands (right). The estimated line is close to the equivalence line (Figure 3.28) and is inside the hyperbolic confidence bands (Figure 3.30-right) provided by BLS or GR. Obviously, the equivalence point is inside both ellipses (Figure 3.30-left). Both CB and both ellipses are very close to each other at the opposite of the systolic blood pressure example. This is because there is no vertical outlier in this data set.
Ideally, the data should be transformed into logarithm but the detailed data are not provided in the related papers where they are analyzed without such transformation [16,17]. We can finally notice, that the ellipses or hyperbolic CB computed by predicting the variances (the standard errors of the means) are very close to the ones computed with locally estimated variances, for both regressions (BLS and GR).

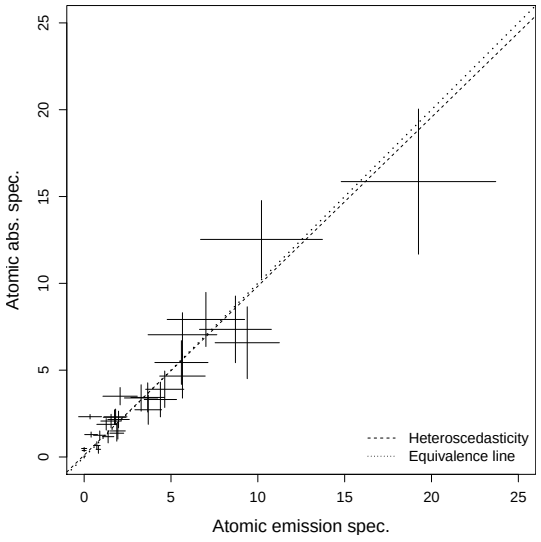

Figure 3.28: *AsO_4 data, scatterplot with the regression line under heteroscedasticity and standard errors of the means*

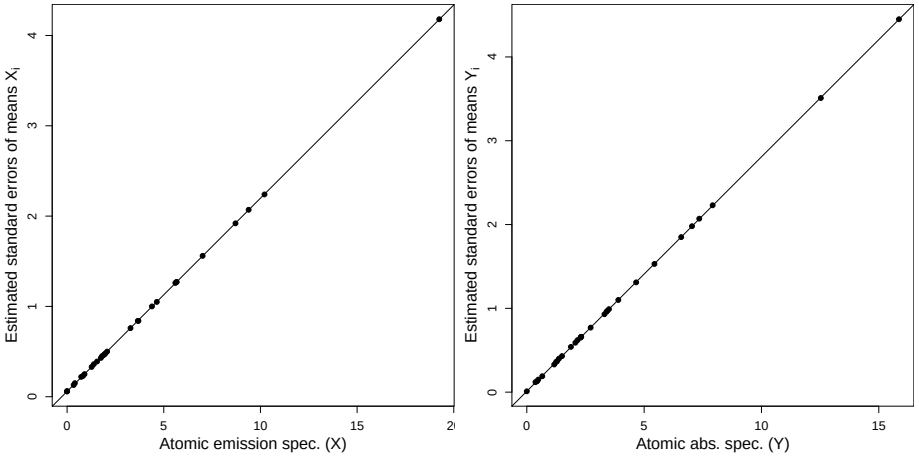

Figure 3.29: *AsO_4 data, scatterplots of the locally estimated standard errors of the means X_i (Y_i) with respect to X_i (Y_i) and their estimated regression OLSv line*

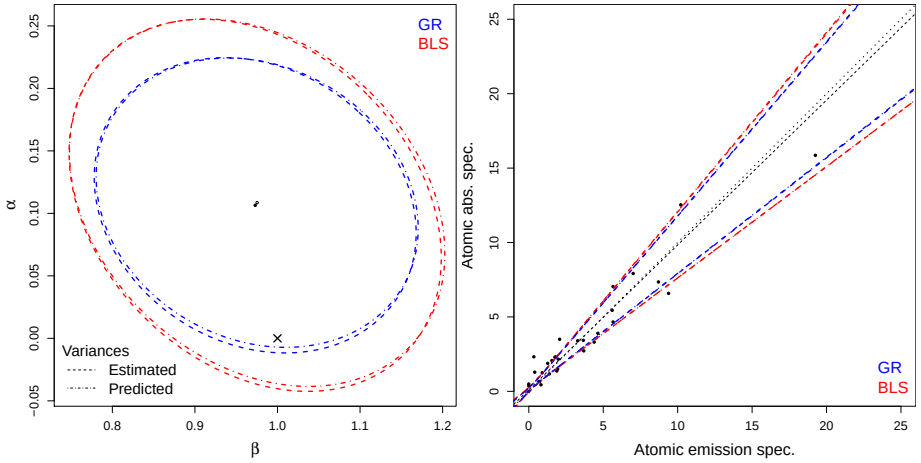

Figure 3.30: *AsO_4 data, confidence ellipses provided by BLS and GR with heteroscedasticity and the corresponding CB*

Chapter 4

Bland and Altman plot: method of differences versus (X, Y) plot

Contents

The most known and widely applied approach in method comparison studies is certainly the one proposed by Bland and Altman which focuses directly on the differences between the two measurement methods of interest [11–13] (their main paper have been cited more than ten thousand times [82]). This chapter reviews the Bland and Altman approach and proposes some improvements. Then, the (X, Y) plot and the Bland and Altman approach will be compared.

4.1 Practical equivalence without replicates

Remind that the goal of practical equivalence is to assess whether the differences observed between two measurement methods, $D = Y - X$, are clinically meaningful or not. In practice, an agreement interval where, for instance, 95% of the future differences are expected to lie, is computed and compared to an acceptance interval $[-\Delta, \Delta]$. If the agreement interval is included inside the acceptance interval, the differences are not clinically important and the two measurement methods are interchangeable in practice. Although the 'agreement intervals' popularized by Bland and Altman [11] are widely applied in practice, it will be explained in this chapter that predictive or tolerance intervals are a better alternative.
In this section, we suppose that X_i and Y_i correspond each to a single measure (there is no replicates).

4.1.1 Horizontal agreement interval

Instead of plotting the Y values with respect to the X values, Bland and Altman propose to plot the differences (on the Y-axis) with respect to the averages (on the X-axis):

$$D_i = Y_i - X_i \text{ and } M_i = \frac{X_i + Y_i}{2} \tag{4.1}$$

We have also:

$$\overline{D} = \frac{1}{N}\sum_{i=1}^{N} D_i \text{ and } S_D^2 = \frac{1}{N-1}\sum_{i=1}^{N}\left(D_i - \overline{D}\right)^2, \tag{4.2}$$

where $\overline{D}$ and S_D^2 are the estimators of, respectively, μ_D and σ_D^2.
Bland and Altman propose to compute an agreement interval (AI, or agreement limits LoA) [11] such that $(100 - \psi)\%$ of the future differences are expected to lie into this interval:

$$\text{AI: } \overline{D} \pm z_{1-\psi/2} S_D \tag{4.3}$$

where $z_{1-\psi/2}$ is the $100(1 - \psi/2)\%$ percentile of the standardized normal distribution.

This interval, very easy to compute, is widely applied in the literature. In practice, the equivalence is not rejected when the acceptance interval is included inside the agreement interval, in other words, the differences between the two measurement methods are then not clinically important and both measurement methods can be interchangeable.
Unfortunately, this interval is computed with the estimated mean difference $\overline{D}$ and standard deviation of the differences S_D and the uncertainty of S_D is not taken into account (as the normal distribution is applied). To improve their approach, Bland and Altman propose then to compute a confidence interval on each limit of the agreement interval [11]. However, the goal of these confidence intervals is not well-argued in their papers (is it used to take into account the uncertainty of S_D or to test that the acceptance interval is significantly inside the agreement interval?) and they are less often applied in practice than the simple agreement interval.
Computing confidence intervals on each limit of an interval may be viewed as absurd but the agreement interval can be interpreted as a predictive interval. So, each limit of the agreement interval can be considered as an estimated quantile: the lower limit of AI is the estimated quantile $q_{\psi/2}$ of the distribution of the differences D_i and the upper limit of AI is the estimated quantile $q_{1-\psi/2}$. Therefore, a confidence interval can be computed on each limit of the agreement interval to assess the uncertainties on the estimated quantiles. In other words, it is analogous to compute confidence intervals for two quantiles. The variance of the limits of the AI is given by the approximate following formula [11]:

$$Var(AI_{unreplicated}) = S_D^2 \left(\frac{1}{N} + \frac{z_{1-\psi/2}^2}{2(N-1)} \right) \tag{4.4}$$

The $100(1-\gamma)\%$ confidence intervals on the agreement limits were initially proposed by Bland and Altman [13] by adding or subtracting the quantity $\sqrt{3S_D^2/N}$ to, respectively, the upper or lower agreement limit. However, this solution was an approximation and they proposed later to compute the $100(1-\gamma)\%$ confidence intervals on each limit of the agreement interval by the following formula [11]:

$$\text{CI(AI): } (AI) \pm t_{1-\gamma/2,N-1} S_D \sqrt{\frac{1}{N} + \frac{z_{1-\psi/2}^2}{2(N-1)}} \tag{4.5}$$

The agreement interval can then be replaced by an interval from the lower bound of the lower agreement limit to the upper bound of the upper limit. This interval will be called, in this thesis, the XL-AI.

4.1.2 Horizontal β-expectation and $\beta-\gamma$ content tolerance intervals

Tolerance intervals are a better alternative than agreement interval. The concept of tolerance interval is not new and some of very well-known statisticians, like Wald or Wolfowitz [83–85], already published papers in the forties dealing with tolerance intervals under normal distribution [86] and later with improvements [87] or under other assumptions.

The $100(1-\psi)\%$ β-expectation tolerance interval (β-TI) is a predictive interval where $100(1-\psi)\%$ of the future differences will lie **on average** into it:

$$E_{\overline{D},S_{D_s}}\left(P\left(\overline{D}-k_eS_{Ds}<D<\overline{D}+k_eS_{Ds}\right)\right)=1-\psi$$

where k_e is a constant. The $100(1-\psi)\%$ β-TI is given by [88]:

$$\overline{D}\pm t_{1-\psi/2,N-1}S_D\sqrt{1+\frac{1}{N}} \tag{4.6}$$

This interval is exact whatever N.

The $100(1-\psi)\%$ $\beta-\gamma$ content tolerance interval ($\beta\gamma$-TI) is an interval where **at least** $100(1-\psi)\%$ of the future differences will lie into it with a $100(1-\gamma)\%$ confidence level (the $\beta\gamma$-TI contains at least $100(1-\psi)\%$ of the future differences in $100(1-\gamma)\%$ of cases):

$$P_{\overline{D},S_{D_s}}\left(P\left(\overline{D}-k_cS_{Ds}<D<\overline{D}+k_cS_{Ds}\right)\geq 1-\psi\right)=1-\gamma$$

where k_c is a constant. The formula of the $100(1-\psi)\%$ $\beta\gamma$-TI is quite complex but it can be computed easily by the following formula:

$$\overline{D}\pm k_cS_D$$

where k_c is given in tables in literature [88]. However, an approximate formula for the $100(1-\psi)\%$ $\beta\gamma$-TI is given by [87]:

$$\overline{D}\pm z_{1-\psi/2}S_D\sqrt{1+\frac{1}{N}}\sqrt{\frac{N-1}{\chi^2_{\gamma,N-1}}} \tag{4.7}$$

where the normal distribution, z, is related to the predictive level while the Chi square distribution, χ^2, is related to the desired confidence level.

4.1.3 Comparison and coverage probabilities of the agreement and tolerance intervals

We did not find in the literature a detailed comparison of the agreement intervals (there is no simulations in the papers published by Bland and Altman)

with the tolerance intervals. In this section, we propose to compare them with simulations.

First, as explained previously, the agreement interval or predictive intervals can be interpreted as quantiles estimators. Let's define the $100(\psi)\%$ quantile of the differences by q^D_ψ. Then, it is easy to check that the $(100-\psi)\%$ agreement interval moves closer to the quantiles $q^D_{\psi/2}$ and $q^D_{1-\psi/2}$ when $N \to \infty$ as $t \to z$, $\overline{D} \to \mu_D$ and $S_D \to \sigma_D$. Moreover, the confidence intervals on the agreement interval collapse on themselves when $N \to \infty$.

To sum-up, when the sample size increases, the agreement interval and the tolerance intervals move closer to each other and to the quantiles $q^D_{\psi/2}$ and $q^D_{1-\psi/2}$.

In order to compare the coverage probabilities of these intervals, we run 100000 simulations per value of N (from 5 to 100: 5, 6, 7, 8, 9, 10, 12, 15, 20, 25, 30, 50, 75, 100) under equivalence with ξ_i values as described in Section 2.5.1 with $\sigma^2_D = \sigma^2_\tau = \sigma^2_\nu = 0.75$ (and therefore $\lambda_{XY} = 1$) unknown and $n_X = n_Y = 1$ (actually, the choice of σ^2_τ and σ^2_ν does not matter in this Section). Note that the variances are supposed unknown and the data are not replicated but the variance σ^2_D can still be estimated by S^2_D. For each simulated data, the agreement interval with its confidence intervals and the tolerance intervals are computed. The predictive level, $1-\psi$, is fixed to 95% and the confidence level for the confidence intervals on the agreement interval or the $\beta\gamma$-TI is set to 50, 60, 70, 80 or 90%. So, for each simulated data set, 12 intervals are computed:

- 1 agreement interval with 95% predictive level
- 5 XL-AI with 95% predictive level and confidence levels from 50 to 90%
- 1 β tolerance interval with 95% predictive level
- 5 $\beta\gamma$ tolerance intervals with 95% predictive level and confidence levels from 50 to 90%

The coverage probabilities of each interval are computed with the normal distribution where the mean equals 0 (as the data sets are simulated under equivalence) and the variance equals $\sigma^2_\tau + \sigma^2_\nu = 1.5$. The confidence level of the upper bound of the agreement interval is also assessed by checking how many times the true quantile $q^D_{0.975}$ of the normal distribution $N(0, \sigma^2_D = \sigma^2_\tau + \sigma^2_\nu)$ lies into the confidence interval of the upper bound of the agreement interval (since the confidence level is assessed on the upper bound, it is unnecessary to check the coverage probabilities of the lower bound). The confidence levels of the $\beta\gamma$ tolerance interval can be computed in two different but equivalent ways: by creating a binary variable during the simulation to indicate whether the proportion covered by the $\beta\gamma$ tolerance interval is higher or not the predictive level or by computing the quantile q_ψ of the computed proportions covered

by the $\beta\gamma$ tolerance interval. We chose the first solution.

Figure 4.1 shows the coverage probabilities of the different intervals. We can notice that the agreement interval is too narrow, as expected, with coverage probabilities too low for small sample sizes (Figure 4.1-top) while its coverage probabilities move closer to the predictive level when N increases.
The coverage probabilities of the β-TI are excellent whatever N (the green curve is nearly superimposed to the horizontal line representing the predictive level).
The coverage probabilities of the XL-AI are all higher than the predictive level, except for small sample sizes with small confidence level, i.e. the lowest red-dashed curve (the lowest curve is given by the lowest confidence level (50%) and the highest curve by the highest confidence level (90%)). Moreover, the coverage probabilities of the XL-AI are, obviously, all higher than those given by the AI. Indeed, the XL-AI is always larger than the AI since it is expanded from the AI (by increasing the confidence level). We can notice that the coverage probabilities of the CI on each bound of the agreement interval are close to their nominal levels (Figure 4.1-middle), except for small sample sizes.
The coverage probabilities of the $\beta\gamma$-TI with 50% confidence level (the lowest blue curve on Figure 4.1-top) are lower than the predictive level but move closer to it when N increases. But, its coverage probabilities are close to the nominal level 50% (4.1-middle). This result is interesting:
it means that the χ^2 approximation in formula (4.7) works perfectly well as in 50% of cases (half of the simulations) the true proportion recovered by the $\beta\gamma$-TI is higher than 95% (the predictive level defined by z in formula (4.7)). But, on average, this proportion is lower than 95% (and moves closer to 95% when N increases). This is because the distribution of a proportion does not follow a symmetric distribution (especially for small sample sizes and extreme proportion as 95%). Note also that the $\beta\gamma$-TI with 50% confidence level is analog, but mathematically not identical, to the β-TI with the same predictive levels (the median of the simulated proportion is not equal to their average).
We can notice on Figure 4.1-top that the coverage probabilities of the $\beta\gamma$-TI increase and move away from the nominal level 95% when the confidence level increases. This is obvious as by increasing the confidence level $1-\gamma$, the $\beta\gamma$-TI must expand to provide intervals which contain at least 95% of the data in more and more cases.
Even if the confidence level of the $\beta\gamma$-TI and the confidence level of the CI computed on each bound of the agreement interval are two different concepts, we can notice on Figure 4.1-middle that for small sample sizes, the confidence level of the $\beta\gamma$-TI is slightly lower than the nominal level.
Finally, it can be noticed that all the coverage probabilities move closer to each other when N increases. Thus, all the intervals (whatever the confidence level for the CI on the AI or the $\beta\gamma$-TI) move closer to each other and to the true

quantile of the differences $q^D_{0.975}$ as we can see on Figure 4.1-bottom with the upper limits of the intervals.

4.2 Practical equivalence with replicates

In Section 4.1, the practical equivalence was assessed by comparing directly the differences between the two measurement methods as there were no replicates. In this section, we consider that replicates are available with $n_{X_i} = n_X$ and $n_{Y_i} = n_Y \ \forall i$ but the goal is to find an interval where, for instance, 95% of the future **single** differences are expected to lie.

4.2.1 Horizontal agreement interval by Bland and Altman

When replicates are available, the X_{ij} and Y_{ik} values can be averaged to X_i and Y_i as explained in Chapter 1. The differences D_i and averages M_i can still be computed with formulas (4.1). Therefore, the mean of the differences and their variance can also still be computed by formulas (4.2). However, the agreement interval cannot be anymore computed by formula (4.3). Indeed, the goal of the agreement interval is to find an interval where $100(1-\psi)\%$ of the **singles** future differences between the two measurement methods will lie (and not an interval where $100(1-\psi)\%$ of the **averages** future differences will lie).

With unreplicated data, the differences D_i are single differences while the differences D_i are averages differences when the data are replicated. So, with replicated data, the variance of the differences in Formula (4.2) must be modified in order to estimate the variance of singles differences instead of averages differences. The variance of the single differences, $\sigma^2_{Ds} = \sigma^2_\tau + \sigma^2_\nu$, was first estimated by Bland and Altman [12] (wrongly) by NS^2_D and they propose later a better estimator with the following formula [11] (with our notation):

$$\hat{\sigma}^2_{Ds} = S^2_D + \left(1 - \frac{1}{n_X}\right) S^2_\tau + \left(1 - \frac{1}{n_Y}\right) S^2_\nu \tag{4.8}$$

Note that this formula is a generalization of the variance estimator in (4.2) as $S^2_D = S^2_{Ds}$ when $n_X = n_Y = 1$ (unreplicated data). The variance of this estimator is given by [11]:

$$\text{Var}\left(\hat{\sigma}^2_{Ds}\right) = \frac{2\sigma^4_D}{N-1} + \frac{2(n_X-1)\sigma^4_\tau}{n^2_X N} + \frac{2(n_Y-1)\sigma^4_\nu}{n^2_Y N} \tag{4.9}$$

The agreement interval is then given by the following formula [11]:

$$AI_{replicated}: \ \overline{D} \pm z_{1-\psi/2}\hat{\sigma}_{Ds} \tag{4.10}$$

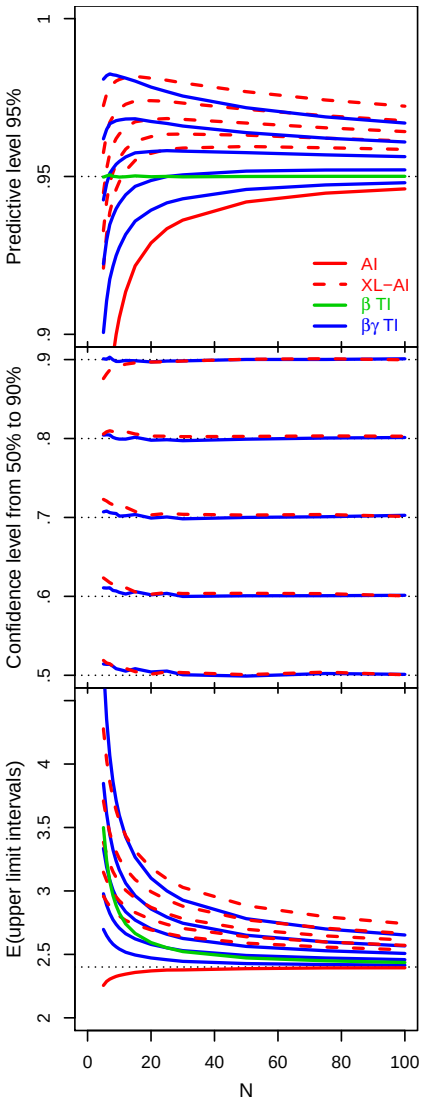

Figure 4.1: *Coverage probabilities of the agreement interval with or without its confidence interval and the tolerance intervals. Predictive level = 95% (top) and confidence level from 50 to 90% (middle), expectation of the upper limits of the intervals (bottom) with respect to N*

The variance of this agreement interval is given by the approximate following formula [11]:

$$Var(AI_{replicated}) \hat{=} \frac{\hat{\sigma}^2_{Ds}}{N} + \frac{z^2_{1-\psi/2}}{2\hat{\sigma}^2_{Ds}} \left(\frac{S^4_D}{N-1} + \frac{(n_X-1)S^4_\tau}{n^2_X N} + \frac{(n_Y-1)S^4_\nu}{n^2_Y N} \right) \tag{4.11}$$

This formula is a generalization of formula (4.4) (both formulas are identical with unreplicated data).
The $100(1-\gamma)\%$ confidence intervals on each limit of this agreement interval is then computed as follows:

$$CI(AI_{replicated}) : \ (AI_{replicated}) \pm z_{1-\gamma/2}\sqrt{Var(AI_{replicated})} \tag{4.12}$$

Note that, curiously, Bland and Altman use a normal distribution to compute these confidence intervals while they use the t-distribution with unreplicated data. Actually, it will be shown with simulations in a further section that the coverage probabilities of formula 4.12 are too high whatever N. Therefore, if a t-distribution was applied, these coverage probabilities would be worst (as the CI would be larger).

4.2.2 Horizontal tolerance interval

Variance of singles differences estimators

We did not find in the literature tolerance intervals with replicates in order to get an interval for singles measures. We propose in this section to modify formulae (4.6) and (4.7) to take into account the variance of singles differences instead of the variance of mean differences.
While Bland and Altman propose to estimate σ^2_{Ds} by $\hat{\sigma}^2_{Ds}$ (formula (4.8)), we propose to estimate this variance simply and more easily by the following estimator:

$$S^2_{Ds} = S^2_\tau + S^2_\nu \tag{4.13}$$

Under normal and independent measurement errors, we have:

$$\mathrm{Var}\left(S^2_{Ds}\right) = \frac{2\sigma^4_\tau}{N(n_X-1)} + \frac{2\sigma^4_\nu}{N(n_Y-1)} \tag{4.14}$$

To simplify the notation, we can also define the following variance:

$$\sigma^2_\kappa = \sigma^2_\tau/n_X + \sigma^2_\nu/n_Y \tag{4.15}$$

Without proportional bias, it is easy to check that $\sigma^2_D = \sigma^2_\kappa$ and so the variance of the differences, σ^2_D, can then be estimated by S^2_D or by the following estimator:

$$S^2_\kappa = S^2_\tau/n_X + S^2_\nu/n_Y \tag{4.16}$$

Both estimators, S_D^2 and S_κ^2 are unbiased.

