MASTER THESIS

Ekaterina Manuilova

Statistical methods for comparison of two inaccurate measurement procedures in experiments with measurement replications

Department of Statistics

University of Munic

Field of study: Biostatistics

Supervision: Prof.Dr.Matthias Schmidt

In cooperation with

Roche Diagnostics GmbH, Penzberg

Supervision: Dr.Fabian Model

April 25th, 2013

Declaration:

Herewith I declare that this thesis is the result of my independent work.

All sources and auxiliary materials used by me in this thesis are cited completely.

Contents

Abstract

In method comparison experiments with replications, two measurement procedures (test method and comparative method) are compared using regression analysis. The ordinary linear regression approach often yields biased estimates of regression coefficients because the predictor (measurements of comparative method) contains random error. Instead, a linear error-in-variables regression model should be used. This work focuses on the approaches based on linear structural error-in-variables model, called averaged Deming regression, and Linnet's weighted Deming regression as they are recommended by CLSI (Clinical and Laboratory Standards Institute) for the analysis of method comparison data. Theoretical foundations of these approaches are analyzed and compared with ML estimations for linear structural error-in-variables model and the performance of both methods is investigated using Monte-Carlo simulations. The performance of different methods of calculation of confidence intervals for the estimates of regression coefficients is also examined using Monte-Carlo simulations. Simulation studies show that the averaged Deming regression and Linnet's weighted Deming regression approaches combined with jackknife or bootstrap quantile confidence intervals yield reasonable estimations of regression coefficients, even if the model assumptions are considerably violated. The case of matrix effect in the test measurement procedure is also considered. Simulation studies suggest biased estimates of regression parameters calculated via averaged Deming regression or Linnet's weighted Deming regression, if there is a strong matrix effect in the test method. In that case, another approach based on the linear structural ME model with error in equation may be used for the estimation of regression coefficients. The simulation studies show proper results of this method if there is a strong matrix effect in the test method.

Chapter 1

Introduction

1.1   Method comparison study

The present master thesis deals with the problem of method comparison studies in cases when there are replicated measurements. This is one of the most common techniques used by both manufacturers and clinical laboratory users to measure the performance of in vitro diagnostic (IVD) measurement procedures. The aim of a method comparison study is to analyze bias between two measurement procedures using paired results from patient samples.

The main guideline document regulating measurement comparison study is called EP9: "Measurement Procedure Comparison and Average Bias Estimation Using Patient Samples" [18]. It was developed by CLSI (Clinical and Laboratory Standards Institute). Realization and analysis of method comparison studies should adhere to recommendations of this document as strictly as possible.

The CLSI EP9 [18] lists the following primary purposes for measurement procedure comparisons:

1. **Establishing the relationship between measurement procedures by the manufacturer.** Manufacturers must establish the relationship of any candidate measurement procedure quantification with a comparative measurement procedure. The primary goal of such study is to analyze average bias between two methods.

2. **Claim Verification by Manufacturer.** The goal of such study is to show that the candidate measurement procedure meets the claims already established in an in-use situation.

3. **Measurement Procedure introduction to the clinical laboratory.** Clinical laboratories typically perform measurement procedure comparison studies when introducing an IVD product in their menu. The candidate measurement procedure typically replaces one that is currently used in the laboratory.

In following we consider only the first purpose of a measurement comparison study.

1.2 Scopes

The main purposes of this work are

- to find theoretical foundations of methods for analysis of method comparison data, which are proposed by CLSI EP9 [18].

- to implement the methods in the statistical programming language R.

- to evaluate the reliability of these methods using simulations.

- to evaluate performance of different methods for calculation of confidence intervals for the estimated parameters of interest.

- to develop methodical recommendations for the analysis of method comparison data in case model assumptions of methods suggested by CLSI EP9 are violated.

1.3 Outline

The master thesis is organized as follows. Chapter 2 deals with requirements concerning design and analysis of measurement comparison studies. In chapter 3 theoretical foundations for analysis of method comparison data are investigated and different methods of analysis are described. Great attention has been focused on the modifications of Deming and weighted Deming regression in case of data with replications. In chapter 4 reliability and performance of the considered methods of analysis are investigated using Monte-Carlo simulations. In chapter 5, two examples of method comparison data are considered. The real world data are used to illustrate the analysis methods from the previous chapters.