The properties of estimator $\hat{\sigma}_{Ds}^2$ (4.8) are not discussed by Bland and Altman but note that both estimators, $\hat{\sigma}_{Ds}^2$ and S_{Ds}^2, are unbiased. Indeed,
$E\left(\hat{\sigma}_{Ds}^2\right)$
$= E\left(S_D^2\right) + \left(1 - \frac{1}{n_X}\right) E\left(S_\tau^2\right) + \left(1 - \frac{1}{n_Y}\right) E\left(S_\nu^2\right)$
$= \sigma_D^2 + \left(1 - \frac{1}{n_X}\right) \sigma_\tau^2 + \left(1 - \frac{1}{n_Y}\right) \sigma_\nu^2$
$= \frac{\sigma_\nu^2}{n_Y} + \frac{\sigma_\tau^2}{n_X} + \left(1 - \frac{1}{n_X}\right) \sigma_\tau^2 + \left(1 - \frac{1}{n_Y}\right) \sigma_\nu^2$
$= \sigma_\tau^2 + \sigma_\nu^2$

while it is obvious that: $E\left(S_{Ds}^2\right) = E\left(S_\tau^2\right) + E\left(S_\nu^2\right) = \sigma_\tau^2 + \sigma_\nu^2$.

The variance of estimator $\hat{\sigma}_{Ds}^2$ (4.8), is lower than the variance of estimator S_{Ds}^2 (4.13), except for high number of replicates as shown in Figure 4.2 where $\mathrm{Var}(S_{Ds}^2)/\mathrm{Var}(\hat{\sigma}_{Ds}^2)$ is displayed with respect to $n_X = n_Y$ for $N = 10$ or $N = 20$.
Indeed, we can notice that for $N = 10$, the variance of S_{Ds}^2 is lower than the variance of $\hat{\sigma}_{Ds}^2$ for $n_X = n_Y \geq 6$ as the curve is lower than 1. For $N = 20$, the variance of estimator (4.13) is not lower than the variance of estimator (4.8) until $n_X = n_Y = 10$ but they move closer each other when the number of replicates increases (for $N = 20$ and $n_X = n_Y = 10$, the variances of the estimators are nearly equal such that the ratio between these variances ≈ 1).

Moreover, our estimator S_{Ds}^2 is completely robust to points outlying the distributions of the differences D_i. While, the estimator $\hat{\sigma}_{Ds}^2$ is not robust as it is computed from the variance of the differences (S_D^2).

β-expectation tolerance interval

In order to develop a β-expectation tolerance interval (β-TI), let's take a look at the distributions of $\overline{D}$ and the future single difference D_{N+1} (without proportional bias). The distribution of $\overline{D}$ is given by the following expression:

$$\overline{D} \sim N\left(\mu_D, \frac{\sigma_D^2}{N} = \frac{\sigma_\kappa^2}{N}\right)$$

while the distribution of D_{N+1}, is given by the following expression:

$$D_{N+1} \sim N\left(\mu_D, \sigma_{Ds}^2\right)$$

Then, it follows that:

$$D_{N+1} - \overline{D} \sim N\left(0, \sigma_\omega^2 = \sigma_{Ds}^2 + \frac{\sigma_D^2}{N} = \sigma_{Ds}^2 + \frac{\sigma_\kappa^2}{N}\right)$$

where σ_ω^2 can be estimated by S_ω^2 or $\hat{\sigma}_\omega^2$ as follows:

$$S_\omega^2 = S_{Ds}^2 + S_\kappa^2/N \tag{4.17}$$

$$\hat{\sigma}_\omega^2 = S_{Ds}^2 + S_D^2/N \tag{4.18}$$

Note that:

$$S_\omega^2 = S_\tau^2 + S_\nu^2 + \frac{S_\tau^2}{n_X N} + \frac{S_\nu^2}{n_Y N} = n'_X S_\tau^2 + n'_Y S_\nu^2$$

where

$$n'_X = \left(1 + \frac{1}{n_X N}\right) \text{ and } n'_Y = \left(1 + \frac{1}{n_Y N}\right)$$

and

$$\hat{\sigma}_\omega^2 = S_\tau^2 + S_\nu^2 + S_D^2/N$$

The distributions and the degrees of freedom of both estimators, S_ω^2 and $\hat{\sigma}_\omega^2$ can be computed by the well-known Welch-Satterthwaite equation to get an approximation of the effective degrees of freedom of a linear combination of independent sample variances. For n sample variances S_j^2 $(j = 1, ..., n)$, each respectively having r_j degrees of freedom, the linear combination $S_k = \sum_{j=1}^p k_j S_j^2$ follows approximately a Chi square distribution where the degrees of freedom, r, is given by the following formula:

$$r \approx \frac{\left(\sum_{j=1}^p k_j S_j^2\right)^2}{\sum_{j=1}^p \frac{\left(k_j S_j^2\right)^2}{r_j}}$$

Then, both estimators, S_ω^2 and $\hat{\sigma}_\omega^2$ follow Chi square distributions with, respectively, r_S and r_σ degrees of freedom given by the following approximate formulae:

$$r_S \approx \frac{\left(n'_X S_\tau^2 + n'_Y S_\nu^2\right)^2}{\frac{\left(n'_X S_\tau^2\right)^2}{(n_X - 1)N} + \frac{\left(n'_Y S_\nu^2\right)^2}{(n_Y - 1)N}} = \frac{S_\omega^4}{\frac{\left(n'_X S_\tau^2\right)^2}{N_X} + \frac{\left(n'_Y S_\nu^2\right)^2}{N_Y}} \tag{4.19}$$

$$r_\sigma \approx \frac{\left(S_\tau^2 + S_\nu^2 + S_D^2/N\right)^2}{\frac{S_\tau^4}{(n_X - 1)N} + \frac{S_\nu^4}{(n_Y - 1)N} + \frac{S_D^4/N^2}{N-1}} = \frac{\hat{\sigma}_\omega^4}{\frac{S_\tau^4}{N_X} + \frac{S_\nu^4}{N_Y} + \frac{S_D^4/N^2}{N-1}} \tag{4.20}$$

where

$$N_X = (n_X - 1)\, N \text{ and } N_Y = (n_Y - 1)\, N$$

It follows that:

$$\frac{D_{N+1} - \overline{D}}{S_\omega} \sim \frac{N\left(0, \sigma_\omega^2\right)}{\sqrt{\frac{\chi_{r_S}^2 \sigma_\omega^2}{r_S}}} = \frac{N\left(0, 1\right)}{\sqrt{\frac{\chi_{r_S}^2}{r_S}}} = t_{r_S}$$

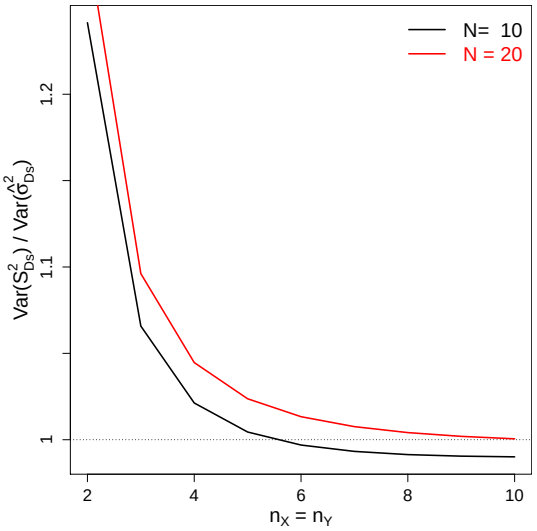

Figure 4.2: *Difference between the variances of the estimators* (4.8) *and* (4.13) *with respect to* $n_X = n_Y$ *for* $N = 10$ *and* $N = 20$

and

$$\frac{D_{N+1} - \overline{D}}{\hat{\sigma}_\omega} \sim \frac{N\left(0, \sigma_\omega^2\right)}{\sqrt{\frac{\chi^2_{r_\sigma}\sigma^2_\omega}{r_\sigma}}} = \frac{N\left(0,1\right)}{\sqrt{\frac{\chi^2_{r_\sigma}}{r_\sigma}}} = t_{r_\sigma}$$

Therefore, we propose to compute the $100(1-\psi)\%$ β-expectation tolerance interval (β-TI) by one of the two following formulae:

$$\overline{D} \pm t_{1-\psi/2,r_S} S_\omega \tag{4.21}$$

or

$$\overline{D} \pm t_{1-\psi/2,r_\sigma} \hat{\sigma}_\omega \tag{4.22}$$

$\beta\gamma$-content tolerance interval

Based on the paper of Graybill [89], a $\beta\gamma$-content tolerance interval ($\beta\gamma$-TI) where the variance, S_k^2, is a linear combination of other variances S_j^2 ($j = 1, ..., p$), can be computed by the following formula:

$$\overline{D} \pm z_{1-\psi/2} S_k \sqrt{1 + \frac{1}{N_e}} \sqrt{1 + \frac{1}{S_k^2}\sqrt{\sum_{j=1}^{p} k_j H_j^2 S_j^4}}$$

where

$$H_j = \frac{1}{F_{\gamma,r_j,\infty}} - 1$$

and N_e is the effective sample size.
As $r_j F_{\gamma,r_j,\infty} = \chi^2_{\gamma,r_j}$ and based on the theoretical development described for the β-TI, we propose to compute the $100(1-\psi)\%$ $\beta\gamma$-TI by one of the two following formulae:

$$\overline{D} \pm z_{1-\psi/2} S_\omega \sqrt{1 + \frac{1}{S^2_{Ds}} \sqrt{n'_X H^2_\tau S^4_\tau + n'_Y H^2_\nu S^4_\nu}} \tag{4.23}$$

or

$$\overline{D} \pm z_{1-\psi/2} \hat{\sigma}_\omega \sqrt{1 + \frac{1}{S^2_{Ds}} \sqrt{H^2_\tau S^4_\tau + H^2_\nu S^4_\nu + H^2_D S^4_D / N}} \tag{4.24}$$

where

$$H_\tau = \frac{N_X}{\chi_{\gamma,N_X}} - 1,\ H_\nu = \frac{N_Y}{\chi_{\gamma,N_Y}} - 1 \text{ and } H_D = \frac{N-1}{\chi_{\gamma,N-1}} - 1$$

4.2.3 Comparison and coverage probabilities of the agreement and tolerance intervals

In order to assess coverage probabilities of the agreement and tolerance intervals, we run 10000 simulations as described in Section 4.1.3 with $\sigma^2_\tau = 1.25 n_X$ and $\sigma^2_\nu = 0.375 n_Y$ unknown and $n_X = n_Y = 2$ or $n_X = n_Y = 4$ or $n_X = 4$, $n_Y = 2$. For each simulated data set, 18 intervals are, then, computed as explained in Section 4.1.3 where additionally the TI are computed with S_ω or $\hat{\sigma}_\omega$. The coverage probabilities of each interval under the normal distribution with mean equal to 0 and variance equal to $\sigma^2_\tau + \sigma^2_\nu = 1.625$ are then computed.

Figure 4.3 shows the coverage probabilities of the different agreement intervals with two replicates per sample and measurement method, where it can be noticed that:

- the agreement interval is too narrow, as expected, with coverage probabilities too low for small sample sizes (Figure 4.3-top), while its coverage probabilities move closer to the predictive level when N increases.
- the coverage probabilities of the XL-AI are all higher than the predictive level, even for small sample sizes (they are higher than those obtained with unreplicated data). As already mentioned in Section 4.1.3, the coverage probabilities of the XL-AI are, obviously, all higher than those given by the AI. But, the coverage probabilities of the CI on each bound of the agreement interval are, here, not anymore close to their nominal levels (Figure 4.3-bottom), as they are all higher than their nominal level.

Figure 4.4 and 4.5 are analogous with, respectively, $n_X = n_Y = 4$ or $n_X = 4$ and $n_Y = 2$. When the variance σ^2_{Ds} is estimated more precisely by increasing N or the number of replicates:

- the coverage probabilities of the agreement interval move closer to the predictive level
- the coverage probabilities of the CI on the bounds of the agreement interval move away to their nominal level

To sum-up, the coverage probabilities of the agreement interval is improved by increasing N or/and the number of replicates but unfortunalety, by increasing n_X or n_Y, the coverage probabilities of the CI of the agreement intervals are too wide.

Figures 4.6, 4.7 and 4.8 display the coverage probabilities of the tolerance intervals for, respectively, $n_X = n_Y = 2$, $n_X = n_Y = 4$ and $n_X = 4$, $n_Y = 2$:

- the β-TI provide excellent coverage probabilities whatever N and n_X or n_Y and the two green curves for S_ω or $\hat{\sigma}_\omega$ are superimposed on the charts (the coverage probabilities are different but very similar)
- the coverage probabilities of the $\beta\gamma$-TI with 50% confidence level (the lowest blue curve on Figures 4.6-top, 4.7-top and 4.8-top) are lower than the predictive level but move closer to it when N increases
- the coverage probabilities of the $\beta\gamma$-TI increase and move away from the nominal level 95% when the confidence level increases as already mentioned in Section 4.1.3
- the coverage probabilities of the confidence level of the $\beta\gamma$-TI are excellent except for small sample sizes where the $\beta\gamma$-TI obtained with S_ω are better than those obtained with $\hat{\sigma}_\omega$ (Figures 4.6-bottom, 4.7-bottom and 4.8-bottom).

4.3 How to regress in a Bland and Altman plot?

In Sections 4.1 and 4.2, the different intervals are horizontal lines around the mean difference but these intervals are restrictive as they assume that there is no proportional bias between the two measurement methods. In order to get better intervals, a regression line must be estimated and the intervals computed around the regression line. This section compares different regression techniques to directly estimate a regression line in a Bland and Altman coordinate system (M, D) under homoscedasticity. The bias and coverage probabilities of some errors-in-variables regressions presented in Chapter 2 will be analyzed in a (M, D) space. Additionally, we will show how to regress adequately in a (M, D) space with the Mandel procedure and we will modify the formulae of the BLS regression in order to get a regression line consistent in a (M, D) space. These modifications are, according to our knowledge, novel.

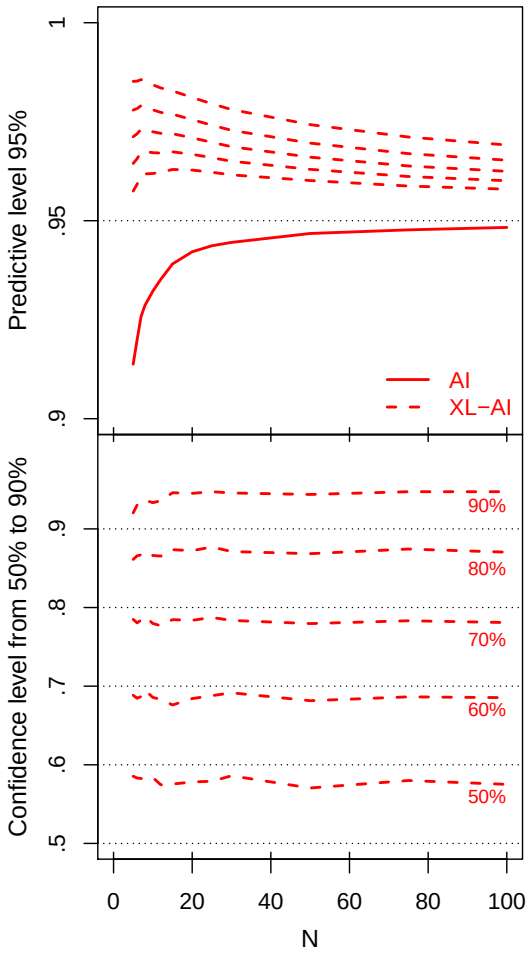

Figure 4.3: *Coverage probabilities of the agreement interval with or without its confidence interval. Predictive level* $= 95\%$ *(top) and confidence level from 50 to 90% (bottom),* $n_X = n_Y = 2$ *with respect to* N

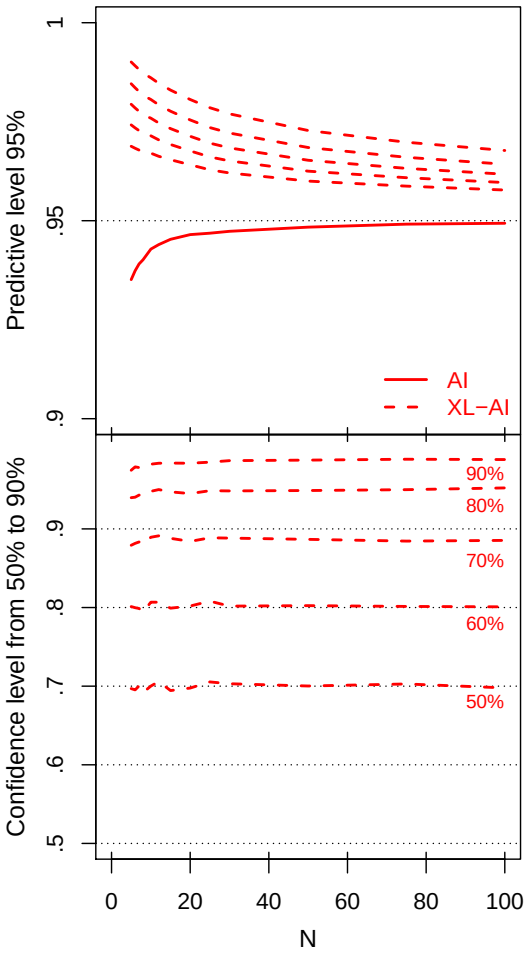

Figure 4.4: *Coverage probabilities of the agreement interval with or without its confidence interval. Predictive level* = 95% *(top) and confidence level from 50 to 90% (bottom),* $n_X = n_Y = 4$ *with respect to* N

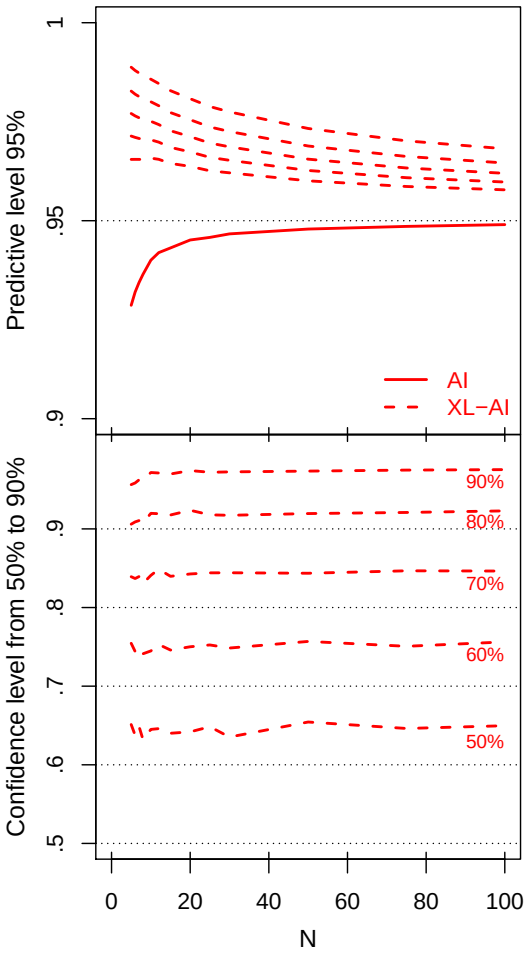

Figure 4.5: *Coverage probabilities of the agreement interval with or without its confidence interval. Predictive level* $= 95\%$ *(top) and confidence level from 50 to 90% (bottom),* $n_X = 4$, $n_Y = 2$ *with respect to* N

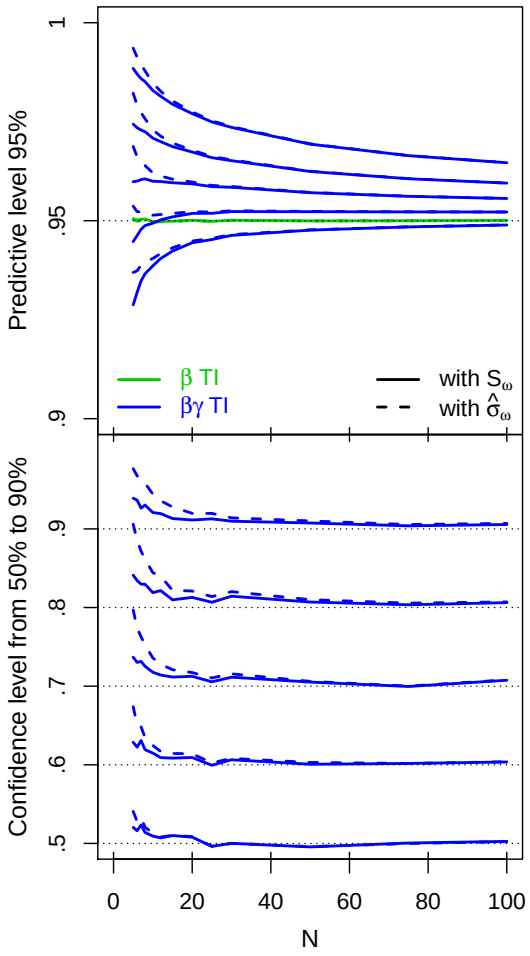

Figure 4.6: *Coverage probabilities of the tolerance intervals. Predictive level* $= 95\%$ *(top) and confidence level from 50 to 90% (bottom),* $n_X = n_Y = 2$ *with respect to* N

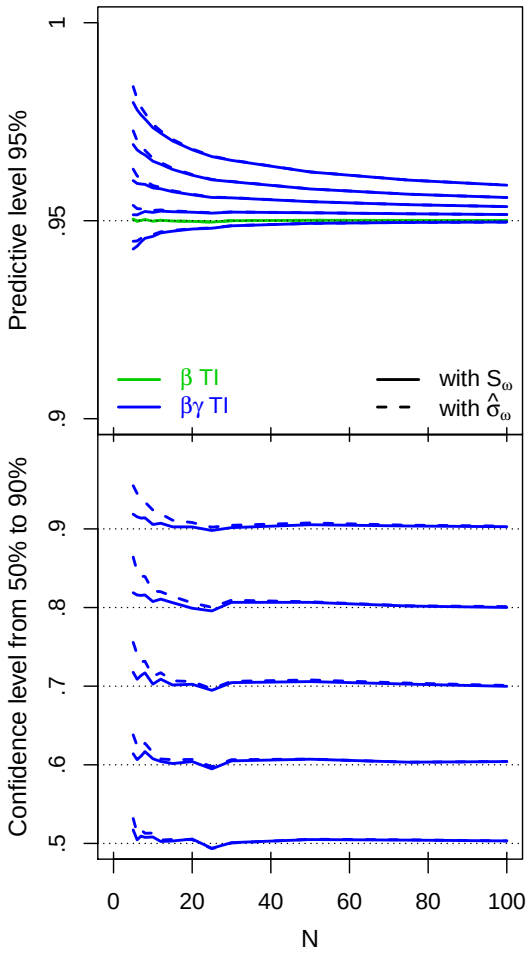

Figure 4.7: *Coverage probabilities of the tolerance intervals. Predictive level* $= 95\%$ *(top) and confidence level from 50 to 90% (bottom),* $n_X = n_Y = 4$ *with respect to* N

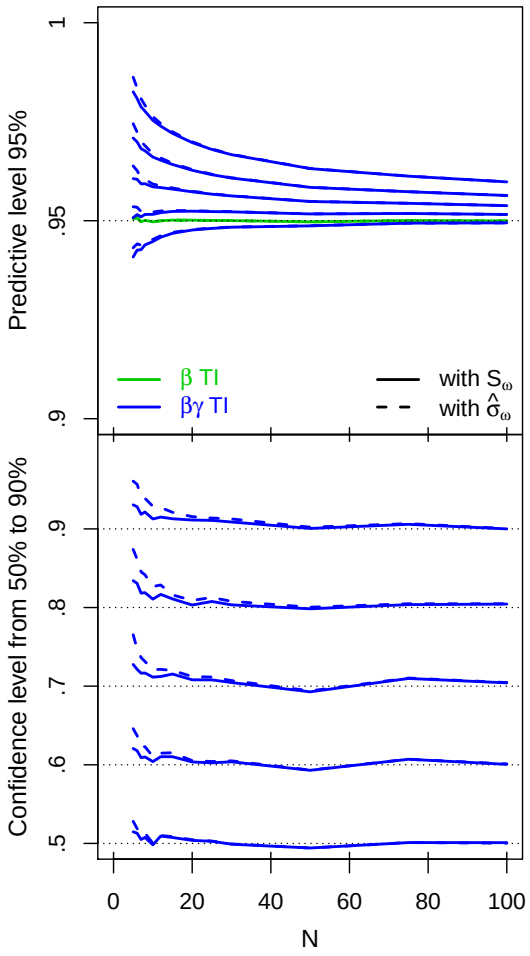

Figure 4.8: *Coverage probabilities of the tolerance intervals. Predictive level* $= 95\%$ *(top) and confidence level from 50 to 90% (bottom),* $n_X = 4$, $n_Y = 2$ *with respect to* N

4.3.1 Bland and Altman model versus (X, Y) model

Remind that the model in a (X, Y) plot is given by the following formulae:

$$X_{ij} = \xi_i + \tau_{ij},\ Y_{ik} = \eta_i + \nu_{ik} \tag{4.25}$$

or with the means of the repeated measures:

$$X_i = \xi_i + \tau_i,\ Y_i = \eta_i + \nu_i \tag{4.26}$$

The literature about Bland and Altman approach does not present any clear model for the data represented in a (M, D) space. We propose here a model which takes into account the specificities of the (M, D) coordinate system, according to the formulae and notations given in Chapter 1. The formulae are, according to our knowledge, novel.
In a Bland and Altman plot, we have:

$$D_i = Y_i - X_i = \eta_i - \xi_i + \nu_i - \tau_i,\ M_i = \frac{X_i + Y_i}{2} = \frac{\xi_i + \eta_i}{2} + \frac{\tau_i + \nu_i}{2} \tag{4.27}$$

In further sections, the following notations will be used:

$$\overline{D} = \frac{1}{N}\sum_{i=1}^{N} D_i \text{ and } \overline{M} = \frac{1}{N}\sum_{i=1}^{N} M_i,$$

$$S_{mm} = \sum_{i=1}^{N}(M_i - \overline{M})^2, S_{dd} = \sum_{i=1}^{N}(D_i - \overline{D})^2 \text{ and } S_{md} = \sum_{i=1}^{N}(M_i - \overline{M})(D_i - \overline{D}).$$

In a (X, Y) plot, the ratio between the two error variances in both axes was defined as λ_{XY} (2.8), while in a (M, D) plot, this ratio becomes:

$$\lambda_{BA} = \frac{\sigma_\tau^2/n_X + \sigma_\nu^2/n_Y}{\left(\sigma_\tau^2/n_X + \sigma_\nu^2/n_Y\right)/4} = \frac{\sigma_\kappa^2}{\sigma_\kappa^2/4} = 4 \tag{4.28}$$

This result can be surprising as this ratio is an important parameter in a (X, Y) plot, sometimes unknown, while it is simply equal to 4 in a (M, D) space. But, this result is actually, obvious, since it is well known that $\mathrm{Var}(X \pm Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)$ and $\mathrm{Var}(aX) = a^2\mathrm{Var}(X)$.