Chapter 2

Design of measurement comparison studies

In this section we consider briefly the recommendations of CLSI EP9 for design and analysis of method comparison studies.

2.1 Requirements concerning the comparative measurement procedure

Ideally, the comparative measurement procedure should be a reference measurement procedure. In this case the desired result of the comparison is no significant average bias between the two procedures. However, often a new (candidate) measurement procedure is developed as an improvement over a comparative measurement procedure. In such situation the primary goal is to analyze average bias between them. Generally an establishment study is successful, if the estimated bias is within a predetermined acceptance criterion. The measurements of the comparative measurement procedure should ideally have the following characteristics:

- lower uncertainty than the measurements of the candidate measurement procedure.

- be free of interferences (matrix effect) if possible.

- use the same units as a candidate measurement procedure,

- have known bias relative to standards or reference measurement procedures, if possible.

Note: More about matrix effect can be found in section 3.4.

2.2   Requirements concerning samples

According to CLSI EP9, it is recommended that at least 100 patient samples with measurand values spanning the common measuring interval of the two measuring procedures are used and that as many influential factors as possible are included in the experimental design. Such factors may include replicate, calibration, run, day, reagent, lot etc. It is generally expected that such factors will not create a systematic shift in bias but only affect the variability of the bias estimate. An experimental design that increases the replication over such factors will decrease the bias estimate uncertainty and thus the confidence interval for the estimate of bias. It is not necessary, that the study design provides enough measurements to truly measure the contribution of each factor to the variability of results. That is not the purpose of the measurement comparison study. Instead, its aim is to provide a robust estimate of bias by sampling all significant sources of variability.

2.3   Requirements concerning measurement replicates

Where possible, two or more replicates of both candidate and comparative measurement procedure should be obtained. CLSI EP9 recommends to use the matched sample-to-sample averaged replications for estimation of average bias performance. The underlying assumption behind of this averaging of replications is the attempt to measure the same unchanging quantity. Then the averaging of replications will reduce the uncertainty of the measurements. So we can conclude that the measurement replications should be exchangeable.

2.4   Recommendations for the analysis of the outcome of measurement comparison study.

CLSI EP9 requires manufacturers to conduct establishment or validation studies via regression analysis. The initial goal of regression analysis in a measurement comparison study is to fit a straight line through the data presented as an X-Y plot with the comparative procedure on the x-axis and the candidate measurement procedure on the y-axis: $y = a + b \cdot x$. The slope b models systematical proportional bias and the intercept gives the systematical additive bias between two methods. We observe $a = 0$ and $b = 1$ if there is no systematical bias between measurement procedures. If a and b are known, the systematical bias between two methods at some medical decision point X_c can be calculated as

$$bias(X_c) = a + b \cdot X_c - X_c = a + (b - 1) \cdot X_c. \tag{2.1}$$

The ordinary linear regression approach (OLS) is not suitable for estimation of regression coefficients a and b because the predictor X (comparative procedure measurement) is subject to

4

measurement error. In such situation one can use errors-in-variables models to estimate slope and intercept.

Figure 2.1 shows an example of simulated data with a true slope of 1 and a true intercept of 0. We can see that the OLS regression fit with intercept of 17.91 and slope of 0.91 is considerably biased compared to the Deming regression fit (intercept of -2.91, slope of 1.01), which allows for measurement error in the predictor variable (Method 1).

Note: We can use OLS regression for the analysis of method comparison data if the measurement error of the comparative method is negligible. In this case, the measurements can be considered to be error free. Such a method is called "gold standard" [3, p. 203].

Linear Regression Fit

Deming Regression
−2.91 + 1.01 * Method1
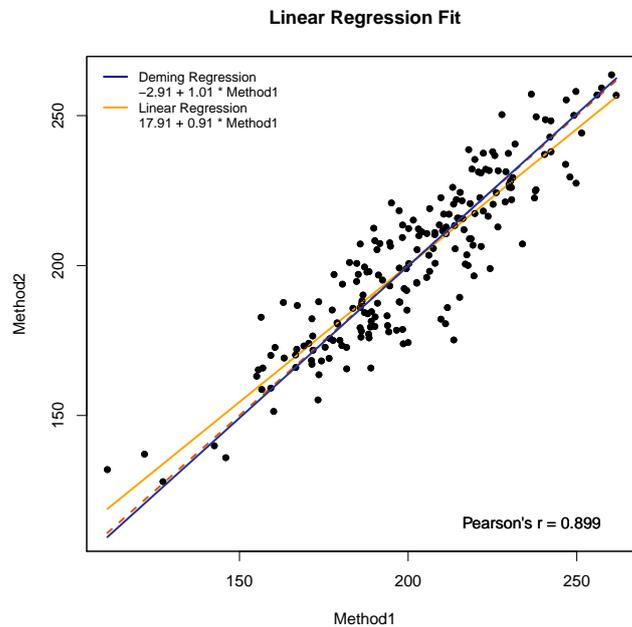Linear Regression
17.91 + 0.91 * Method1