So, in a (X, Y) plot, the ratio of the two error variances, λ_{XY}, can vary from 0 to ∞ and it is known when the variances σ_τ^2 and σ_ν^2 are known, otherwise it is unknown. In a (M, D) plot, the ratio between the two error variances on both axes, λ_{BA} is always known and equal to 4.

Unfortunately, while the measurement errors on both axes are uncorrelated in a (X, Y) plot (the measurement errors are independent (1.3): $\rho_{XY} = 0$), the measurement errors on both axes in a (M, D) plot can be correlated. Indeed, it is easy to check that:

$$\rho_{BA}\left(\frac{\tau_i + \nu_i}{2}, \nu_i - \tau_i\right) = \frac{\lambda_{XY} - 1}{\lambda_{XY} + 1} \tag{4.29}$$

where ρ_{BA} is the correlation between the measurement errors on both axes for a given sample in a (M, D) space. Figure 4.9 shows the correlation ρ_{BA} with respect to λ_{XY}. We can notice in Figure 4.9 or in formula (4.29) that the measurement errors in a (M, D) plot are uncorrelated when $\lambda_{XY} = 1$ (i.e. $\sigma_\tau^2/n_X = \sigma_\nu^2/n_Y$). Otherwise, these measurement errors are correlated, as defined in (4.29), positively for $\lambda_{XY} > 1$ and negatively for $\lambda_{XY} < 1$ (the opposite would have been observed if we had defined the differences D_i inversely like $D_i = X_i - Y_i$). Note that the correlation ρ_{BA}, is not insignificant in practice as, for instance, $\rho_{BA} = 0.5$ for $\lambda_{XY} = 3$ (a measurement method three times more precise than the other, with $n_X = n_Y$).
Figure 4.10 shows three examples of measurement errors where the (true) 95% ellipses are displayed in a (X, Y) plot on the left and in a (M, D) plot on the right. We can notice that whatever the value of λ_{XY}, the axes of the ellipses are parallel to the X and Y-axes in a (X, Y) plot as $\rho_{XY} = 0$. In a (M, D) plot, the ratio of the lengths between the axes of the ellipses is constant as $\lambda_{BA} = 4$ but the ellipses can be tilted on the right or on the left as $\rho_{BA} \neq 0$ for $\lambda_{XY} \neq 1$.

While the true but unobservable measures, ξ_i and η_i, are linked by the linear regression $\eta_i = \alpha^{XY} + \beta^{XY}\xi_i$ in a (X, Y) plot, we have in a (M, D) plot:

$$\eta_i - \xi_i = \alpha^{BA} + \beta^{BA}\left(\xi_i + \eta_i\right)/2 \tag{4.30}$$

and while the regression is applied in a (X, Y) plot with the model (under homoscedasticity) $Y_i = \alpha^{XY} + \beta^{XY}X_i + \epsilon_i^{XY}$, the model in a (M, D) plot becomes:

$$D_i = \alpha^{BA} + \beta^{BA}M_i + \epsilon_i^{BA} \tag{4.31}$$

where α^{BA}, the intercept and β^{BA}, the slope are estimated respectively by $\hat{\alpha}^{BA}$ and $\hat{\beta}^{BA}$.
The two regression models are actually equivalent and the relationships between the parameters are given by the following formulae:

$$\alpha^{BA} = \frac{2\alpha^{XY}}{1 + \beta^{XY}} \text{ and } \beta^{BA} = \frac{2\beta^{XY} - 2}{1 + \beta^{XY}} \tag{4.32}$$

or inversely:

$$\alpha^{XY} = \frac{2\alpha^{BA}}{2 - \beta^{BA}} \text{ and } \beta^{XY} = \frac{\beta^{BA} + 2}{2 - \beta^{BA}} \tag{4.33}$$

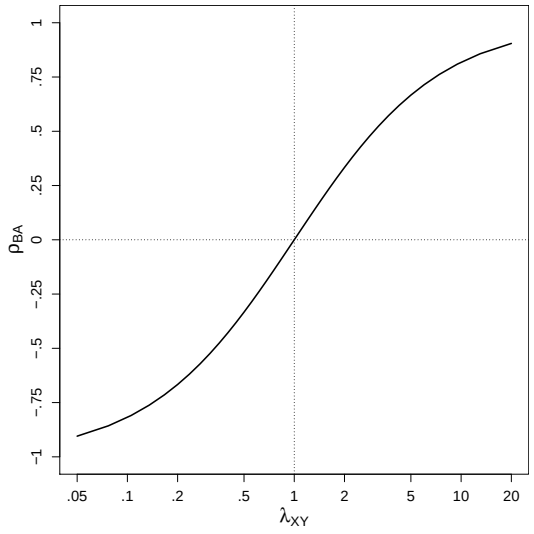

Figure 4.9: *The correlation between the measurement errors in a* (M, D) *plot,* ρ_{BA}*, with respect to* λ_{XY}

So, the equivalence test based on the hypotheses $H_0^\alpha : \alpha^{XY} = 0$ and $H_0^\beta : \beta^{XY} = 1$ in a (X, Y) plot becomes $H_0^\alpha : \alpha^{BA} = 0$ and $H_0^\beta : \beta^{BA} = 0$ in a (M, D) plot.
The comparison of both models, (X, Y) plot or (M, D) plot is summarized in Table 4.1.

4.3.2 Classical regressions

Bland and Altman propose simply to estimate a regression line in a (M, D) plot by the OLSv technique [11]. Ludbrook [48,82,90,91] proposes also the OLSv technique (or the MR in a (X, Y) plot or a weighted MR). Carstensen [92] proposes also an OLSv approach by converting the estimated line in a (X, Y) to (M, D) plot and vice-versa. We did not find in the literature papers dealing with other regressions than OLSv in a (M, D) plot (except the weighted OLSv technique).

We explained in Chapter 2 that DR and BLS regressions are consistent in order to estimate model (1.11). The easiest solution to estimate a regression line in a (M, D) space is then probably to convert the DR and BLS estimates computed in the (X, Y) plot, $\hat{\alpha}_{DR}^{XY} = \hat{\alpha}_{BLS}^{XY}$ and $\hat{\beta}_{DR}^{XY} = \hat{\beta}_{BLS}^{XY}$, to the corresponding estimates in the (M, D) plot by using the conversion formulae (4.32).

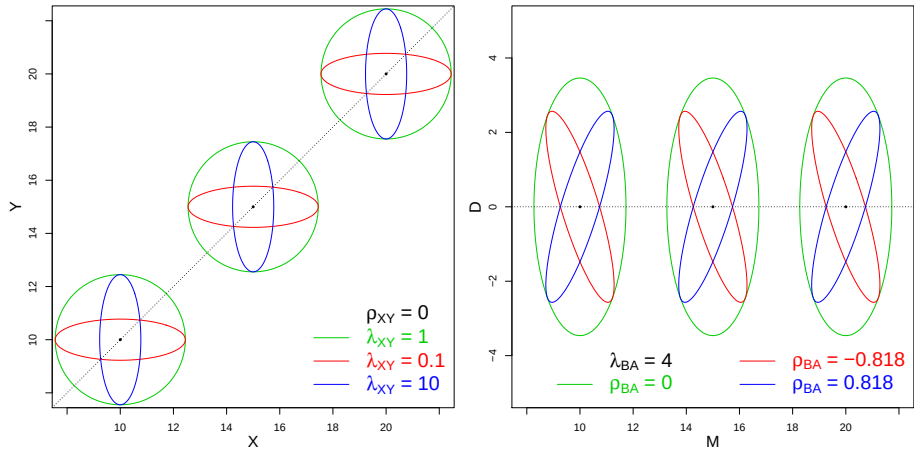

Figure 4.10: *Three examples of measurement errors in a* (X, Y) *plot (left) and the corresponding measurement errors in a* (M, D) *plot (right)*

	(X, Y)	(M, D)
Errors	Both axes	Both axes
Ratio Variances Errors	$\lambda_{XY} \in [0, \infty]$ known or unknown	$\lambda_{BA} = 4$
Correlation between errors	$\rho_{XY} = 0$	$\rho_{BA} \in [-1, 1]$ known or unknown
Proportional bias	$H_0^{\beta^{XY}} : \beta^{XY} = 1$	$H_0^{\beta^{BA}} : \beta^{BA} = 0$
Constant bias	$H_0^{\alpha^{XY}} : \alpha^{XY} = 0$	$H_0^{\alpha^{BA}} : \alpha^{BA} = 0$
Bias	$H_0 : \theta^{XY} = (0, 1)'$	$H_0 : \theta^{BA} = (0, 0)'$

Table 4.1: *Comparison of* (X, Y) *and* (M, D) *models*

However, the existing Bland and Altman literature does not use this approach. This section proposes to compare different regression techniques to estimate the regression line directly in the (M, D) space.

Figure 4.11-left shows the different regression techniques given in Chapter 2 in a (X, Y) plot and in a (M, D) plot in Figure 4.11-right, under equivalence. Actually, the OLSh is the worst regression in a (M, D) plot. Indeed, without proportional bias, the slope is equal to zero: $\beta^{BA} = 0$ and the OLSh slope is then equal to infinity: $\beta^{BA}_{OLSh} = \infty$ as it minimizes the horizontal distances between the points to the line which is horizontal in that case (the distances between the points and the line are, then, equal to infinity). Without proportional bias, the expectation of OLSh slope estimator, is then undefined (and follow a Cauchy distribution). Note that without proportional bias, the variables D and M are independent and $\beta^{BA} = 0$. The OLSh will not be considered anymore in order to estimate a regression line in a (M, D) plot.
As explained in Chapter 2, the slope given by MR is the geometric mean of the two OLS slopes. As the OLSh is the worst regression in a (M, D) plot, it can be expected that the MR is not suitable to estimate a regression line in a (M, D) plot. Moreover, we can notice in Figure 4.11-right that the areas of the MR rectangles are equal to infinity without proportional bias. The MR will not be considered anymore in order to estimate a regression line in a (M, D) plot.
Without proportional bias, the OR and OLSv should provide similar parameters estimates as the orthogonal distances from the points to the line are vertical.
The DR and BLS regressions can still be considered, replacing λ_{XY} by $\lambda_{BA} = 4$ and note that the oblique direction given by $-\lambda_{BA}/\hat{\beta}$ is, then, here equal to infinity and therefore, DR (and BLS) should also provide similar parameters estimates than OR and OLSv.

4.3.3 Correlated-errors-in-variables regressions

As explained in Section 4.3.1, the errors in both axes in a (M, D) are correlated (for $\lambda_{XY} \neq 1$). This correlation, ρ_{BA} must be taken into account in order to estimate a regression line in a (M, D) plot. Indeed, ignoring the correlation in the errors (in a given point) in a regression can affect the regression estimates and predictions as shown in Schaalje and Butts (in the framework of the Deming regression in a (X, Y) plot where they add a correlation) [93].

In order to estimate a regression line in a (M, D) plot, we propose two correlated-errors-in-variables regression (CEIV) based on two existing regressions modified to get a suitable and consistent estimated regression line in a (M, D) coordinate system by taking into account its assumptions. These formulae are, according

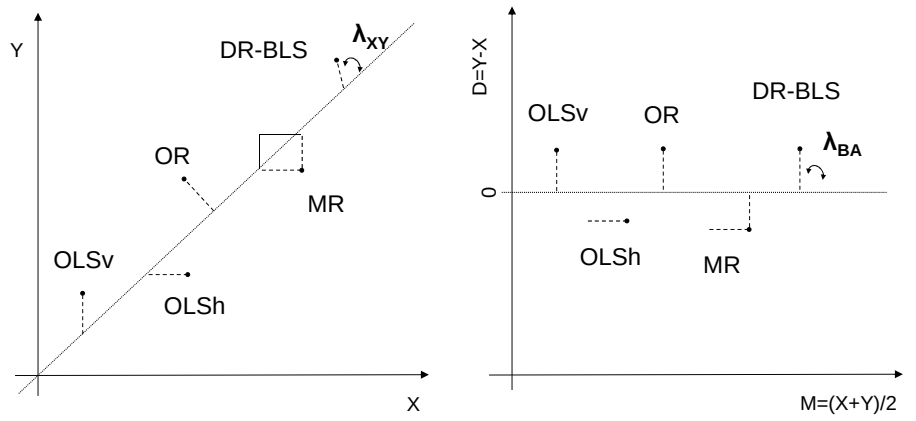

Figure 4.11: *The different regressions in a* (X, Y) *plot (left) or* (M, D) *plot (right)*

to our knowledge, novel.

CEIV regression based on Mandel procedure

In order to estimate a regression line in a (M, D) space, we propose a CEIV regression based on the Mandel procedure [56] described in Chapters 2 and 3. Remind that the Mandel procedure was proposed in the context of inter-laboratories studies and can take into account the errors and their correlation on both axes. This procedure consists in transforming the (X_i, Y_i) data into (U_i, V_i) data, or in the Bland and Altman context: (M_i, D_i) data into $(U_i = M_i + kD_i, V_i = D_i - \hat{\beta}_{Mandel} M_i)$ data.
The parameters are given by the following formulae:

$$\hat{\beta}^{BA}_{Mandel} = \frac{S_{md} + kS_{dd}}{S_{mm} + kS_{md}} \text{ and } \hat{\alpha}^{BA}_{Mandel} = \overline{D} - \hat{\beta}_{Mandel}\overline{M}$$

with

$$k = \frac{\hat{\beta}^{BA}_{Mandel} - \rho\sqrt{\lambda}}{\lambda - \hat{\beta}^{BA}_{Mandel}\rho\sqrt{\lambda}}$$

where λ can be replaced by $\lambda_{BA} = 4$ and ρ by ρ_{BA}:

$$k = \frac{\hat{\beta}^{BA}_{Mandel} - 2\rho_{BA}}{4 - 2\hat{\beta}^{BA}_{Mandel}\rho_{BA}} \tag{4.34}$$

After some algebraic manipulations, the parameters $\hat{\beta}^{BA}_{CEIV} = \hat{\beta}^{BA}_{Mandel}$ and $\hat{\alpha}^{BA}_{CEIV} = \hat{\alpha}^{BA}_{Mandel}$ are then given by:

$\hat{\beta}^{BA}_{CEIV}$

$$= \frac{S_{dd} - 4S_{mm} + \sqrt{(S_{dd} - 4S_{mm})^2 - 4(S_{md} - 2\rho_{BA}S_{mm})(2\rho_{BA}S_{dd} - 4S_{md})}}{2(S_{md} - 2\rho_{BA}S_{mm})}$$

$$\text{and } \hat{\alpha}^{BA}_{CEIV} = \overline{D} - \hat{\beta}^{BA}_{CEIV}\overline{M} \tag{4.35}$$

In practice, if λ_{XY} is unknown, ρ_{BA} is therefore also unknown as σ^2_τ or/and σ^2_ν is/are unknown, they can be estimated from replicated data and replaced by S^2_τ or/and S^2_ν. The estimator of ρ_{BA} is given by $\hat{\rho}_{BA}$.

The CI for α^{BA}, β^{BA} and θ^{BA} can then be computed by formulae given in Section 3.1.2 where k can be replaced by formula (4.34) and the X and Y variables by, respectively M and D (the CI for β^{BA} can be computed by a t-distribution like the CI for α^{BA}). The confidence bands can be derived as explained in Chapter 3.
The null hypothesis $H_0^{CEIV} : \theta^{BA} = (0,0)'$ becomes $H_0^{UV-BA-CEIV} : \theta^{UV-BA} = (0, -\hat{\beta}^{BA}_{CEIV})'$ in the (U,V) axis. Remind that the Mandel procedure is also applied to compute the joint-CI for OR, and the null hypothesis $H_0^{OR} : \theta^{BA} = (0,0)'$ becomes $H_0^{UV-BA-OR} : \theta^{UV-BA} = (0, -\hat{\beta}^{BA}_{OR})'$ in the (U,V) axis.

CEIV regression based on BLS procedure

In order to estimate a regression line in a (M,D) plot, we propose in this section a CEIV regression based on the BLS regression described in Chapters 2 and 3. The BLS formulae (2.11) presented in Chapters 2 and 3 do not contain a term related to the correlation between the errors. Actually, the original papers dealing with the BLS regression present its full formulae with a term related to the covariance between the errors in both axes but this term is always set to zero to simplify the notations and to fulfill the assumptions (1.3) of a (X,Y) plot. Moreover, there is a confusion in the literature as this covariance term is, erroneously, written as $cov(X_i, Y_i)$ [58] while this covariance term is related to the errors. We propose in this section to write properly the BLS formulae according to the Bland and Altman model and its assumptions under homoscedasticity.

The covariance between the errors is given by the following formula:

$$\iota_{BA} = Cov\left(\frac{\tau_i + \nu_i}{2}, \nu_i - \tau_i\right) = \rho_{BA}\sqrt{\left(\frac{\sigma^2_\kappa}{4}\right)}\sqrt{(\sigma^2_\kappa)} = \rho_{BA}\frac{\sigma^2_\kappa}{2} \tag{4.36}$$

CBLS (Correlated-BLS) minimizes the criterion C_{CBLS}: the sum of weighted residuals:

$$C_{CBLS} = \frac{1}{W_{CBLS}} \sum_{i=1}^{N} (D_i - \hat{\alpha}^{BA} - \hat{\beta}^{BA} M_i)^2 = (N-2) S^2_{CBLS}$$

with

$$W_{CBLS} = \sigma^2_\epsilon = \sigma^2_\kappa + \hat{\beta}^{2BA} \sigma^2_\kappa / 4 - 2\hat{\beta}^{BA} \iota_{BA}$$

The CBLS regression is, then, given by the following formulae:

$$b = \begin{pmatrix} \hat{\alpha}^{BA}_{CBLS} \\ \hat{\beta}^{BA}_{CBLS} \end{pmatrix} = R^{-1} g \tag{4.37}$$

where

$$R = \frac{1}{W_{CBLS}} \begin{pmatrix} N & \sum_{i=1}^{N} M_i \\ \sum_{i=1}^{N} M_i & \sum_{i=1}^{N} M_i^2 \end{pmatrix}$$

and

$$g = \frac{1}{W_{CBLS}} \begin{pmatrix} \sum_{i=1}^{N} D_i \\ \sum_{i=1}^{N} \left(M_i D_i + \left(\hat{\beta}^{BA}_{CBLS} \sigma^2_\kappa - \iota \right) \frac{(D_i - \hat{\alpha}^{BA}_{CBLS} - \hat{\beta}^{BA}_{CBLS} M_i)^2}{W_{CBLS}} \right) \end{pmatrix}$$

If σ^2_κ is unknown, it can be estimated with replicated data (with S^2_τ and S^2_ν).

The CI for α^{BA}_{CBLS}, β^{BA}_{CBLS} and θ^{BA}_{CBLS} can then be computed as explained in Chapters 2 and 3: the variance-covariance matrix of the parameters is computed by $\hat{V}(b) = S^2_{CBLS} R^{-1}$ and the separated confidence intervals computed, by default, with a t distribution. The confidence bands can be derived as explained in Chapter 3.

Relationships between the two CEIV and EIV regressions

Firstly, it can be proved that both CEIV regressions (based on the Mandel or BLS procedures) previously described are mathematically equivalent to estimate the regression line but the variances of the parameters are computed differently:

$$\hat{\beta}^{BA}_{CEIV} = \hat{\beta}^{BA}_{CBLS} \text{ and } \hat{\alpha}^{BA}_{CEIV} = \hat{\alpha}^{BA}_{CBLS}$$

Secondly, by using the conversion formulae (4.32), it can be proved that estimating the regression line directly in a (X, Y) plot by $\hat{\beta}^{XY}_{DR} = \hat{\beta}^{XY}_{BLS}$ and $\hat{\alpha}^{XY}_{DR} = \hat{\alpha}^{XY}_{BLS}$ is mathematically equivalent to estimating the regression line directly in a (M, D) plot by $\hat{\beta}^{BA}_{CEIV} = \hat{\beta}^{BA}_{CBLS}$ and $\hat{\alpha}^{BA}_{CEIV} = \hat{\alpha}^{BA}_{CBLS}$:

$$\hat{\beta}^{XY}_{DR} = \hat{\beta}^{XY}_{BLS} \equiv \hat{\beta}^{BA}_{CEIV} = \hat{\beta}^{BA}_{CBLS} \text{ and } \hat{\alpha}^{XY}_{DR} = \hat{\alpha}^{XY}_{BLS} \equiv \hat{\alpha}^{BA}_{CEIV} = \hat{\alpha}^{BA}_{CBLS}$$

While λ_{XY} is the 'key' parameter in a (X, Y), ρ_{BA} is the 'key' parameter in a (M, D) space.
As DR and BLS regressions are consistent in a (X, Y) plot (under homoscedasticity), the two CEIV regressions proposed in this section are, then, also consistent in a (M, D) plot.
However, as the separated or joint confidence intervals are computed approximately in each coordinate system, a confidence interval computed in a (X, Y) plot is not exactly equivalent to the corresponding one computed in a (M, D) space and vice-versa.
Figure 4.12 shows the relationships between the CEIV and EIV regressions. Our formula (4.37) which comes from BLS regression, takes into account a constant correlation but it can be generalized to a CEIV regression where the correlation can change from point to point. On the contrary, the Mandel procedure cannot be generalized. That's why, we consider in Figure 4.12 that CBLS includes the Mandel regression.
Such regression where the correlation is not constant must be applied to estimate a regression line in a (M, D) plot under heteroscedasticity. Such regression exists in the literature (and the BLS formulae could also be modified to deal with such model structure) but we think that it is easier to estimate the regression line in a (X, Y) plot under heteroscedasticity as explained in Chapter 3 and to convert it into the (X, Y) plot with the conversion formulae (4.32).
Finally, for a given coordinate system, the CEIV regressions (based on Mandel or BLS procedures) includes the EIV regressions when there is no correlation between the errors as illustrated in Figure 4.12.

The exact CI for the slope in CEIV regression

In order to assess whether there is a proportional bias or not between the two measurement methods, we think that the easiest solution consists to compute the exact confidence interval for the slope with the Deming regression in (X, Y) plot (as described in Chapter 2) since both models, (X, Y) plot and (M, D) space, are equivalent. If the null hypothesis $H_0^\beta : \beta^{XY} = 1$ is not rejected, it does not mean that there is no proportional bias but horizontal intervals can be computed by simplicity in a (M, D) space.

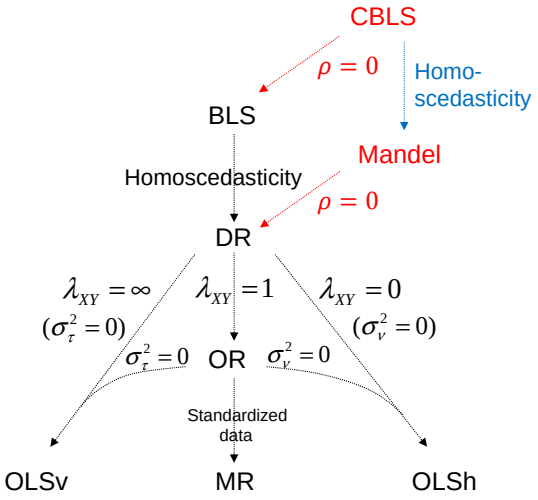

Figure 4.12: *Relationships between correlated-errors-in-variables regressions and errors-in-variables regressions*

4.4 Comparison of bias estimators under equivalence

4.4.1 Comparison of the OLSv estimators in a (X, Y) plot or (M, D) space

As already mentioned in Section 4.3.2, Bland and Altman propose to estimate the regression line in a (M, D) plot by the OLSv technique. Unfortunately, as explained in Section 4.3.3, to get a suitable regression line in a (M, D) plot, the errors in both axes must be taken into account together with their correlation. OLSv estimators are then obviously biased in a (M, D) plot. In order to assess and compare the OLSv biases in the two coordinate systems, each data set simulated in Section 2.3 in a (X, Y) plot were moved into a (M, D) space and then the OLSv technique were applied. The OLSv regression line computed in the (M, D) space is, then, moved back into the (X, Y) plot with the conversion formulae. Figure 4.13 is identical to Figure 2.2 but the means of the computed values of $\hat{\alpha}_{OLSv}^{BA}$ and $\hat{\beta}_{OLSv}^{BA}$ converted into the (X, Y) plot, are added.

We can notice in Figure 4.13 (where $n_X = n_Y = 1$) that the OLSv estimates computed in a (M, D) space are obviously biased, except for $\lambda_{XY} = 1$. Moreover, when the OLSv estimates in a (M, D) plot are moved back into a (X, Y) plot, it can be observed that the obtained biases (magenta curves) are between

those of the MR and OR (the magenta curves are between the blue curves (OR) and the dashed-blue curves (MR)). This is still valid with replicated data (charts not given).
To sum-up, the OLSv applied in a (M, D) space, as proposed by Bland and Altman, provides biased estimates (between MR and OR) and, by comparison, MR provides slightly better estimates in a (X, Y) plot.

4.4.2 Bias estimators in a (M, D) space with classical and CEIV regressions

In order to compare the estimators $\hat{\alpha}^{BA}$ and $\hat{\beta}^{BA}$ given by three classical regressions (OLSv, OR, DR with $\lambda_{XY} = 4$) and two equivalent correlated-errors-in-variables regressions (Mandel and Correlated-BLS), each data set simulated in Section 2.3 in a (X, Y) plot were moved into a (M, D) space and then the different regression techniques were applied.
For each converted data set, $\hat{\alpha}^{BA}$ and $\hat{\beta}^{BA}$ were computed for the five regressions (three classical and two CEIV regressions). Thereafter, the mean of the 100000 values of $\hat{\alpha}^{BA}$ and $\hat{\beta}^{BA}$ were computed per value λ_{XY} and regression techniques and displayed in Figure 4.14 (which is then analogous to Figure 2.2 in Chapter 2) with respect to λ_{XY} or equivalently with respect to ρ_{BA}. For the replicated data, the results are displayed in Figure 4.15 (analogous to Figure 2.3 for $n_X = n_Y = 2$), in Figure 4.16 (analogous to Figure 2.4 for $n_X = n_Y = 3$) and in Figure 4.17 (analogous to Figure 2.5 for $n_X = 4, n_Y = 2$).
Dotted-lines correspond to the true parameters ($\beta^{BA} = 0$ or $\alpha^{BA} = 0$), λ_{XY}=1 or equivalently $\rho_{BA} = 0$.

We can observe on Figure 4.14 that the biases of the classical regressions (OLSv, OR and DR with $\lambda_{XY} = 4$) are very similar (the curves are nearly superimposed on the chart) but OLSv is slightly less biased. Actually, as explained in Section 4.3.2, the minimization criteria of the classical regressions OLSv, OR and DR are similar without proportional bias in a (M, D) space. These classical regressions provide unbiased estimates for $\lambda_{XY} = 1$, otherwise the biases increase when λ_{XY} moves away from 1. Actually, when $\lambda_{XY} = 1$, the measurement errors are uncorrelated as $\rho_{BA} = 0$ but for $\lambda_{XY} \neq 1$, the measurements errors are correlated according to formula (4.29). Ignoring this correlation, leads, then, to biased estimated parameters.
Both correlated-errors-in-variables regressions, Mandel and Correlated-BLS regressions, are identical and asymptotically not biased (their estimators are consistent) whatever λ_{XY}, so whatever the correlation ρ_{BA} (the orange lines move closer to the theoretical lines when N increases) but the biases are lower when λ_{XY} moves closer to 1, so when ρ_{BA} moves closer to 0.
In the case of replicated data (Figure 4.15, Figure 4.16 and 4.17), all these findings are still valid: the shapes of the lines and their respective positions are

identical but the value of λ_{XY} or ρ_{BA} must be adapted.

4.5 Comparison of the coverage probabilities of the different CI under equivalence

In order to compare the separated and joint confidence intervals for α^{BA} and β^{BA} provided by the different regressions (OLS, OR, DR with $\lambda_{XY} = 4$, Mandel and correlated-BLS), each data set simulated in Section 2.5.1 in a (X, Y) plot were moved into a (M, D) space and then the different regression techniques were applied. The coverage probabilities were computed (at a nominal level = 95%) per value of λ_{XY} and displayed in Figure 4.18 for $n_X = n_Y = 1$ (which is then analogous to Figure 2.8 in Chapter 2).
The coverage probabilities obtained by replicated data (with λ_{XY} unknown) are displayed in Figure 4.19 (analogous to Figure 2.9) for $n_X = n_Y = 2$, Figure 4.20 (analogous to Figure 2.10) for $n_X = n_Y = 4$ and Figure 4.21 (analogous to Figure 2.11) for $n_X = 4, n_Y = 2$.

We can notice in Figure 4.18 that the coverage's probabilities (for α^{BA}, β^{BA} and their confidence region) collapse drastically when λ_{XY} moves away from 1, or equivalently when ρ_{BA} moves away from 0, and when the sample size, N, increases, for the classical regressions (OLSv, OR, DR with $\lambda_{XY} = 4$). On the contrary, the coverage probabilities are slightly lower but close to the nominal level for the two correlated-errors-in-variables regressions (Mandel and Correlated-BLS) whatever λ_{XY}, or equivalently whatever ρ_{BA}, and their coverage probabilities are very close to each other. With replicated data, all these findings are still valid but the value of λ_{XY} must be adapted.

4.6 Applications

The systolic blood pressure data have already been analyzed under homoscedasticity in Chapter 2 and under heteroscedasticity in Chapter 3. This section proposes to analyze this data set under homoscedasticity in a (M, D) space and to compare the results given by both models, the (X, Y) plot and (M, D) space.

4.6.1 Systolic blood pressure (replicated data) under homoscedasticity

Horizontal intervals

The systolic blood pressure data has been analyzed in a (X, Y) plot in Section 2.7.1 where the hypothesis H_0^{β}, no proportional bias, was not rejected. We

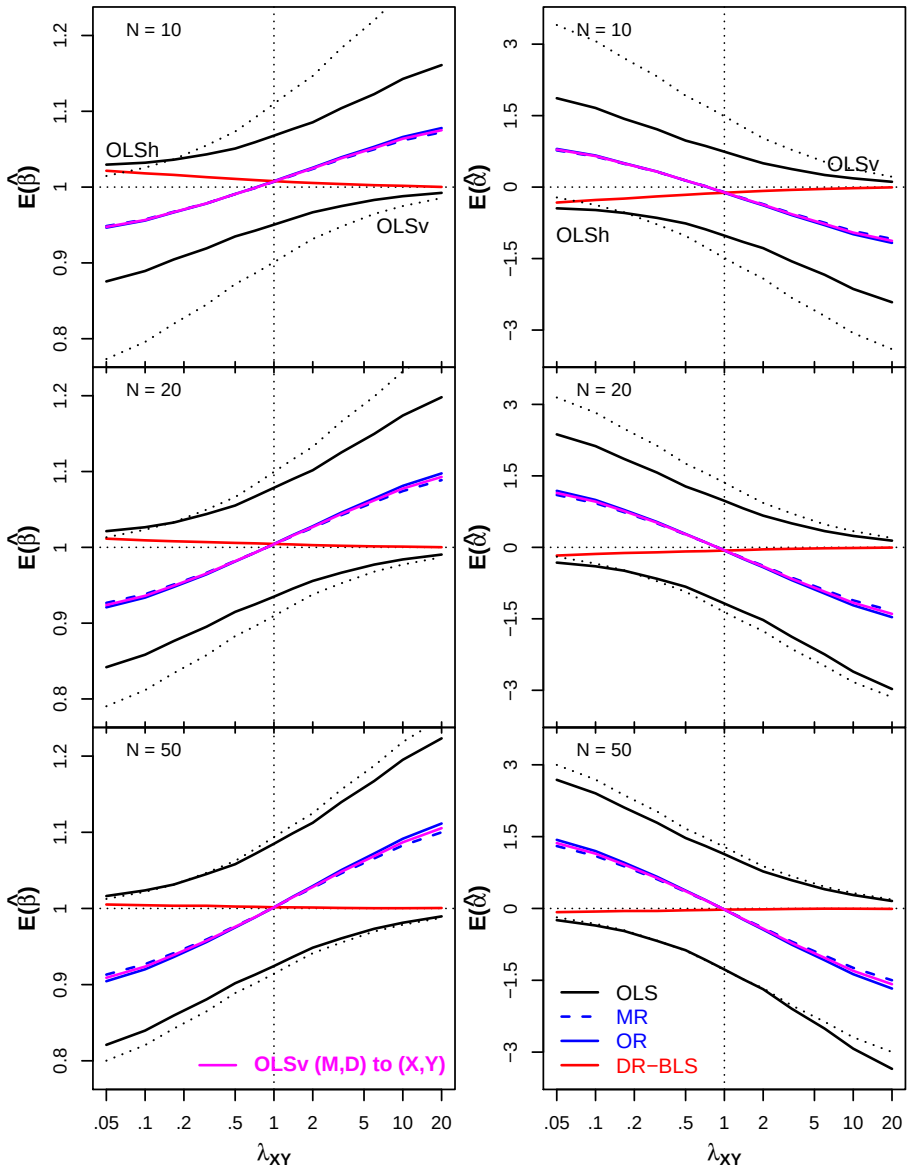

Figure 4.13: *See Figure 2.2 where* $\hat{\alpha}^{BA}_{OLSv}$ *and* $\hat{\beta}^{BA}_{OLSv}$ *converted into the* (X, Y) *plot are added*

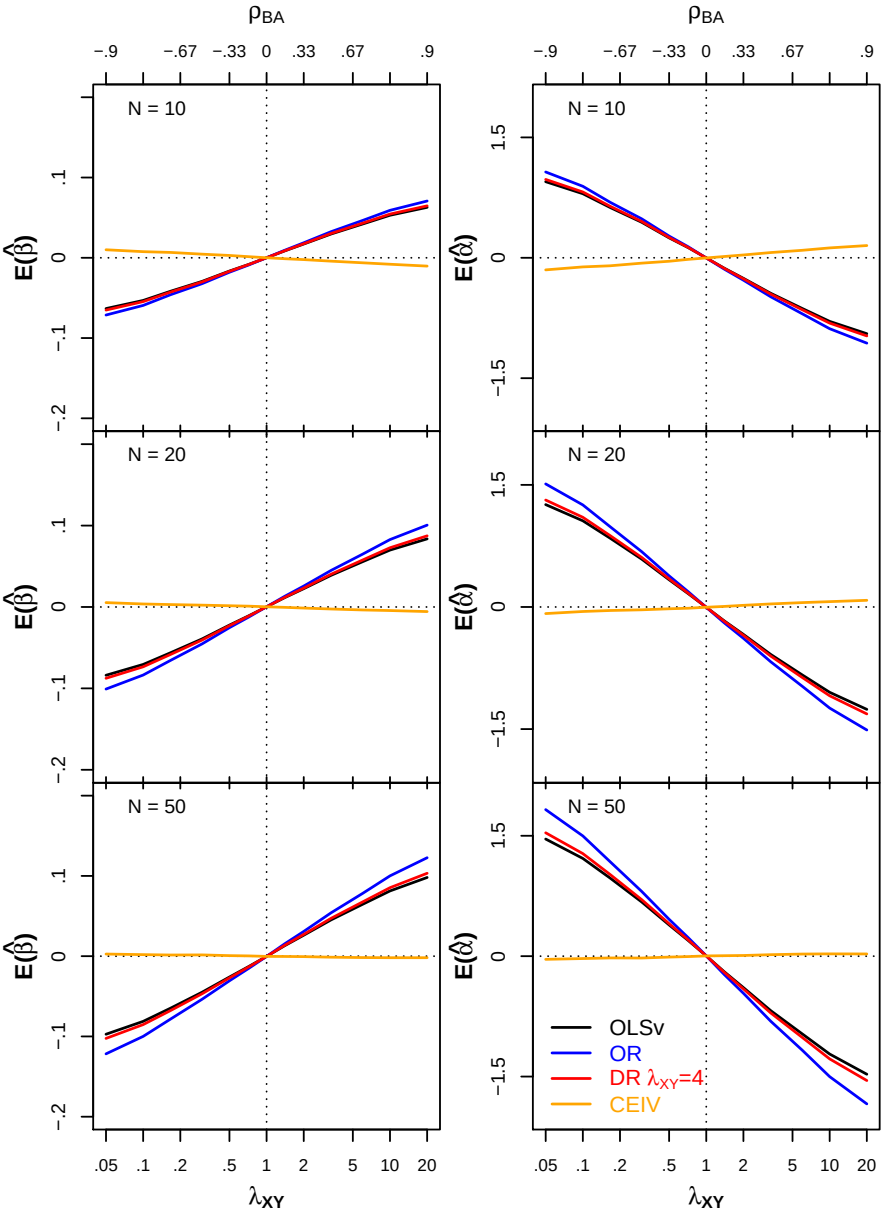

Figure 4.14: *Bias of the slope β^{BA} (left) and intercept α^{BA} (right) with respect to λ_{XY} or ρ_{BA}, for N=10 (top), N=20 (middle) and N=50 (bottom), $n_X = n_Y = 1$, see also by analogy Figure 2.2*

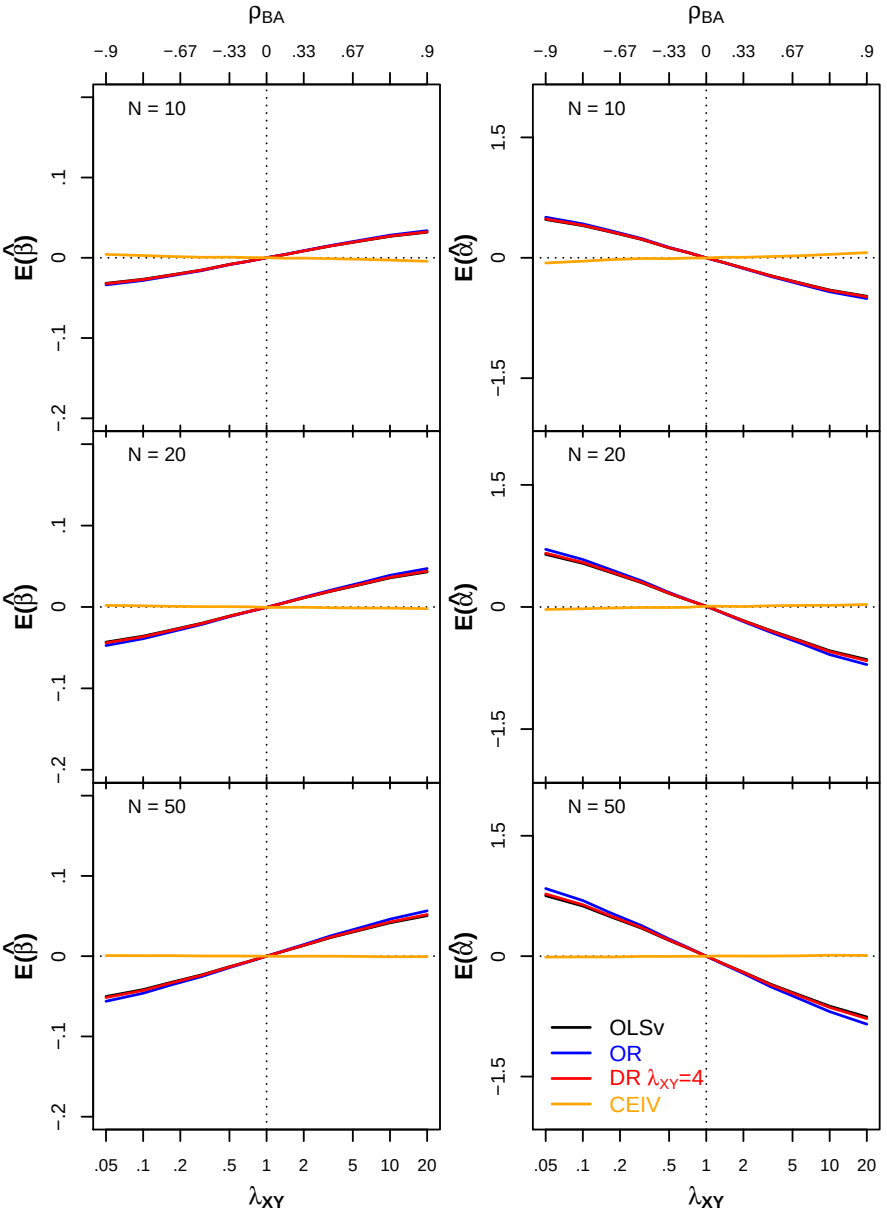

Figure 4.15: *Bias of the slope β^{BA} (left) and intercept α^{BA} (right) with respect to λ_{XY} or ρ_{BA}, for N=10 (top), N=20 (middle) and N=50 (bottom), $n_X = n_Y = 2$, see also by analogy Figure 2.3*

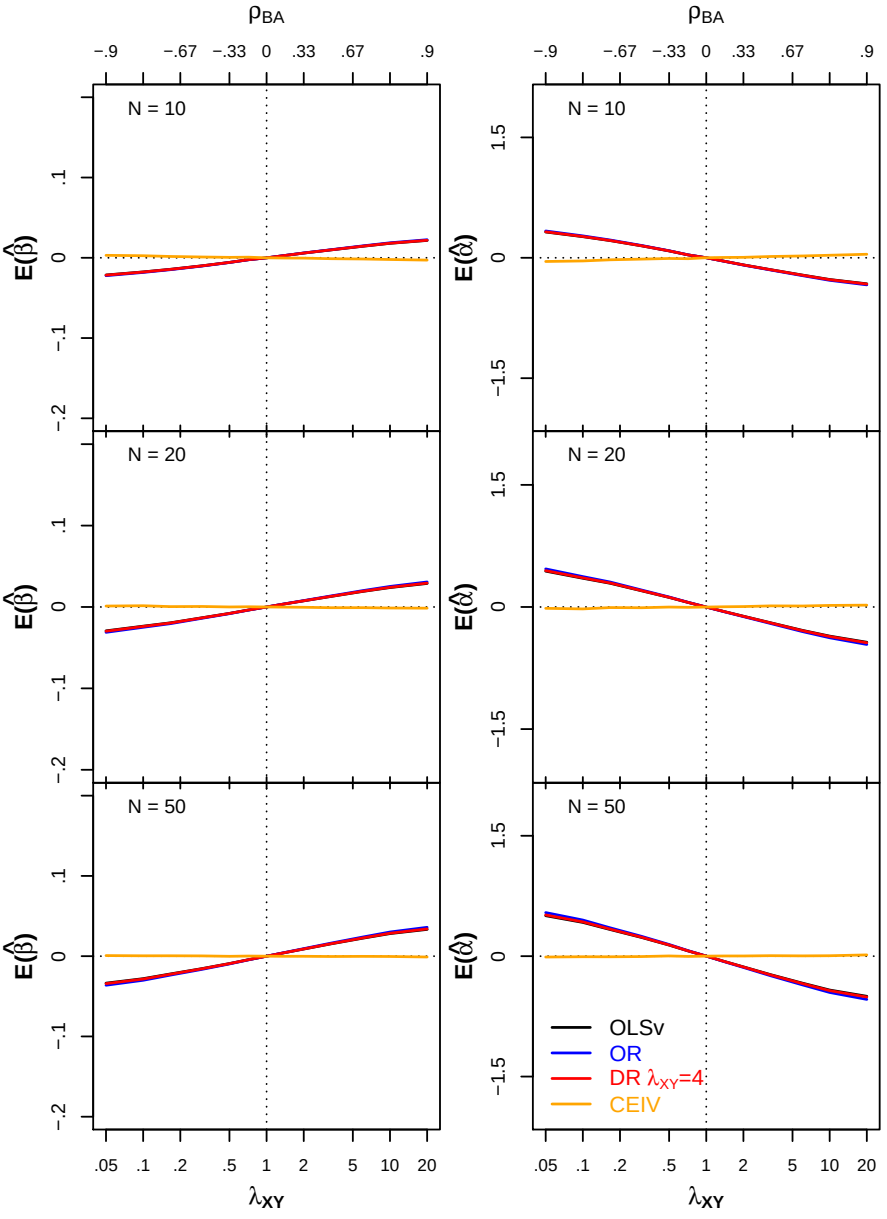

Figure 4.16: *Bias of the slope β^{BA} (left) and intercept α^{BA} (right) with respect to λ_{XY} or ρ_{BA}, for N=10 (top), N=20 (middle) and N=50 (bottom), $n_X = n_Y = 3$, see also by analogy Figure 2.4*