Pearson's r = 0.899

Figure 2.1: Example of simulated data with errors in both variables. The "true" concentrations are normally distributed with mean 200 and standard deviation 25. The both simulated measurements have normally distributed errors with mean of 0 and standard deviation of 5. The true values of the data are slope of 1 and intercept of 0. The Deming regression fit (blue line) is close to the identity line (red dotted line), the OLS regression fit (orange line) is strongly biased.

The choice of an appropriate error-in-variables regression method depends on the sample data. Generally three situations are possible:

1. **Constant standard deviation.** In this case the variation of differences between the two methods does not change along the whole measurement range. An equivalent statement: the variance of the measurement errors of both measurement procedures stays constant within the whole measurement range. For such data the CLSI EP9 [18] suggests the Deming regression

approach (look at section 3.2.2).

2. **Constant coefficient of variation (CV).** In this case, the variability of measurement errors of both measurement procedures is proportional to the measurement level. The method recommended for the analysis of this situation is Linnet's weighted Deming regression (look at section 3.3.4.

3. An other form of heteroscedasticity or presence of outliers. In this case, the CLSI EP9 [18] suggests a nonparametric approach: Passing-Bablok regression (which is not subject of this work).

For regression coefficients and estimates of systematical bias at some medical decision point X_c, confidence intervals (typically 95%) should be computed. The analytical formulas to compute CIs are given only for unweighted Deming regression. For other regression types, resampling techniques such as jackknife or bootstrap should be used.

Chapter 3

Theoretical part

Measurement error models (ME) are a generalization of standard regression models. For the simplest ME model, the goal is to estimate a straight line fit between the two variables, both of which are measured with error. One can find several solutions for this problem: method of grouping, using of instrumental variables, using of variance components in replicated cases, Berkson model, estimation via cumulants, etc. (an overview of these methods can be found for example at Madansky [18]).

Following the recommendations of CLSI EP9, we concentrate on structural error-in-variables models and corresponding maximum likelihood estimation of regression coefficients under the assumption that the ratio of variances of measurement errors of both variables is known. It is the most common approach used for regression analysis of bivariate data in biometry, known as "Deming regression" [7]. Then, analogously linear structural error-in-variables model with replications and corresponding ML estimations are considered. Connection between ML-estimations and estimations of averaged- and Linnet's weighted Deming regression are assessed. The last two methods are recommended by the CLSI EP9 document for the analysis of method comparison data with replications.

The theoretical part is organized as follows:

1. First, general methods of parameter estimation, mentioned in this work (maximum likelihood method and method of moments) are described (section 3.1). The description of the method of moments is based on the book of Lagutin [12], for the description of the maximum likelihood method the book of Held [4] is used.

2. In the second section (3.2), we consider linear structural models without replications and concentrate on the Deming regression approach. The main source of information used here is the book of Cheng and Ness [1] and the book of Miller [3].

   (a) In subsection 3.2.1, the linear structural ME model is introduced.

(b) In subsection 3.2.2, the Deming regression approach is described. ML estimations, their important properties and analytical formulas for calculation of variances of estimations (derived by Strike, cited by Looney, [13] and Kelly [11]) are given.

3. The third section (3.3) deals with the linear structural models with replications, which are closely connected to the same models without replications described in the previous section.

   (a) In subsection 3.3.1, the general model formulation is introduced and ML estimations of parameters of this model are given, derived from Chan and Mak [6].

   (b) In subsection 3.3.2, averaged Deming regression is described. This method can be deduced from the linear structural model with replications. The sources used for the description are papers of Chan and Mak [6] and the book of Miller [3].

   (c) The subsection (3.3.3) deals with the weighted version of averaged Deming regression, which is recommended for the analysis of heteroscedastic data with constant CV. In the following, we call this approach Linnet's weighted Deming regression, because it was developed by Linnet and introduced in his papers [7] and [8].