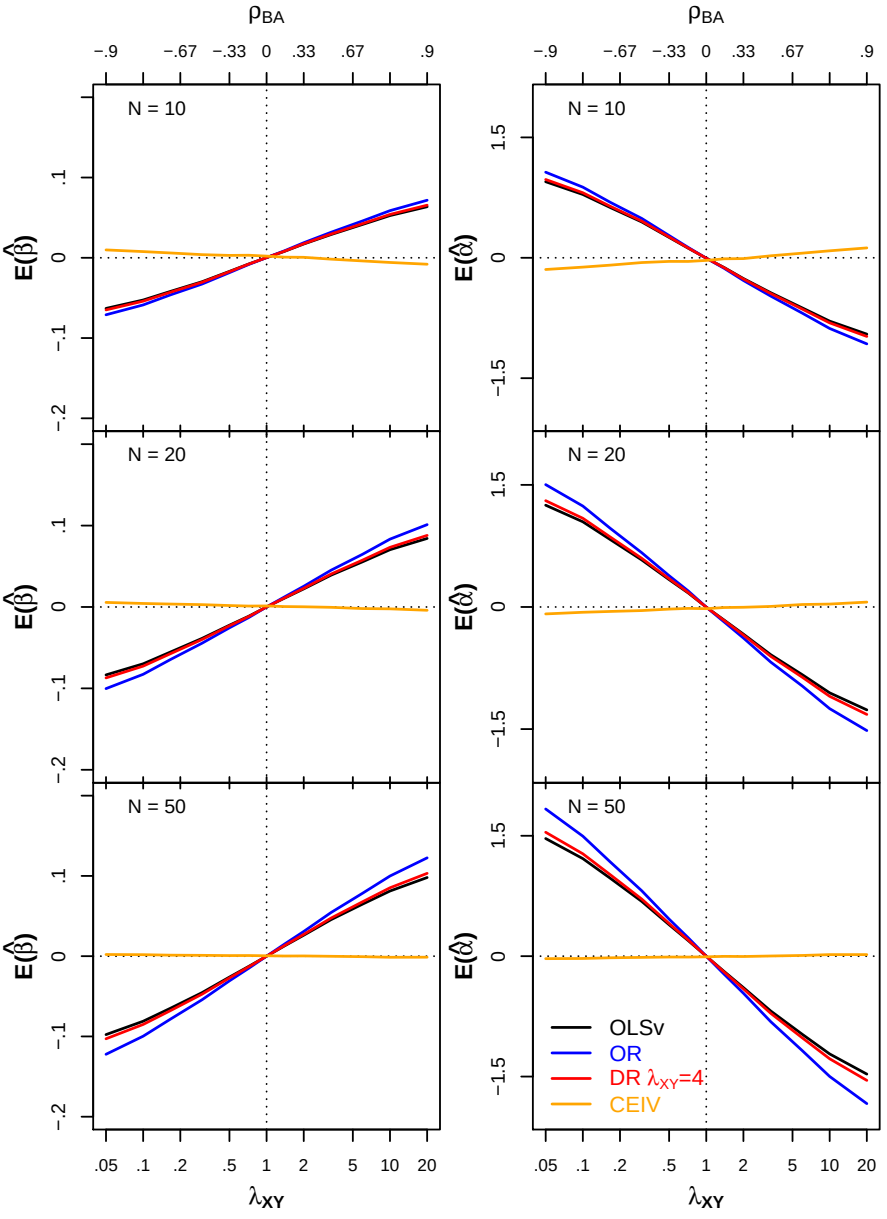

Figure 4.17: *Bias of the slope β^{BA} (left) and intercept α^{BA} (right) with respect to λ_{XY} or ρ_{BA}, for N=10 (top), N=20 (middle) and N=50 (bottom), $n_X = 4, n_Y = 2$, see also by analogy Figure 2.5*

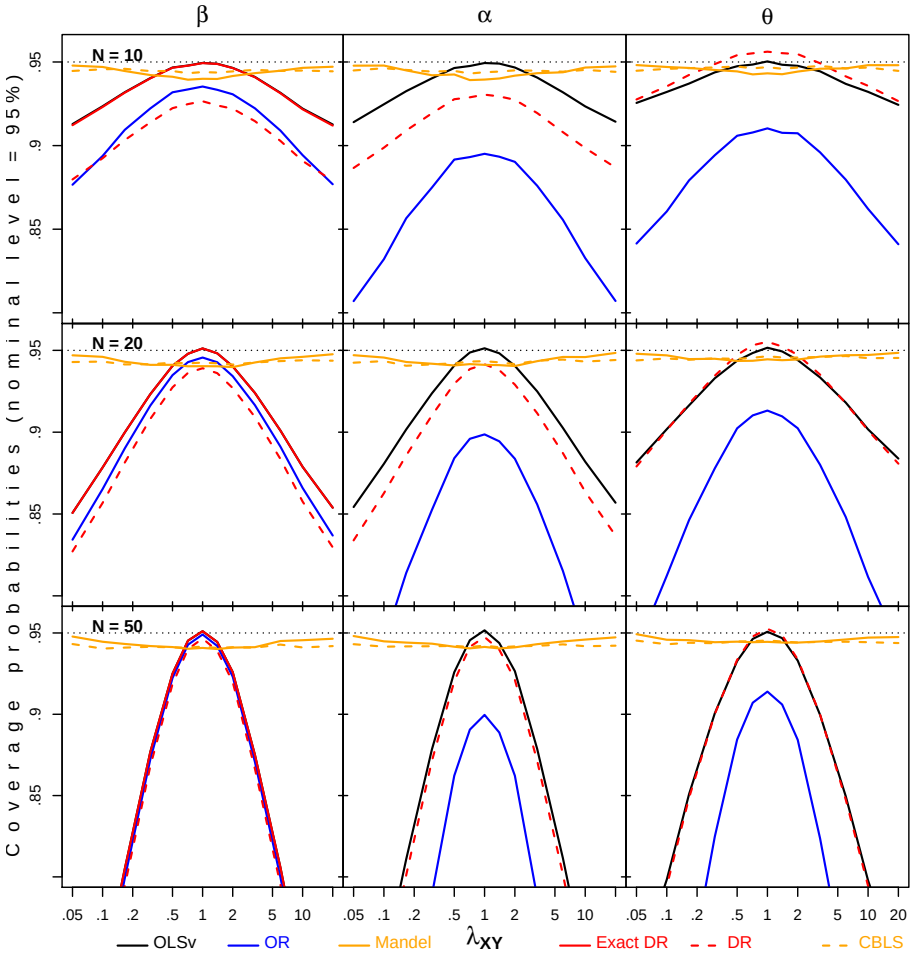

Figure 4.18: *Coverage probabilities of the CI for β (left), α (middle) and the joint-CI (right) with respect to λ_{XY} or ρ_{BA} in a logarithmic scale, for N=10 (top), N=20 (middle) and N=50 (bottom), $n_X = n_Y = 1$, see also by analogy Figure 2.8*

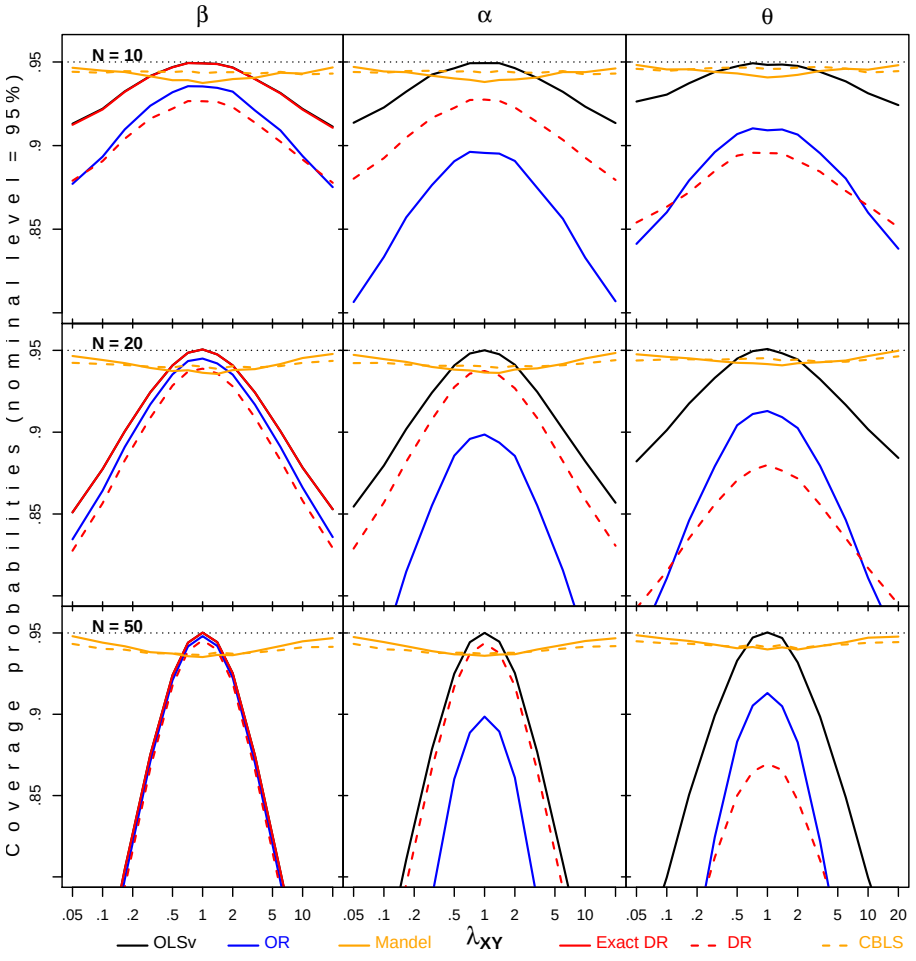

Figure 4.19: *Coverage probabilities of the CI for β (left), α (middle) and the joint-CI (right) with respect to λ_{XY} or ρ_{BA} in a logarithmic scale, for N=10 (top), N=20 (middle) and N=50 (bottom), $n_X = n_Y = 2$, see also by analogy Figure 2.9*

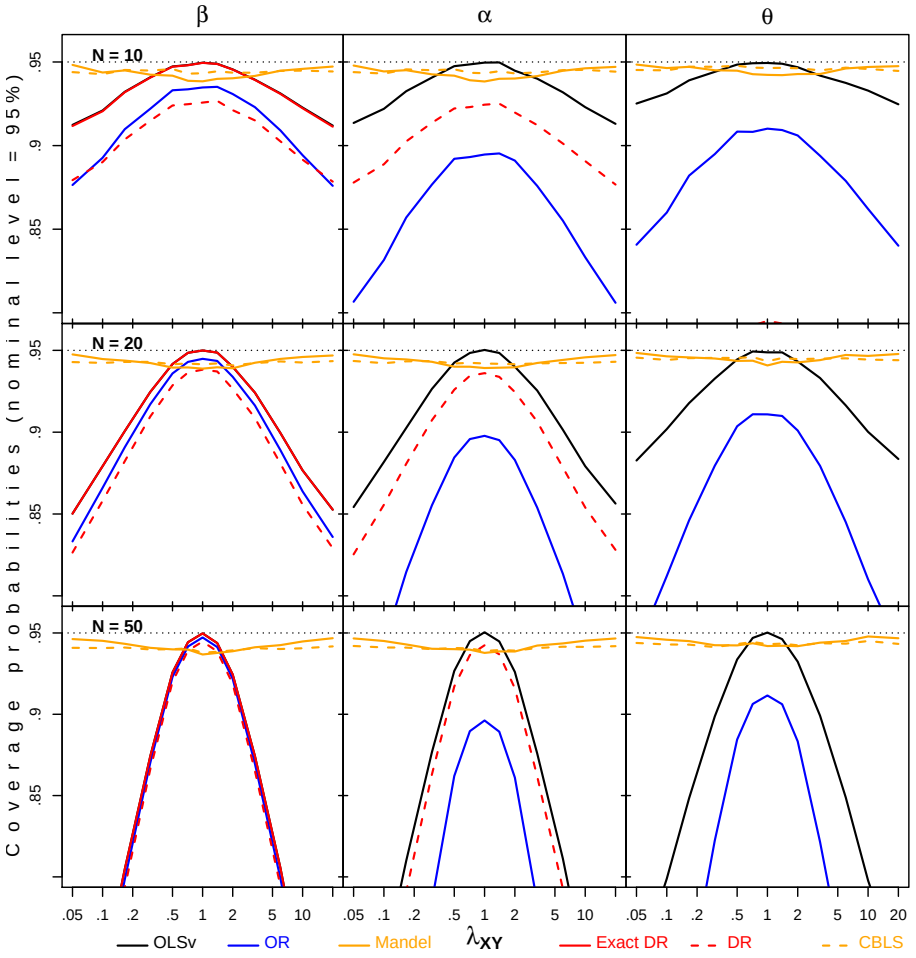

Figure 4.20: *Coverage probabilities of the CI for β (left), α (middle) and the joint-CI (right) with respect to λ_{XY} or ρ_{BA} in a logarithmic scale, for N=10 (top), N=20 (middle) and N=50 (bottom), $n_X = n_Y = 4$, see also by analogy Figure 2.10*

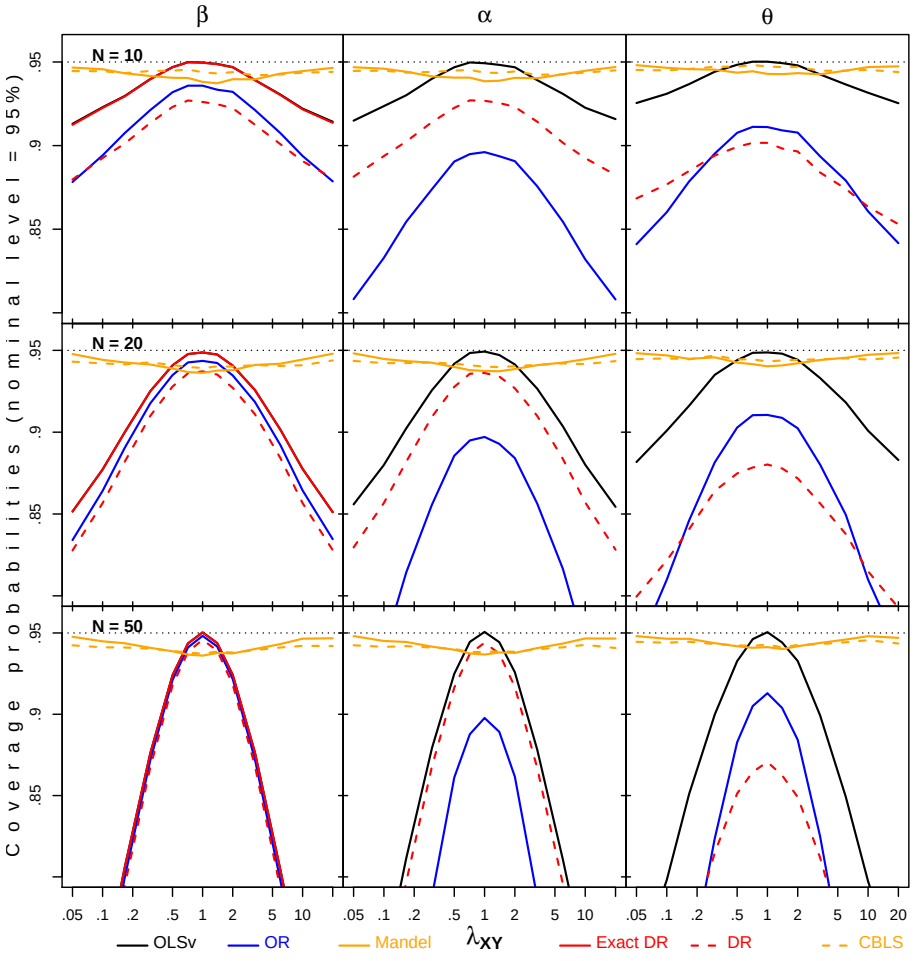

Figure 4.21: *Coverage probabilities of the CI for β (left), α (middle) and the joint-CI (right) with respect to λ_{XY} or ρ_{BA} in a logarithmic scale, for N=10 (top), N=20 (middle) and N=50 (bottom), $n_X = 4, n_Y = 2$, see also by analogy Figure 2.11*

propose then to compute the horizontal agreement and tolerance intervals in a (M, D) plot. These intervals are displayed in Figure 4.22. We suppose that differences up to 10 mmHg are not meaningful in a clinical point of view, and then, the acceptance interval is: $[-10, 10]$. The predictive level is set to 95%: $1 - \psi = 0.95$ and the confidence level to 80%: $1 - \gamma = 0.80$.
According to formulae given in the previous sections, we have:
$\overline{D} = 15.620$, $S_D^2 = 358.493$, $\hat{\sigma}_{Ds}^2 = 438.859$ and $S_{Ds}^2 = 120.549$.
We can notice that the estimates $\hat{\sigma}_{Ds}^2$ and S_{Ds}^2 are greatly different. Actually, as explained in Chapter 2, there are many outliers in SBP data set, and these outliers increase the variance of the differences, and so increase the value of S_D^2 and by the way leads to an overestimated value of σ_{Ds}^2 provided by the estimator $\hat{\sigma}_{Ds}^2$. In contrast, the estimator S_{Ds}^2 is not influenced by such outliers as its formula does not depend of S_D^2.
The agreement limits are: $[-25.440, 56.679]$ and the XL-AI: $[-29.871, 61.110]$, the β tolerance interval is: $[-6.030, 37.269]$ with S_ω and $[-6.357, 37.596]$ with $\hat{\sigma}_\omega$, the $\beta\gamma$ tolerance interval is: $[-6.757, 37.996]$ with S_ω and $[-7.248, 38.488]$ with $\hat{\sigma}_\omega$.

We can notice in Figure 4.22 that the agreement intervals (and the XL-AI) are much larger than the tolerance intervals. Indeed, as explained in the previous paragraph, σ_{Ds}^2 is overestimated by $\hat{\sigma}_{Ds}^2$ because of the outliers while S_ω^2 is not influenced by such outliers and $\hat{\sigma}_\omega^2$ is less influenced by outliers than $\hat{\sigma}_{Ds}^2$ as S_D^2 is divided by N in the formula of $\hat{\sigma}_\omega^2$. The tolerance intervals are close to each other when computing with S_ω or $\hat{\sigma}_\omega$. As the sample size is large, the β-TI and $\beta\gamma$-TI (with 80% confidence level) are close to each other.
Finally, it can be noticed that none of these intervals are inside the acceptance interval. This means that it can be expected that the future single differences between the measurement methods can be higher than 10 mmHg in absolute values. So, the measurement methods are not interchangeable.

Regression analysis

As explained in Section 2.7.1, in the SBP data, λ (=λ_{XY} as $n_X = n_Y$) is estimated by $\hat{\lambda} = 2.223$ and therefore the correlation between the measurement errors in a (M, D) space is estimated by $\hat{\rho}_{BA} = 0.380$.
For the different regression's lines in the (M, D) space, the estimated coefficients are: $\hat{\beta}_{OLSv} = 0.036$ and $\hat{\alpha}_{OLSv} = 10.809$, $\hat{\beta}_{OR} = 0.059$ and $\hat{\alpha}_{OR} = 7.590$, $\hat{\beta}_{DR} = 0.040$ and $\hat{\alpha}_{DR} = 10.272$ with λ=4, $\hat{\beta}_{Mandel} = \hat{\beta}_{CBLS} = -0.045$ and $\hat{\alpha}_{Mandel} = \hat{\alpha}_{CBLS} = 21.708$

The different regression's lines are displayed in Figure 4.23-top left with the equivalence line $D = 0$. We can observe that the OLSv and DR with λ=4 are very close to each other (superimposed in the chart) while the OR is also close

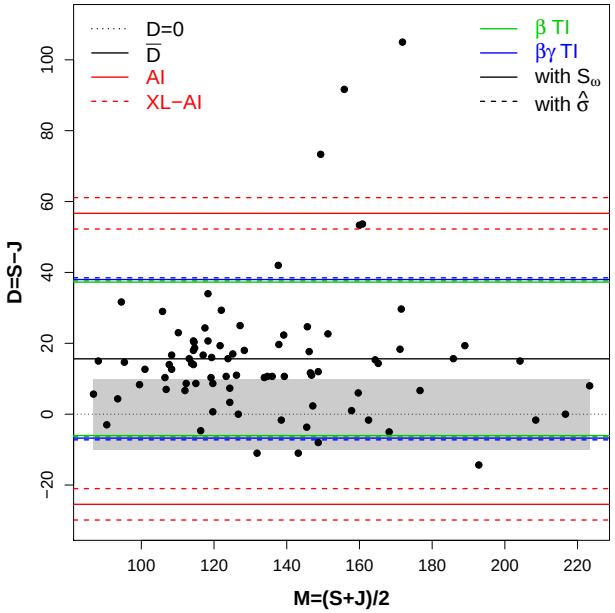

Figure 4.22: *SBP data ($n_X = n_Y = 3$) in a (M, D) space with horizontal agreement and tolerance intervals, and the acceptance interval $[-\Delta, \Delta] = [-10, 10]$ in grey area*

to the OLSv and DR. The two black lines are the 'extremes' or 'outside' lines and correspond to the correlated-errors-in-variables regressions with $\lambda_{BA} = 4$ and $\rho = 1$ or $\rho = -1$. It is easy to check that these two lines are the analogous of the two OLS lines in the (X, Y) plot:

$$\hat{\theta}^{XY}_{OLSv} \equiv \hat{\theta}^{BA}_{CEIV} \text{ with } \rho = 1$$

$$\hat{\theta}^{XY}_{OLSh} \equiv \hat{\theta}^{BA}_{CEIV} \text{ with } \rho = -1$$

The Mandel and the correlated-BLS regressions confounded (under homoscedasticity) are, obviously, inside the two extreme lines and are the best estimated lines because the measurement errors in both axes with their correlation are taken into account the best.

By analogy to Figure 2.14, Figure 4.23 shows the full information about the separated or joint confidence intervals for α and β whatever ρ_{BA}. In Figure 4.23-top right, we can observe that the CI given by Mandel or correlated-BLS regressions for β are very close to each other. We also display on the chart the horizontal line $\beta = 0$ corresponding to the null hypothesis $H_0^\beta : \beta^{BA} = 0$ and a vertical line at the estimated value $\hat{\rho}_{BA}$ with the corresponding 95% CI in wheat color (the confidence interval for ρ is derived from the one for λ).

As the lines $\beta = 0$ and $\hat{\rho}_{BA}$ cross inside the curves of Mandel or Correlated-BLS confidence intervals, the null hypothesis of no proportional bias is not rejected. This is still valid from the lower to the upper bounds of the CI for ρ_{BA}. Similarly, on Figure 4.23-bottom left, the null hypothesis of no constant bias, $H_0^\alpha : \alpha^{BA} = 0$, is not rejected at $\hat{\rho}_{BA}$, except on the lower bound of its CI.
Figure 4.23-bottom right shows the approximate confidence ellipses from $\rho = -1$ to $\rho = 1$. As the 'equivalence point' $(\alpha = 0, \beta = 0)$ is outside the approximate orange ellipse given by Mandel or Correlated-BLS at $\hat{\rho}_{BA}$, the null hypothesis $H_0 : \theta^{BA} = (0, 0)'$ (no bias between the two devices) is rejected. Moreover, this null hypothesis is rejected whatever the measurement errors correlation.

To sup-up, these full charts with the separated or joint confidence intervals are equivalent for the estimated parameters as $\hat{\theta}^{XY} \equiv \hat{\theta}^{BA}$ and very similar for the confidence intervals (the confidence intervals are computed approximately, so they are not equivalent but very similar between the (X, Y) plot or the (M, D) space).

The hyperbolic confidence bands can be derived from the confidence ellipse as explained in Section 3.1.2, according to the following formula:

$$x'\theta \in x'\hat{\theta} \pm \sqrt{c}\sqrt{x'\hat{\Sigma}^{-1}x}$$

Figure 4.24-top displays the scatterplot of the SBP data in a (M, D) space with the Mandel and Correlated-BLS regressions (confounded) and with their confidence bands which are very close to each other (superimposed in the chart). The confidence ellipses provided by Mandel or Correlated-BLS are equivalently very similar and superimposed in the chart (Figure 4.24-bottom). It can be noticed that the equivalence line $D = 0$ is not inside the confidence bands or equivalently that the equivalence point $\theta^{BA} = (0, 0)'$ is outside the confidence ellipses.

4.6.2 Systolic blood pressure (unreplicated data) under homoscedasticity

Horizontal intervals

The systolic blood pressure data with unreplicated data has been analyzed in a (X, Y) plot in Section 2.7.2 where the hypothesis H_0^{β}, no proportional bias, was not rejected. We propose then to compute the horizontal agreement and tolerance intervals in a (M, D) plot. These intervals are displayed in Figure 4.25. The acceptance interval, the confidence and the predictive levels are identical to those of Section 4.6.1.
We can notice in Figure 4.25 that the agreement and tolerance intervals are close to each other. Indeed, in this example, there is no outlier. The $\beta\gamma$ tolerance interval is, obviously slightly larger than the β tolerance interval which is, obviously, slightly larger than the agreement interval.
Finally, it can be noticed that the acceptance interval is inside all these intervals. This means that it can be expected that the future single differences between the measurement methods can be higher than 10 mmHg in absolute values. So, the measurement methods (remind that in this example, we do not compare two measurement methods but two set of measures provided by the same device) are not interchangeable. However, it can be noticed that both measurement methods would be interchangeable whether the acceptance was equal to $[-20, 20]$.

Regression analysis

As explained in Section 2.7.2, to deal with unreplicated data, we consider only the two first sets of measures given by the manual device. The measurement errors and their correlation are therefore unknown. For the different regression's lines, the estimated coefficients are: $\hat{\beta}_{OLSv} = 0.002$ and $\hat{\alpha}_{OLSv} = -1.543$, $\hat{\beta}_{OR} = 0.002$ and $\hat{\alpha}_{OR} = -1.566$, $\hat{\beta}_{DR} = 0.002$ and $\hat{\alpha}_{DR} = -1.548$ with λ=4

The correlated-errors-in-variables regressions Mandel and Correlated-BLS cannot be applied because ρ_{BA} is unknown and inestimable.
Figure 4.26 is analogue to Figure 2.16. The OLSv, OR and DR (with $\lambda = 4$)

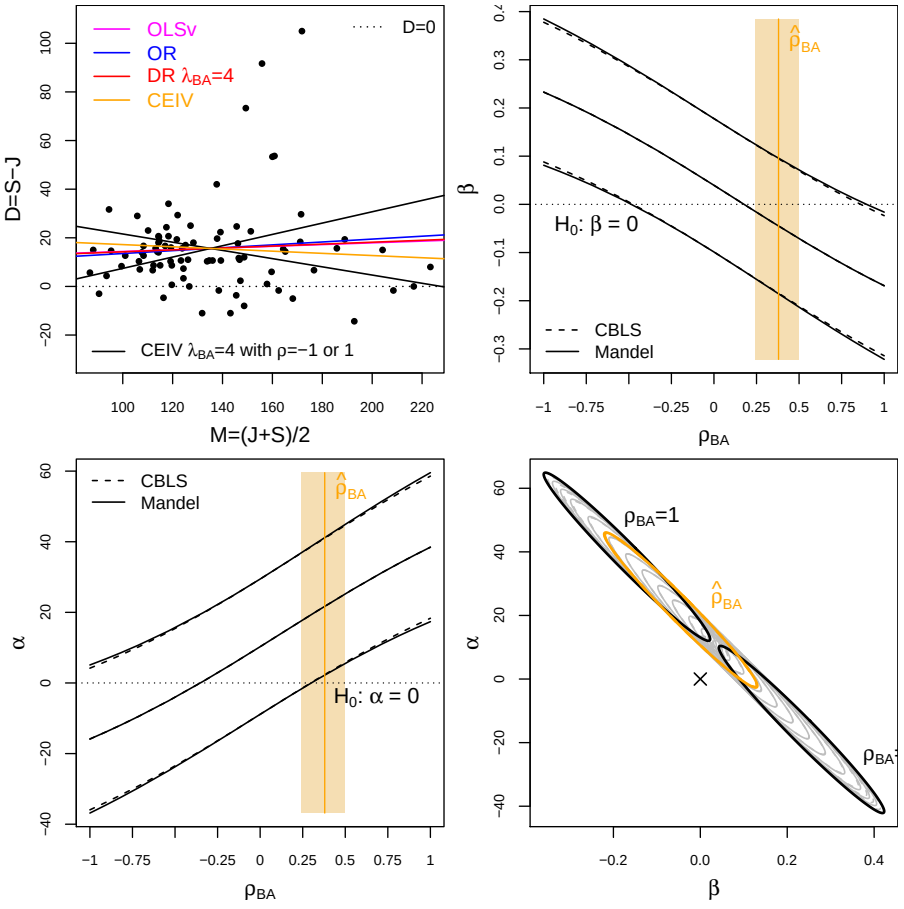

Figure 4.23: *SBP data ($n_X = n_Y = 3$), the different regressions lines in a (M, D) space (Top-left) and their CI for the parameters (Top-right for β, Bottom-left for α and Bottom-right for the joint-CI), see also by analogy Figure 2.14*

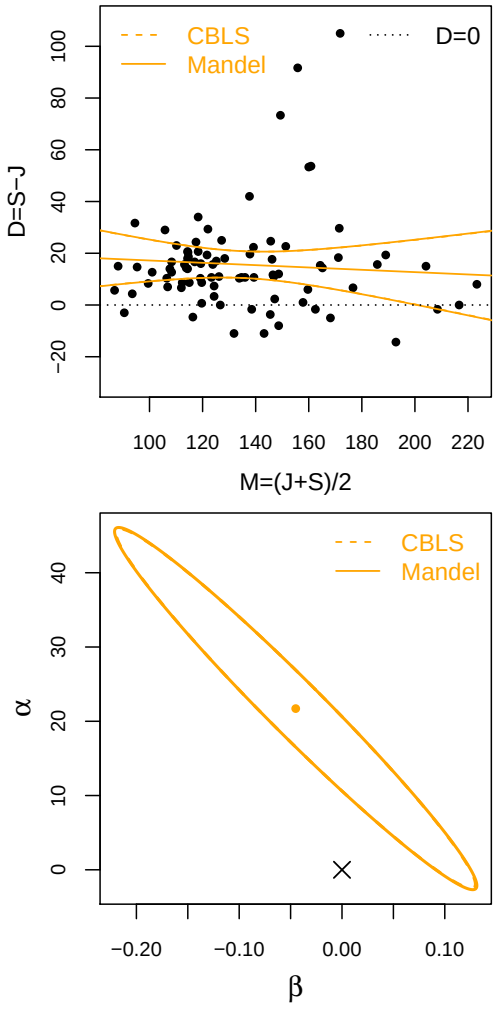

Figure 4.24: *SBP data ($n_X = n_Y = 3$) in a (M, D) space with the confidence bands (top) or confidence ellipses (bottom) provided by Mandel and Correlated BLS regressions*

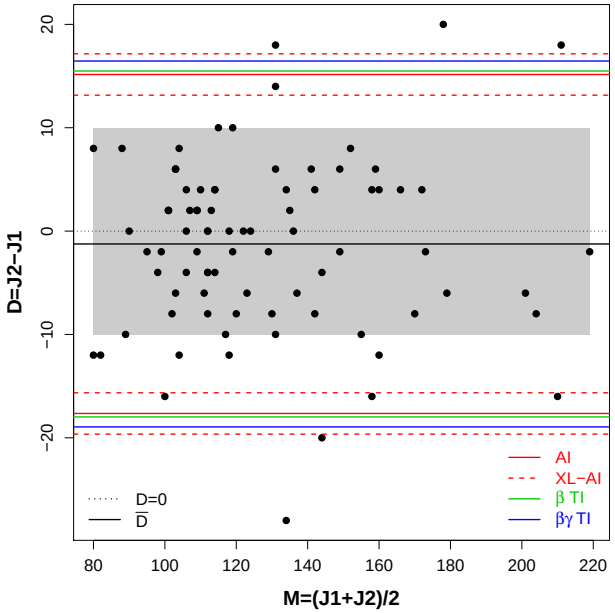

Figure 4.25: *SBP data ($n_X = n_Y = 1$) in a (M, D) space with horizontal agreement and tolerance intervals, and the acceptance interval $[-\Delta, \Delta] = [-10, 10]$ in grey area*

lines are very close to each other, such that it is not possible to distinguish them on the plot.
The CI for α or β are displayed from $\rho_{BA} = -1$ to $\rho_{BA} = 1$ as previously explained. The separated null hypotheses (1.12) $H_0^{\alpha} : \alpha^{BA} = 0$ and $H_0^{\beta} : \beta^{BA} = 0$ are not rejected whatever ρ_{BA}. When these two hypotheses are tested jointly, it can be noticed that the 'equivalence point' $(\alpha^{BA} = 0, \beta^{BA} = 0)$ is inside all the ellipses from $\rho_{BA} = -1$ to $\rho_{BA} = 1$, which means that the joint hypothesis is also not rejected whatever ρ_{BA}. In other words, we do not reject the hypothesis that the first measures are equivalent to the second, as already mentioned in Section 2.7.2.

4.7 Prediction Intervals in a (X, Y) plot and (M, D) space

4.7.1 Prediction Intervals for single measurements

Papers dealing with prediction intervals are rarer. Carstensen [92] proposes approximate prediction intervals based on OLSv procedure (or bayesian methodology) but he neglects the usual curvature term $(X - \overline{X})^2$ of the prediction. Moreover, to deal with replicate data and be able to predict single measures, he proposes to plot all the pairs (X_{ij}, Y_{ik}) and regress them. Unfortunately, these pairs of points are correlated and he proposes to neglect this correlation. Note that Francq [19] shows that estimating the regression with the means (X_i, Y_i) or all the pairs (X_{ij}, Y_{ik}) provides the same estimated line, the confidence intervals on the parameters are different but with similar coverage probabilities. Unfortunately, a problem of effective sample size occurs when regressing all the pairs with $n_X \neq n_Y$ [19]. Bland and Altman [11] propose straight lines by approximation based on OLSv. All these approaches will, then, not be analyzed in this thesis.
Del Rio and al. [58] give formulae to compute the prediction interval or the confidence interval for a given value X_0. Unfortunately, the theoretical development is not clear (it is based on the OLSv), the concept of the prediction is not well-argued and their formulae cannot be generalized to predict singles measures. Mandel [56] gives the formula to compute the variance of the expectation of the predicted value $\hat{Y}_0$ for a given value X_0. Unfortunately, his formula cannot be generalized and none interval is given.

We propose in this section hyperbolic predictive intervals to predict single measures for $n_X = n_Y$ or $n_X \neq n_Y$ for an observed value X_0. In other words, when a single measure is observed by the measurement method X, the single measure that will be provided by the measurement method Y is expected to lie in that interval. The following development is based on the BLS regression and

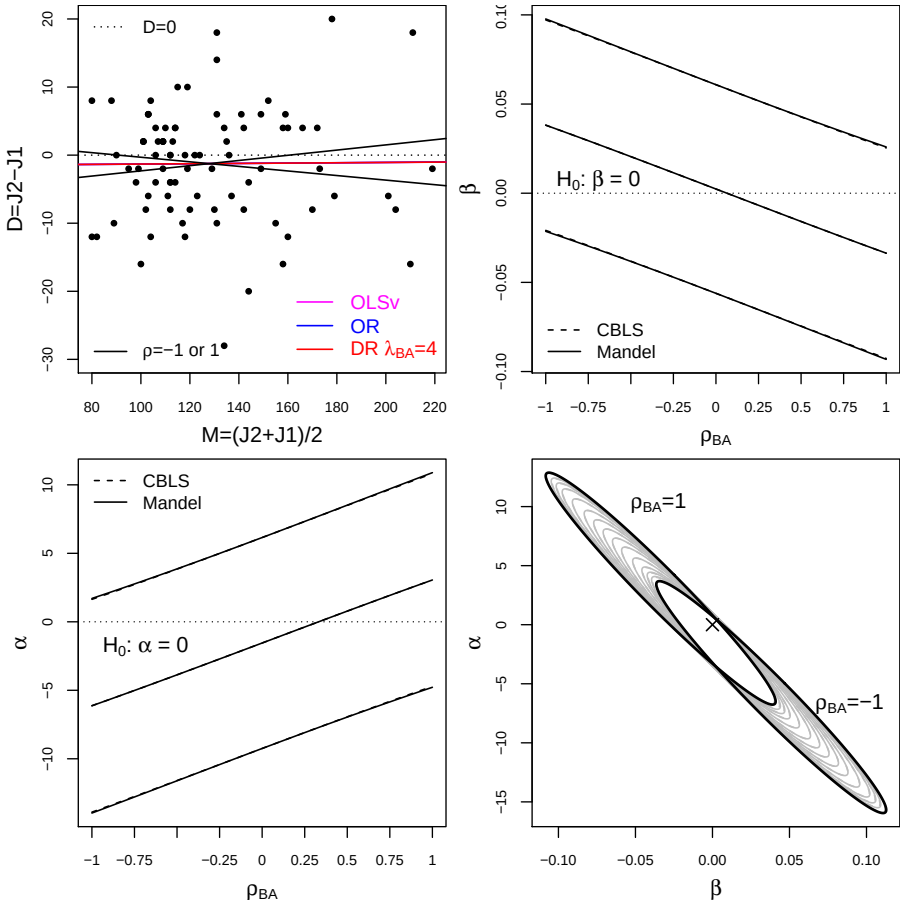

Figure 4.26: *SBP data ($n_X = n_Y = 1$), the different regressions lines in a (M, D) space (Top-left) and their CI for the parameters (Top-right for β, Bottom-left for α and Bottom-right for the joint-CI), see also by analogy Figure 2.16*

is, according to our knowledge, novel.

The regression model is given by:

$$Y_i = \alpha_{BLS} + \beta_{BLS} X_i + \epsilon_i$$

For an observed measure X_0, the corresponding predicted value is given by:

$$\hat{Y}_0 = \hat{\alpha}_{BLS} + \hat{\beta}_{BLS} X_0 \tag{4.38}$$

Under OLSv assumption, the variance of the expectation of a prediction is given by: $\mathrm{Var}(E(\hat{Y}_0)) = S^2_{\hat{\alpha}_{BLS}} + S^2_{\hat{\beta}_{BLS}} X_0^2 + 2X_0 S_{\hat{\alpha},\hat{\beta}_{BLS}}$.
But, to take into account the error associated with the observed measurement X_0, the square of the slope must be introduced [58]. In this section, as the goal is to predict a future single measure by the measurement method Y, the measurement error variance of the Y device can, also, be added to the variance of the expectation of a prediction, to get:

$$Var(\hat{Y}_0) = S^2_{\hat{Y}_0} \approx \hat{\beta}^2_{BLS}\sigma^2_\tau + S^2_\zeta + \sigma^2_\nu \tag{4.39}$$

with

$$S^2_\zeta = S^2_{\hat{\alpha}_{BLS}} + S^2_{\hat{\beta}_{BLS}} X_0^2 + 2X_0 S_{\hat{\alpha},\hat{\beta}_{BLS}} \tag{4.40}$$

where σ^2_τ and σ^2_ν can be replaced by S^2_τ and S^2_ν if needed.

The prediction interval (PI) is then given by the following formulae:

- with σ^2_τ and σ^2_ν known:

$$PI : \hat{Y}_0 \pm z_{1-\gamma/2} S_{\hat{Y}_0} \tag{4.41}$$

- when σ^2_τ and σ^2_ν are replaced, respectively, by S^2_τ and S^2_ν:

$$PI : \hat{Y}_0 \pm t_{1-\gamma/2,N_P} S_{\hat{Y}_0} \tag{4.42}$$

  where the degrees of freedom, N_P, are computed with the Welch-Satterthwaite equation (see Section 4.2.2, where $\hat{\beta}_{BLS}$ is, here, considered constant):

$$N_P = \frac{(S^2_\zeta + S^2_\nu + \hat{\beta}^2_{BLS} S^2_\tau)^2}{\frac{S^4_\zeta}{N-2} + \frac{S^4_\nu}{N_Y} + \frac{\hat{\beta}^4_{BLS} S^4_\tau}{N_X}}$$

These formulae can easily be adapted to a (M, D) space according to the formulae of the correlated-BLS regression.
A prediction in a (M, D) space for an observed mean M_0 is given by the following formula:

$$\hat{D}_0 = \hat{\alpha}^{BA}_{CBLS} + \hat{\beta}^{BA}_{CBLS} M_0 \tag{4.43}$$

and the variance of the prediction is given by:

$$Var(\hat{D}_0) = S^2_{\hat{D}_0} \approx \hat{\beta}^{2BA}_{CBLS}(\sigma^2_\tau + \sigma^2_\nu)/4 + S^{2BA}_\zeta + (\sigma^2_\nu + \sigma^2_\tau) \tag{4.44}$$

with

$$S^{2BA}_\zeta = S^2_{\hat{\alpha}_{CBLS}} + S^2_{\hat{\beta}_{CBLS}} M_0^2 + 2M_0 S_{\hat{\alpha},\hat{\beta}_{CBLS}} \tag{4.45}$$

where σ^2_τ and σ^2_ν can be replaced by S^2_τ and S^2_ν if needed.

The prediction interval (PI) is then given by the following formulae:

- with σ^2_τ and σ^2_ν known:

$$PI : \hat{D}_0 \pm z_{1-\gamma/2} S_{\hat{D}_0} \tag{4.46}$$

- when σ^2_τ and σ^2_ν are replaced, respectively, by S^2_τ and S^2_ν:

$$PI : \hat{D}_0 \pm t_{1-\gamma/2, N_P^{BA}} S_{\hat{D}_0} \tag{4.47}$$

  where the degrees of freedom, N_P^{BA}, are computed with the Welch-Satterthwaite equation (see Section 4.2.2, where $\hat{\beta}^{BA}_{CBLS}$ is, here, considered constant):

$$N_P^{BA} = \frac{(S^{2BA}_\zeta + (1 + \hat{\beta}^{2BA}_{CBLS}/4)S^2_\nu + (1 + \hat{\beta}^{2BA}_{CBLS}/4)S^2_\tau)^2}{\frac{S^{4BA}_\zeta}{N-2} + \frac{(1+\hat{\beta}^{2BA}_{CBLS}/4)^2 S^4_\nu}{N_Y} + \frac{(1+\hat{\beta}^{2BA}_{CBLS}/4)^2 S^4_\tau}{N_X}}$$

4.7.2 Coverage probabilities of the prediction intervals

In order to assess the coverage probabilities, 100000 simulations were run as described in Section 2.3:

- with $\sigma^2_\tau = 0.5, 0.5, 0.75$ and the corresponding value of $\sigma^2_\nu = 0.75, 0.5, 0.5$,
- with $n_X = n_Y = 1$, $n_X = n_Y = 2$, $n_X = n_Y = 4$ and $n_X = 4, n_Y = 2$,
- the ξ_i values were fixed from 10 to 20 by step 2 (6 values of ξ_i, under equivalence: $\xi_i = \eta_i$) and, then, 6 X_i values are simulated ($X_i = \xi_i + \tau_i$),
- the corresponding predicted values, $\hat{Y}_i$ are computed with formula 4.38 and the prediction interval computed with formula 4.41 or 4.42,
- then, it is checked whether η_i lies into the corresponding prediction interval or not.

This simulation scheme is synthesized as follows:

$$\xi_i \text{ and } \eta_i \xrightarrow[\eta_i+\nu_i]{\xi_i+\tau_i} X_i \text{ and } Y_i \xrightarrow{\hat{\alpha}_{BLS}+\hat{\beta}_{BLS}X_i} \hat{Y}_i \xrightarrow{4.41 \text{ or } 4.42} PI(\hat{Y}_i) \rightarrow Y_i \in ? PI$$

Each simulated data set is also converted into a (M, D) space and the prediction interval computed with formula 4.46 or 4.47.

Figure 4.27 displays the coverage probabilities of the prediction intervals with respect to ξ (from 10 to 20) for $n_X = n_Y = 1$ into a (X, Y) plot (left) or with respect to $(\xi + \eta)/2$ into a (M, D) space (right). Figure 4.28, Figure 4.29 and Figure 4.30 are the analogous for, respectively, $n_X = n_Y = 2$, $n_X = n_Y = 4$ and $n_X = 4, n_Y = 2$.
It can be noticed that the coverage probabilities of the approximate prediction intervals are very close to each other for the different values of λ and close to the nominal level. They are slightly better when the sample size increases. The coverage probabilities obtained in a (X, Y) plot are very similar to those obtained into a (M, D) space.

4.7.3 Application: prediction intervals for single measurements in SBP data - replicated data

Figure 4.31 displays the prediction intervals computed for the systolic blood pressure (with $n_X = n_Y = 3$) in the (X, Y) plot or (M, D) space with, respectively, the DR-BLS regressions line or the correlated-BLS regression line.
It can be noticed that the prediction intervals are not inside the acceptance interval (where the acceptance interval is $Y = X \pm 10$ in the (X, Y) plot) which means that both measurement methods are not interchangeable. The prediction intervals are very similar between both models, the (X, Y) plot or the (M, D) space. As the null hypothesis of no proportional bias was not rejected, the horizontal β tolerance interval (previously computed) is also displayed in the (M, D) space for didactic purposes, as the prediction interval is the analogous of the β tolerance interval. We can notice that, indeed, the prediction interval and the β tolerance interval coincide very well.
For instance, if a measurement is taken for a new patient with the manual device (X (J)) and equal to 100 mmHg, then the prediction for a measurement taken with the semi-automatic device is:
$\hat{Y}$ (S) $= 21.230 + 0.956 \cdot 100 = 116.827$ mmHg and the 95% prediction interval is: [94.850,138.803] (see the arrow in Figure 4.31-top).
The interpretation is less easy in a (M, D) plot, but for the mean measurements between 100 and 200 mmHg, the maximum of the differences (in absolute values) is equal to 39.721 mmHg. This means that the two devices can differ at most up to 39.721 mmHg in this range (see the arrows in Figure 4.31-bottom).

4.8 Practical recommendations

In this section, we provide some guidelines for the practitioners.
As the (X, Y) plot and (M, D) space are analogous, we recommend to test

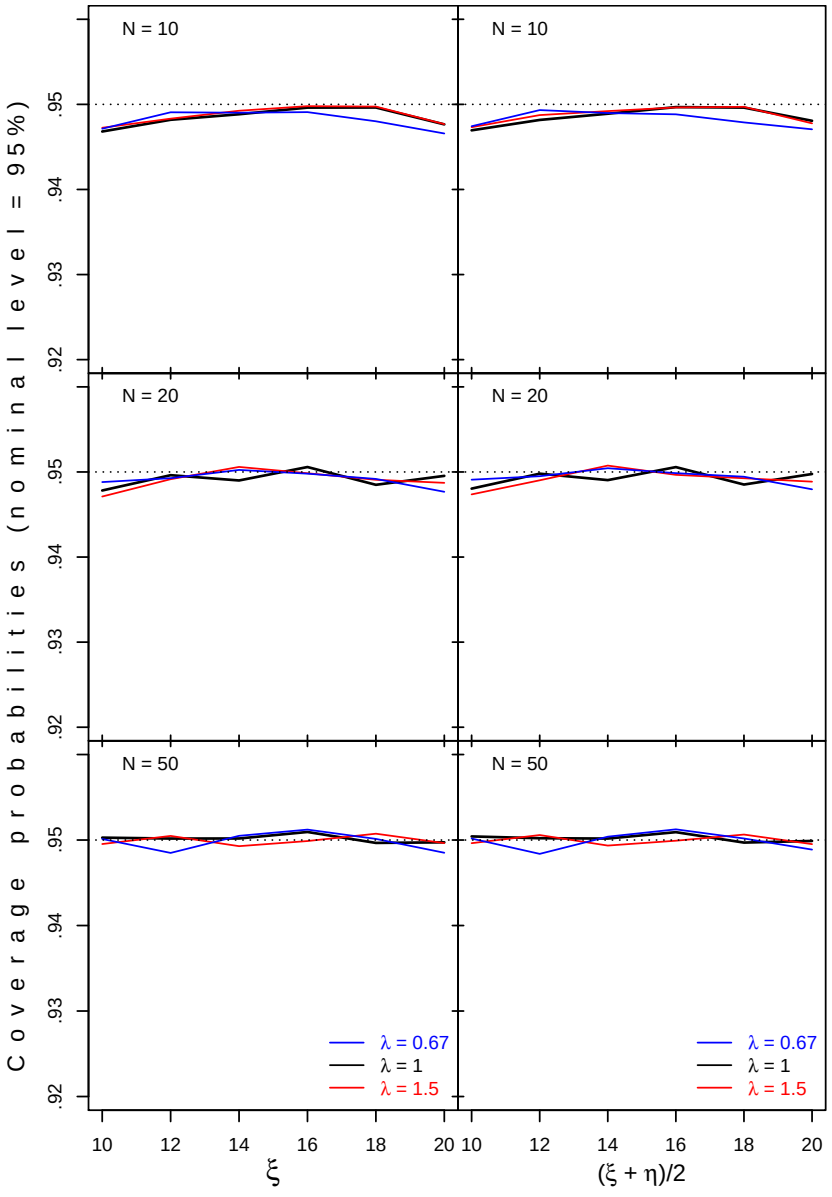

Figure 4.27: *Coverage probabilities of the prediction intervals for* $n_X = n_Y = 1$ *and* $N = 10$ *(top),* $N = 20$ *(middle) and* $N = 50$ *(bottom), with respect to* ξ *(left) into a* (X, Y) *plot or with respect to* $(\xi + \eta)/2$ *into a* (M, D) *space (right)*

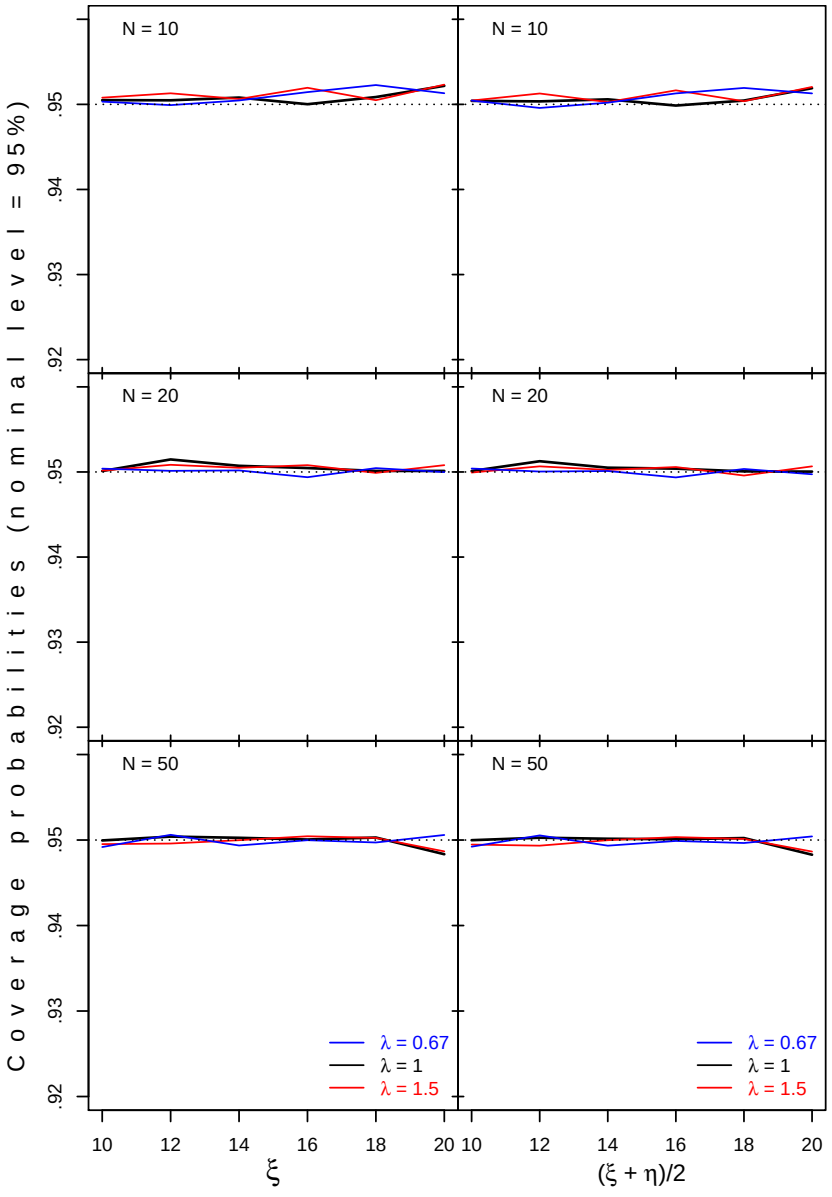

Figure 4.28: *Coverage probabilities of the prediction intervals for* $n_X = n_Y = 2$ *and* $N = 10$ *(top),* $N = 20$ *(middle) and* $N = 50$ *(bottom), with respect to* ξ *(left) into a* (X, Y) *plot or with respect to* $(\xi + \eta)/2$ *into a* (M, D) *space (right)*

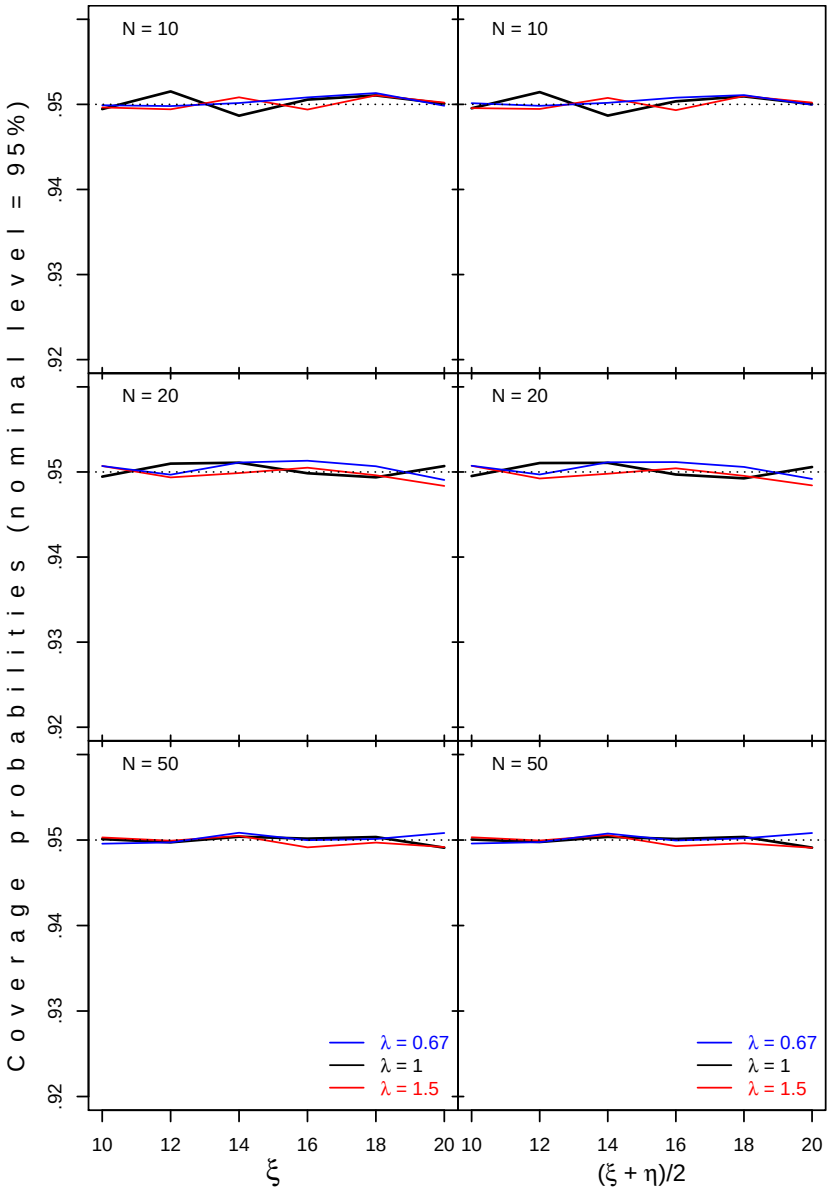

Figure 4.29: *Coverage probabilities of the prediction intervals for* $n_X = n_Y = 4$ *and* $N = 10$ *(top),* $N = 20$ *(middle) and* $N = 50$ *(bottom), with respect to* ξ *(left) into a* (X, Y) *plot or with respect to* $(\xi + \eta)/2$ *into a* (M, D) *space (right)*

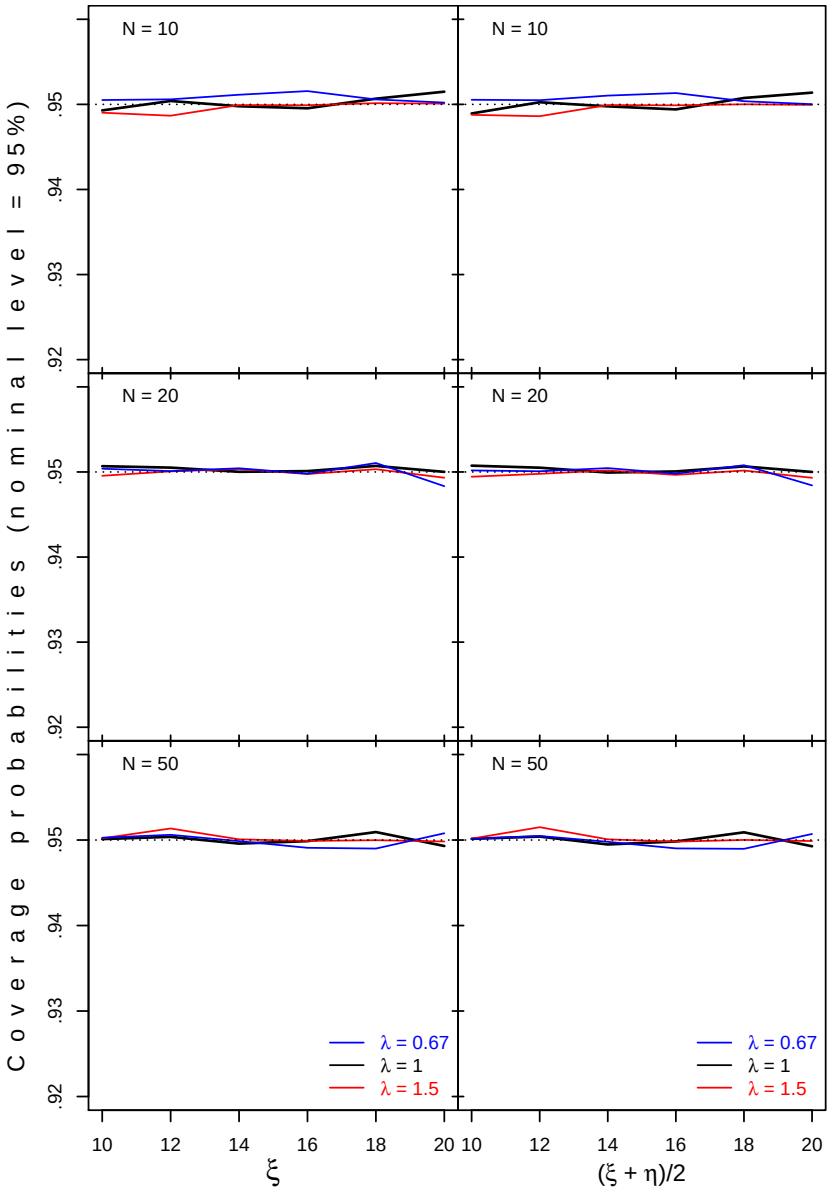

Figure 4.30: *Coverage probabilities of the prediction intervals for* $n_X = 4, n_Y = 2$ *and* $N = 10$ *(top),* $N = 20$ *(middle) and* $N = 50$ *(bottom), with respect to* ξ *(left) into a* (X, Y) *plot or with respect to* $(\xi + \eta)/2$ *into a* (M, D) *space (right)*

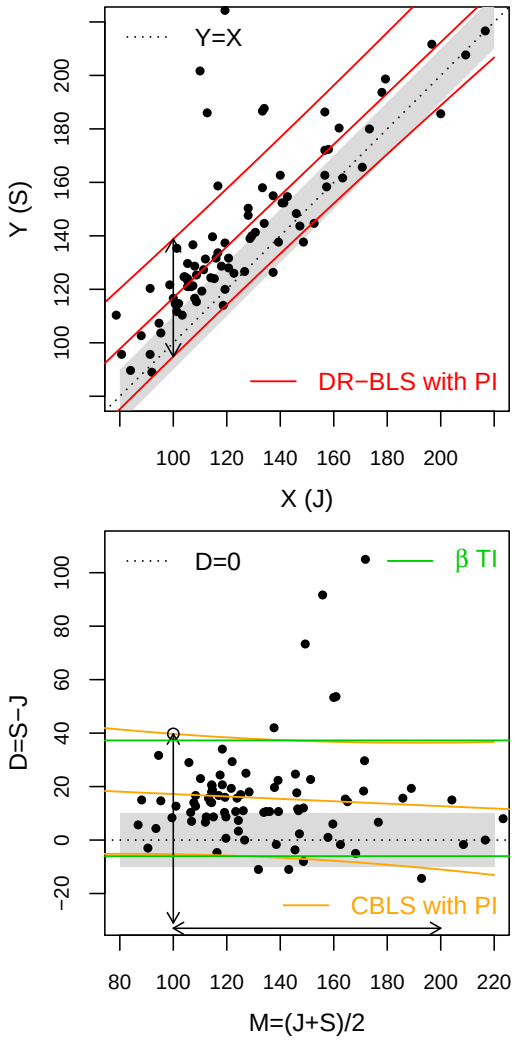

Figure 4.31: *SBP data ($n_X = n_Y = 3$), in a (X, Y) plot (top) or (M, D) space (bottom) with respectively, the DR-BLS or Correlated-BLS regressions lines and the prediction intervals*

the null hypothesis $H_0^\beta : \beta^{XY} = 1$ with the exact confidence interval given by Tan and Iglewicz [28]. If this hypothesis is not rejected, we recommend, by simplicity, to compute and display in the (M, D) space the horizontal tolerance intervals (β or $\beta\gamma$) and to compare the tolerance interval to the acceptance interval $[-\Delta, \Delta]$. If the tolerance interval is inside the acceptance interval, then the two measurement methods are interchangeable.
If the null hypothesis $H_0^\beta : \beta^{XY} = 1$ is rejected, a regression line must be estimated:
DR or BLS regression lines in a (X, Y) plot (to take into account the measurement errors) or Mandel or correlated-BLS regression lines in a (M, D) space (to take into account the measurement errors and their correlation).
Then, the prediction interval can be computed and compared to the acceptance interval: $Y = X \pm \Delta$ in a (X, Y) plot or $[-\Delta, \Delta]$ in a (M, D) space. If the prediction interval is inside the acceptance interval (for a given range of values of X where it is expected that the future measures will lie in practice, i.e. $[80, 200]$ mmHg for the SBP data), then the two measurement methods can be considered interchangeable.
To assess the strict equivalence, full charts can be displayed in a (X, Y) plot or equivalently, in a (M, D) space as described in Sections 2.7 and 4.6.

Chapter 5

Conclusion

Concluding remarks

The main goal of this thesis was to compare two methodologies available in the literature to assess the equivalence of two measurement methods: the errors-in-variables regressions in a classical (X, Y) measures space or the Bland and Altman procedure based on a plot of the measurement differences with respect to their averages in a (M, D) space. While the first one focuses on confidence intervals to test statistically the equivalence with strict hypotheses, the second one focuses on predictive (or agreement) intervals to test practically the equivalence with flexible hypotheses. These two methodologies are, very often, presented separately in the literature. However, this thesis reconciles the two methodologies and shows that they are inseparable and complement each other. Indeed, it is essential to deal with (correlated)-errors-in-variables regression in the (M, D) space in order to develop predictive intervals which are not horizontal while the predictive intervals can also be computed and displayed in a (X, Y) plot.
Therefore, the main message of this thesis is that, errors-in-variables can (unfortunately) not be avoided in method comparison studies, although the Bland and Altman plot was, initially, applied to avert the complexity of errors-in-variables regressions.
The theoretical background of the Bland and Altman model is, actually, often neglected in the literature. However, when the model is properly written, the regressions are equivalent for the two methodologies and the approximate confidence intervals very similar, as well as the predictive intervals. Indeed, both models are, actually, equivalent: the coordinate system (X, Y) plot can be moved into a (M, D) space and vice-versa.

In Chapter 2, the relationships between seven well-known regressions applied in the context of measurement methods comparison are summarized and com-

pared, with or without replicated data in a (X, Y) plot: the Ordinary Least Square regressions (OLSv and OLSh), the geometric Mean regression (MR), the Orthogonal Regression (OR), the Deming Regression (DR), the Bivariate Least Square regression (BLS) and the non-parametric regression of Passing and Bablok (PB). The ratio of the measurement errors variances in both axes, λ or λ_{XY} (when taking into account the number of replicates), is a very important parameter. It has been used as a leitmotiv in order to compare the different regressions and to display the results of the simulations. An unbiased estimator of λ_{XY} is proposed in order to compare the precisions of both devices. OLSv, OLSh, MR, OR and PB regressions provide biased estimated lines when their respective assumption are violated and consequently, their coverage probabilities collapse. DR or BLS regressions (confounded under homoscedasticity) are the most general regressions and provide (asymptotically) unbiased estimated lines whatever λ_{XY}. However, their small sample biases are lower for $\lambda_{XY} > 1$. The exact confidence interval for the slope in the DR is excellent whatever λ_{XY} while BLS provides slightly better coverage probabilities for $\lambda_{XY} > 1$. Full diagrams with the regression parameters confidence intervals from OLSh to OLSv are very useful in practice. Indeed, the hypothesis of equivalence can sometimes be generalized by being rejected or not rejected whatever λ_{XY}. This is very useful when λ_{XY} is unknown and inestimable because no repeated measures are available and the measurement errors are then unknown.

Chapter 3 focuses on the joint regression parameters confidence intervals given by DR and BLS regressions, and additionally, the Mandel and the Galea-Rojas et al. procedures. The confidence ellipses are expressed equivalently, as confidence bands. These four methodologies provide very close coverage probabilities but the ellipses (or confidence bands) can be different in the presence of outliers. We eventually recommend the BLS for its advantages: it can be generalized to a correlated-BLS regression if needed, its formulae are quite simple without the need to convert the data into another coordinate system (as Mandel does), and its confidence intervals are computed very easily with analogies to OLSv. Under heteroscedasticity, the coverage probabilities of DR and BLS collapse when the variances are unknown. These coverage probabilities can, then, be improved by modeling the locally estimated variances with suitable variance functions.

Chapter 4 deals with the Bland and Altman method. The advantages of working in the (M, D) space are, firstly, to focus directly on the differences which are obviously emphasized in a (M, D) space and secondly, to be able to compute horizontal intervals if one can assume that there is no proportional bias. The horizontal agreement intervals are, first, compared to horizontal tolerance intervals. Tolerance intervals are interesting alternatives to the agreement intervals of Bland and Altman as they provide coverage probabilities closer to

the nominal level. Moreover, this chapter gives a robust estimator, S^2_{Ds}, of the future single differences (σ^2_{Ds}), while the one given by Bland and Altman, $\hat{\sigma}^2_{Ds}$, is influenced by outliers. When a proportional bias is observed, a regression line must be estimated in the (M, D) space. Unfortunately, the classical regressions are biased and their coverage probabilities collapse for $\lambda_{XY} \neq 1$. Indeed, the disadvantage of the (M, D) space is the complexity of the model where the correlation between the error terms must be taken into account in order to estimate a consistent regression line. This chapter proposes two consistent regressions (confounded under homoscedasticity) based on the BLS or Mandel procedures, with coverage probabilities close to each other and close to the nominal level whatever λ_{XY}. These two new regressions take into account the errors in both axes and also their correlation. These correlated-errors-in-variables regressions are, indeed, equivalent to DR or BLS regressions in a (X, Y) plot. Finally, this chapter gives the formulae to compute prediction intervals in a (X, Y) plot or in a (M, D) space in order to predict a single measurement or a single difference. The coverage probabilities of these prediction intervals are very close to each other and close to the nominal level.

Guidelines

This thesis is not only intended to develop theoretical models but also offers to the practitioners tools to investigate method comparison studies. Figure 5.1 synthesizes this thesis by providing some rules and guidelines for the practitioners. The R code can be provided on request by the author and a R package is in preparation.

The choice between strict or flexible equivalence must be discussed between the practitioners and the statistician. When comparing different measurement methods, flexible hypotheses can be preferred when the differences are not clinically important up to a given threshold. On the other hand, when comparing measures provided by two observers or raters with the same device, or two identical devices, strict hypotheses could be preferred.

With strict hypotheses and under heteroscedasticity, the equivalence can be assessed in a classical (X, Y) plot with BLS regression. The equivalence is rejected if the equivalence point $(\beta = 1, \alpha = 0)$ lies outside the confidence ellipse for the parameters of the BLS regression. The confidence ellipses can equivalently be represented by confidence bands around the regression line. These confidence bands are easier to display and to interpret, and the equivalence is rejected if the equivalence line $Y = X$ intercepts the hyperbolic confidence bands. Under heteroscedasticity, we strongly recommend to model the variances by variance functions if possible.

With strict hypotheses and under homoscedasticity, the equivalence can be assessed as well with confidence ellipses for the parameters or equivalently the

confidence bands by either:

- the BLS regression line for λ_{XY} known or $\hat{\lambda}_{XY}$ with its confidence interval. We also recommend to display full charts with confidence intervals for the parameters from OLSv to OLSh in order to assess the hypotheses $H_0 : \alpha^{XY} = 0$ and $\beta^{XY} = 1$.
- or the correlated-BLS regression line for $\lambda_{BA} = 4$ and ρ_{BA} known or $\hat{\rho}_{BA}$ with its confidence interval. We also recommend to display full charts with confidence intervals for the parameters from $\rho_{BA} = -1$ to $\rho_{BA} = 1$ in order to assess the hypotheses $H_0 : \alpha^{BA} = 0$ and $\beta^{BA} = 0$.

These charts are especially useful when there is no replicate. The conclusions can, then, be usually generalized whatever the measurement errors.
With flexible equivalence, we recommend to test the presence of proportional bias by the exact confidence interval for the slope with the (X, Y) data. Horizontal intervals can, then, be computed and displayed in a (M, D) space when the hypothesis of no proportional bias is not rejected. We recommend tolerance intervals which are a better alternative than agreement intervals. Both devices are, then, interchangeable when the tolerance interval is inside the acceptance interval $[-\Delta, \Delta]$.
When the assumption of no proportional bias cannot be fulfilled, a regression line must be estimated: the BLS regression in a (X, Y) plot or equivalently the correlated-BLS regression in a (M, D) space. The predictive interval can, then, be computed and displayed around the regression line. The equivalence is, then, assessed by comparing the predictive interval to the acceptance interval: $Y = X \pm \Delta$ in a (X, Y) plor, or $\pm\Delta$ in a (M, D) space. The measurement methods are then interchangeable if the predictive interval lies inside the acceptance interval.

Future research

Some challenges remain, of course, to be overcome. Two main issues are first, developing $\beta - \gamma$ content tolerance interval around an errors-in-variable regression line and more generally predictive intervals under heteroscedasticity in order to predict single measure. The variance functions are essential to deal with heteroscedasticity in order to get predictive intervals which are not 'over-fitted' with locally estimated variances as those sometimes given in the literature [58]. Note that the predictive intervals given in this thesis are computed, by default, point by point (for a given X_0 or M_0). Predictive bands can, then, be investigated in order to cover with a given probability the values of one or more future observations. Secondly, there is, obviously, a need to develop statistical methods to compare more than two measurement methods. Some papers already exist in the literature [94] but the theoretical backgrounds are not always clear and the methodologies can still be improved.

Furthermore, as explained in the introductory chapter, four different methodologies coexist in order to monitor and inspect the analytical methods or devices in the laboratories: the validation, the transfert, the interlaboratories studies and the method comparison studies investigated in detail in this thesis. There is, then, a need, to step back and to investigate the relationships between these four methodologies to get an 'overall picture'. Indeed, the method comparison studies and the interlaboratories are, for instance, closely linked.
More complicated designs of experiments for method comparison studies should also be investigated. It is, indeed, essential to take into account the sampling effect, repeated measures, different observers,... Such model will include additional random effects and become more complex to estimate. Bayesian statistical modeling [92] can, therefore, be promising.

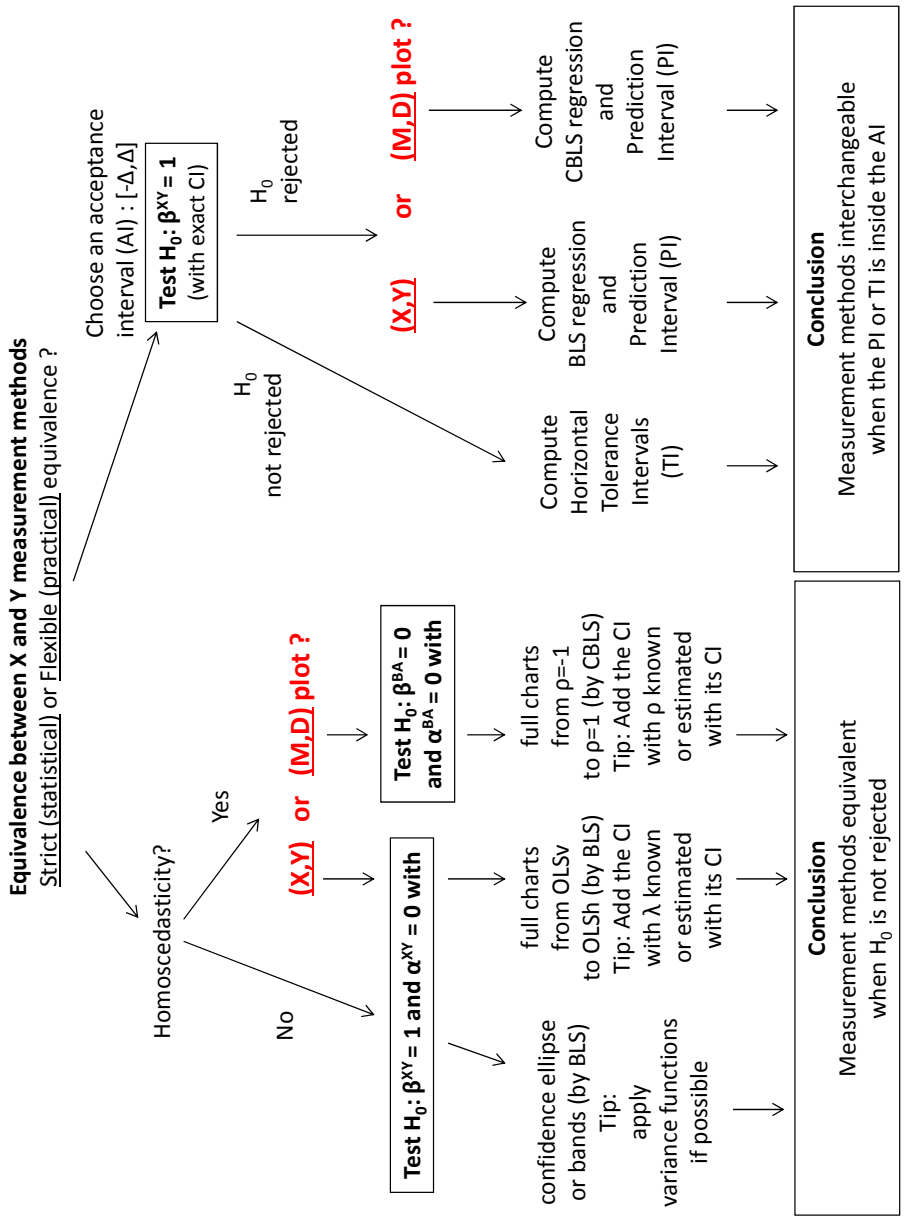

Figure 5.1: *synthetic guidelines*

Bibliography

[1] J. O. Westgard and M. R. Hunt, "Use and interpretation of common statistical tests in method-comparison studies," *Clinical Chemistry*, vol. 19, no. 1, pp. 49–57, 1973.

[2] K. Van Uytfanghe, D. Stöckl, J. M. Kaufman, T. Fiers, H. A. Ross, A. P. De Leenheer, and L. M. Thienpont, "Evaluation of a candidate reference measurement procedure for serum free testosterone based on ultrafiltration and isotope dilution-gas chromatography-mass spectrometry," *Clinical Chemistry*, vol. 50, no. 11, pp. 2101–2110, 2004.

[3] S.-E. Bäck, C. G. M. Magnusson, L. K. Norlund, H. H. von Schenck, M. E. Menschik, and P. E. S. Lindberg, "Multiple-site analytic evaluation of a new portable analyzer, HemoCue hb 201+, for Point-of-Care testing," *Point of Care*, vol. 3, no. 2, pp. 60–65, 2004.

[4] L. I.-K. Lin, "A concordance correlation coefficient to evaluate reproducibility," *Biometrics*, vol. 45, pp. 255–268, 1989.

[5] L. I.-K. Lin, "A note on the concordance correlation coefficient," *Biometrics*, vol. 56, pp. 324–325, 2000.

[6] R. T. St. Laurent, "Evaluating agreement with a gold standard in method comparison studies," *Biometrics*, vol. 54, no. 2, pp. 537–545, 1998.

[7] D. V. Lindley, "Regression lines and the linear functional relationship," *Journal of the Royal Statistical Society*, vol. 9, no. 2, pp. 218–244, 1947.

[8] A. Martínez, F. J. del Río, J. Riu, and F. X. Rius, "Detecting proportional and constant bias in method comparison studies by using linear regression with errors in both axes," *Chemometrics and Intelligent Laboratory Systems*, vol. 49, pp. 181–195, 1999.

[9] B. G. Francq and B. B. Govaerts, "Measurement methods comparison with errors-in-variables regressions. from olsv to olsh, review and new perspectives," *Chemometrics and Intelligent Laboratory Systems*, 2013. Submitted.

[10] B. G. Francq and B. B. Govaerts, "Hyperbolic confidence bands of errors-in-variables regression lines applied to method comparison studies," *Journal de la Soçiété Française de Statistique*, 2013 (submitted and accepted).

[11] J. M. Bland and D. G. Altman, "Measuring agreement in method comparison studies," *Statistical Methods in Medical Research*, vol. 8, pp. 135–160, 1999.

[12] D. G. Altman and J. M. Bland, "Measurement in medicine: the analysis of method comparison studies," *The Statistician*, vol. 32, pp. 307–317, 1983.

[13] J. M. Bland and D. G. Altman, "Statistical methods for assessing agreement between two methods of clinical measurement," *The Lancet*, vol. i, pp. 307–310, 1986.

[14] June 2013. http://en.wikipedia.org/wiki/Bland%E2%80%93Altman_plot.

[15] B. Francq and B. Govaerts, "How to accept the equivalence of two measurement methods? comparison and improvements of the bland and altman's approach and errors-in-variables regressions," *12th European Symposium on Statistical Methods for the Food Industry*, pp. 281–290, 2012.

[16] B. D. Ripley and M. Thompson, "Regression techniques for the detection of analytical bias," *Analyst*, vol. 112, pp. 377–383, 1987.

[17] M. Galea-Rojas, M. V. de Castilho, H. Bolfarine, and M. de Castro, "Detection of analytical bias," *Analyst*, vol. 128, pp. 1073–1081, 2003.

[18] P.-L. Docquier, L. Paul, R. Menten, O. Cartiaux, B. Francq, and X. Banse, "Measurement of bone cyst fluid volume using k-means clustering," *Magnetic Resonance Imaging*, vol. 27, pp. 1430–1439, 2009.

[19] B. Francq, "Développement d'outils statistiques pour tester l'équivalence entre deux méthodes de mesure analytique," 2008. Mémoire de DEA. Université Catholique de Louvain.

[20] S. K. Hanneman, "Design, analysis, and interpretation of method-comparison studies," *AACN Advanced Critical Care*, vol. 19, no. 2, pp. 223–234, 2008.

[21] A. Madansky, "The fitting of straight lines when both variables are subject to error," *Journal of the American Statistical Association*, vol. 54, no. 285, pp. 173–205, 1959.

[22] V. D. Barnett, "Fitting straight lines-the linear functional relationship with replicated observations," *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, vol. 19, no. 2, pp. 135–144, 1970.

[23] W. A. Fuller, *Measurement Error Models*. New York: Wiley, 1987.

[24] H. Passing and W. Bablok, "A new biometrical procedure for testing the equality of measurements from two different analytical methods. application of linear regression procedures for method comparison studies in clinical chemistry, part i," *Journal of Clinical Chemistry & Clinical Biochemistry*, vol. 21, no. 11, pp. 709–720, 1983.

[25] A. Martínez, J. Riu, and F. X. Rius, "Lack of fit in linear regression considering errors in both axes," *Chemometrics and Intelligent Laboratory Systems*, vol. 54, pp. 61–73, 2000.

[26] D. Stöckl, K. Dewitte, and L. M. Thienpont, "Validity of linear regression in method comparison studies: is it limited by the statistical model or the quality of the analytical input data?," *Clinical Chemistry*, vol. 44, no. 11, pp. 2340–2346, 1998.

[27] J. M. Bland and D. G. Altman, "A note on the use of the intraclass correlation coefficient in the evaluation of agreement between two methods of measurement," *Computers in Biology and Medicine*, vol. 20, no. 5, pp. 337–340, 1990.

[28] C. Y. Tan and B. Iglewicz, "Measurement-methods comparisons and linear statistical relationship," *Technometrics*, vol. 41, no. 3, pp. 192–201, 1999.

[29] T. W. Anderson, "Estimation of linear functional relationships: Approximate distributions and connections with simultaneous equations in econometrics," *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, vol. 38, no. 1, pp. 1–36, 1976.

[30] A. Wald, "The fitting of straight lines if both variables are subject to error," *The Annals of Mathematical Statistics*, vol. 11, no. 3, pp. 284–300, 1940.

[31] J. Riu and F. X. Rius, "Univariate regression models with errors in both axes," *Journal of Chemometrics*, vol. 9, pp. 343–362, 1995.

[32] M. S. Bartlett, "Fitting a straight line when both variables are subject to error," *Biometrics*, vol. 5, no. 3, pp. 207–212, 1949.

[33] S. D. Hodges and P. G. Moore, "Data uncertainties and least squares regression," *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, vol. 21, no. 2, pp. 185–195, 1972.

[34] M. Halperin, "Fitting of straight lines and prediction when both variables are subject to error," *Journal of the American Statistical Association*, vol. 56, no. 295, pp. 657–669, 1961.

[35] M. Y. Wong, "Likelihood estimation of a simple linear regression model when both variables have error," *Biometrika*, vol. 76, no. 1, pp. 141–148, 1989.

[36] J. W. Gillard and T. C. Iles, "An historical overview of linear regression with errors in both variables," tech. rep., Cardiff University School of Mathematics, 2006.

[37] C. F. Gauss, *Theoria Motus Corporum Coelestium in sectionibus conicis solem ambientium.* 1809.

[38] A.-M. Legendre, *Nouvelles méthodes pour la détermination des orbites des comètes. Appendice sur la méthodes des moindres carrés.* Paris: Firmin-Didot, 1805.

[39] P. J. Cornbleet and N. Gochman, "Incorrect least-squares regression coefficients in method-comparison analysis," *Clinical Chemistry*, vol. 25, no. 3, pp. 432–438, 1979.

[40] N. R. Draper and H. Smith, *Applied Regression Analysis.* Wiley, 3 ed., 1981.

[41] C. Hartmann, J. Smeyers-Verbeke, W. Penninckx, and D. L. Massart, "Detection of bias in method comparison by regression analysis," *Analytica Chimica Acta*, vol. 338, pp. 19–40, 1997.

[42] T. A. Jones, "Fitting straight lines when both variables are subject to error. i. maximum likelihood and least-squares estimation," *Mathematical Geology*, vol. 11, no. 1, pp. 1–25, 1979.

[43] P. Sprent and G. R. Dolby, "The geometric mean functional relationship," *Biometrics*, vol. 36, no. 3, pp. 547–550, 1980.

[44] P. Dagnelie, *Statistique théorique et appliquée. Tome2. Inférence statistique à une et à deux dimensions.* Bruxelles: De Boeck, 2 ed., 2011.

[45] F. Barker, Y. C. Soh, and R. J. Evans, "Properties of the geometric mean functional relationship," *Biometrics*, vol. 44, no. 1, pp. 279–281, 1988.

[46] H. Passing and W. Bablok, "Comparison of several regression procedures for method comparison studies and determination of sample sizes. application of linear regression procedures for method comparison studies in clinical chemistry, part ii," *Journal of Clinical Chemistry & Clinical Biochemistry*, vol. 22, no. 6, pp. 431–445, 1984.

[47] R. J. Smith, "Use and misuse of the reduced major axis for line-fitting," *American Journal of Physical Anthropology*, vol. 140, pp. 476–486, 2009.

[48] J. Ludbrook, "Linear regression analysis for comparing two measurers or methods of measurement: But which regression?," *Clinical and Experimental Pharmacology and Physiology*, vol. 37, pp. 692–699, 2010.

[49] A. H. Kalantar, R. I. Gelb, and J. S. Alper, "Biases in summary statistics of slopes and intercepts in linear regression with errors in both variables," *Talenta*, vol. 42, no. 4, pp. 597–603, 1995.

[50] G. Casella and R. L. Berger, *Statistical Inference*. Duxbury Press, 2nd ed., 2001.

[51] K. Linnet, "Necessary sample size for method comparison studies based on regression analysis," *Clinical Chemistry*, vol. 45, no. 6, pp. 882–894, 1999.

[52] K. Linnet, "Performance of deming regression analysis in case of misspecified analytical error ratio in method comparison studies," *Clinical Chemistry*, vol. 44, no. 5, pp. 1024–1031, 1998.

[53] K. Linnet, "Estimation of the linear relationship between the measurements of two methods with proportional errors," *Statistics in Medicine*, vol. 9, pp. 1463–1473, 1990.

[54] K. Linnet, "Evaluation of regression procedures for methods comparison studies," *Clinical Chemistry*, vol. 39, no. 3, pp. 424–432, 1993.

[55] J. W. Gillard and T. C. Iles, "Method of moments estimation in linear regression with errors in both variables," tech. rep., Cardiff University School of Mathematics, 2005.

[56] J. Mandel, "Fitting straight lines when both variables are subject to error," *Journal of Quality Technology*, vol. 16, no. 1, pp. 1–14, 1984.

[57] J. M. Lisý, A. Cholvadová, and J. Kutej, "Multiple straight-line least-squares analysis with uncertainties in all variables," *Computers & Chemistry*, vol. 14, no. 3, pp. 189–192, 1990.

[58] F. J. del Río, J. Riu, and F. X. Rius, "Prediction intervals in linear regression taking into account errors on both axes," *Journal of Chemometrics*, vol. 15, pp. 773–788, 2001.

[59] J. Riu and F. X. Rius, "Assessing the accuracy of analytical methods using linear regression with errors in both axes," *Analytical Chemistry*, vol. 68, pp. 1851–1857, 1996.

[60] A. Martínez, J. Riu, and F. X. Rius, "Evaluating bias in method comparison studies using linear regression with errors in both axes," *Journal of Chemometrics*, vol. 16, pp. 41–53, 2002.

[61] P. D. Wentzell, D. T. Andrews, D. C. Hamilton, K. Faber, and B. R. Kowalski, "Maximum likelihood principal component analysis," *Journal of Chemometrics*, vol. 11, pp. 339–366, 1997.

[62] P. D. Wentzell, "Other topics in soft-modeling: Maximum likelihood-based soft-modeling methods," in *Comprehensive Chemometrics* (S. D. Brown, R. Tauler, and B. Walczak, eds.), vol. 2, pp. 507–558, Amsterdam: Elsevier, 2009.

[63] D. York, N. M. Evensen, M. L. Martinez, and J. D. B. Delgado, "Unified equations for the slope, intercept, and standard errors of the best straight line," *American Journal of Physics*, vol. 72, no. 3, pp. 367–375, 2004.

[64] M. Schuermans, I. Markovsky, P. D. Wentzell, and S. Van Huffel, "On the equivalence between total least squares and maximum likelihood PCA," *Analytica Chimica Acta*, vol. 544, pp. 254–267, 2005.

[65] D. T. Andrews, L. Chen, P. D. Wentzell, and D. C. Hamilton, "Comments on the relationship between principal components analysis and weighted linear regression for bivariate data sets," *Chemometrics and Intelligent Laboratory Systems*, vol. 34, pp. 231–244, 1996.

[66] R. R. Sokal and J. F. Rohlf, *Biometry: the principles and practice of statistics in biological research.* New York: W. H. Freeman and Co., 4 ed., 2012.

[67] T. W. Anderson and T. Sawa, "Exact and approximate distributions of the maximum likelihood estimator of a slope coefficient," *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, vol. 44, no. 1, pp. 55–62, 1982.

[68] M. A. Creasy, "Confidence limits for the gradient in the linear functional relationship," *Journal of the Royal Statistical Society. Series B (Statistical Methodology)*, vol. 18, no. 1, pp. 65–69, 1956.

[69] J. W. Gillard and T. C. Iles, "Variance covariance matrices for linear regression with errors in both variables," tech. rep., Cardiff University School of Mathematics, 2006.

[70] F. A. Graybill and T. L. Connell, "Sample size required to estimate the ratio of variances with bounded relative error," *Journal of the American Statistical Association*, vol. 58, no. 304, pp. 1044–1047, 1963.

[71] S. Reiter, "Estimates of bounded relative error for the ratio of variances of normal distributions," *Journal of the American Statistical Association*, vol. 51, no. 275, pp. 481–488, 1956.

[72] A. Goldman, "Sample size for a specified width confidence interval on the ratio of variances from two independent normal populations," *Biometrics*, vol. 19, no. 3, pp. 465–477, 1963.

[73] J. G. Booth and P. Hall, "Bootstrap confidence regions for functional relationships in errors-in-variables models," *The Annals of Statistics*, vol. 21, no. 4, pp. 1780–1791, 1993.

[74] W. Liu, S. Lin, and W. W. Piegorsch, "Construction of exact simultaneous confidence bands for a simple linear regression model," *International Statistical Review*, vol. 76, no. 1, pp. 39–57, 2008.

[75] W. Liu, *Simultaneous Inference in Regression.* CRC Press, 1 ed., 2010.

[76] H. Working and H. Hoteling, "Applications of the theory of error to the interpreatation of trends," *Journal of the American Statistical Association*, vol. 24, pp. 202–217, 1929.

[77] R. F. Martin, "General deming regression for estimating systematic bias and its confidence interval in method-comparison studies," *Clinnical Chemistry*, vol. 46, no. 1, pp. 100–104, 2000.

[78] M. Fekri and A. Ruiz-Gazen, "Robust weighted orthogonal regression in the errors-in-variables model," *Journal of Multivariate Analysis*, vol. 88, no. 1, pp. 89–108, 2004.

[79] R. Zamar, "Robust estimation in the errors in variables model," *Biometrika*, vol. 76, pp. 149–160, 1989.

[80] R. Maronna, "Principal components and orthogonal regression based on robust scales," *Technometrics*, vol. 47, pp. 264–273, 2005.

[81] M. Hubert, P. J. Rousseeuw, and S. Van Aelst, "High-breakdown robust multivariate methods," *Statistical Science*, vol. 23, no. 1, pp. 92–119, 2008.

[82] J. Ludbrook, "Confidence in altman-bland plots: A critical review of the method of differences," *Clinical and Experimental Pharmacology and Physiology*, vol. 37, pp. 143–149, 2010.

[83] A. Wald and J. Wolfowitz, "Tolerance limits for a normal distribution," *Annals of Mathematical Statistics*, vol. 17, no. 2, pp. 208–215, 1946.

[84] A. Wald, "Setting of tolerance limits when the sample is large," *Annals of Mathematical Statistics*, vol. 13, pp. 389–399, 1942.

[85] A. Wald, "An extension of wilk's method for setting tolerance limits," *Annals of Mathematical Statistics*, vol. 14, pp. 45–55, 1943.

[86] S. S. Wilks, "Statistical prediction with special reference to the problem of tolerance limits," *Annals of Mathematical Statistics*, vol. 13, pp. 400–409, 1942.

[87] W. G. Howe, "Two-sided tolerance limits for normal populations, some improvements," *Journal of the American Statistical Association*, vol. 64, no. 326, pp. 610–620, 1969.

[88] C. Lentner, K. Diem, and J. Seldrup, *Geigy scientific tables: Introduction to statistics, statistical tables, mathematical formulae*, vol. 2. Ciba-Geigy, 8 ed., 1982.

[89] F. A. Graybill and C.-M. Wang, "Confidence intervals on nonnegative linear combinations of variances," *Journal of the American Statistical Association*, vol. 75, no. 372, pp. 869–873, 1980.

[90] J. Ludbrook, "Comparing methods of measurement," *Clinical and Experimental Pharmacology and Physiology*, vol. 24, pp. 193–203, 1997.

[91] J. Ludbrook, "Statistical techniques for comparing measurers and methods of measurement: a critical review," *Clinical and Experimental Pharmacology and Physiology*, vol. 29, pp. 527–536, 2002.

[92] B. Cartsensen. http://BendixCarstensen.com/MethComp/Ancona.2011.

[93] G. B. Schaalje and R. A. Butts, "Some effects of ignoring correlated measurement errors in straight line regression and prediction," *Biometrics*, vol. 49, no. 4, pp. 1262–1267, 1993.

[94] M. de Castro, M. Galea-Rojas, H. Bolfarine, and M. V. de Castilho, "Detection of analytical bias when comparing two or more measuring methods," *Journal of Chemometrics*, vol. 18, pp. 431–440, 2004.