   (d) The last subsection (3.3.4) describes the approach of Barnett [15]. He derived ML estimation for the linear functional model with replications, assuming that measurement errors for each item can have different variances. This approach is not mentioned in CLSI EP9.

4. The forth section (3.4) deals with modeling of matrix effects in test method with the help of linear structural ME model with error in equation.

   (a) First (subsection 3.4.1), the term of matrix effect is briefly explained.

   (b) In subsection 3.4.2, the linear structural ME model with error in equation is formulated, based on paper of Dunn et.al. [10] and book of Fuller [2].

   (c) In subsection 3.4.3, MM estimations of parameters of the linear structural ME model with error in equation for the replicated case are given, which were derived from Oman et.al. [9].

5. In section 3.5 different diagnostic plots are described, which can support correct model choice. In this section ideas of Bland and Altman [14], Linnet [7], Oman et.al. [9], Carstensen [16] and Fahrmeir [15] are introduced.

6. In section 3.6 some relevant nonparametric and semiparametric methods of calculation of confidence intervals of parameter estimations are considered. The sources which were used here, are the book of Efron and Tibshirany [5] and papers of Linnet [7] and Oman et.al. [9].

(a) In subsection 3.6.1, calculation of variance of estimations via jackknife method and parametric and nonparametric bootstrap are discussed.

(b) In subsection 3.6.2, bootstrap quantile and BC_a methods are described.

7. In the last section (3.7), conclusions are formulated.

3.1 Methods of estimation

In this section, we consider briefly the methods which are used for derivation of parameter estimations. They are

1. method of moments,

2. maximum likelihood estimation.

The description of the method of moments can be found at Lagutin [12]. For the description of the maximum likelihood method the book of Held [4] was used.

3.1.1 Method of moments

The k-th moment of a distribution of any random variable X is a quantity

$$\alpha_k = E(X^k).$$

Not all distributions have finite moments. It is easy to show that if $E(|X|^m) < \infty$ then all moments $\{\alpha_k \mid k = 1, 2, ..., m\}$ of the distribution of X exist.

We consider a random sample of independent and identically distributed variables $\{X_1, X_2, ..., X_n\}$. Assume that $A_k = \frac{1}{n}\sum_{i=1}^{n} X_i^k$. If the moment α_k exists then, according to the law of large numbers

$$n \to \infty \ \Rightarrow \ A_k \xrightarrow{P} \alpha_k.$$

Suppose that we observe a realization $\{x_1, x_2, ..., x_n\}$ of $\{X_1, X_2, ..., X_n\}$. If n is large enough we can assume that

$$a_k = \frac{1}{n}\sum_{i=1}^{n} x_i^k \approx \alpha_k,$$

which means that each k-th empirical moment $a_k, k = 1, 2, ...m$ is close to the corresponding k-th theoretical moment α_k. This reason is the foundation of the method of moments.

Assume that the distribution of X has m unknown parameters $\theta_1, \theta_2, ...\theta_m$, where $\theta = (\theta_1, \theta_2, ...\theta_m) \in \Theta \subseteq R^m$ and

$$\forall \theta \in \Theta \quad E_\theta(|X|^m) < \infty.$$

Then, all theoretical moments $\{\alpha_k \mid k = 1, 2, ..., m\}$ of the distribution of X exist. Thus we get a system of m equations

$$\alpha_k(\theta) = A_k, \quad k = 1, 2, ...m. \tag{3.1}$$

If this system has a unique solution $\hat{\theta} = (\hat{\theta}_1, \hat{\theta}_2, ...\hat{\theta}_m)$ continuously dependent of $\{\alpha_k \mid k = 1, 2, ...m\}$, then the components $\hat{\theta}_i$ of $\hat{\theta}$ are called estimations of method of moments (MM estimations).

MM estimations have the following properties [12, p.112]:

- they are consistent,

- if the dependence between $\hat{\theta}$ and $\{\alpha_k \,|\, k = 1, 2, ...m\}$ is smooth, then the MM estimations are asymptotically normal distributed.

  In particular

$$\sqrt{n}(A_k - \alpha_k) \xrightarrow{d} \xi \sim N(0, \alpha_{2k} - \alpha_k^2).$$

The MM estimations are easy to derive analytically in most cases but their asymptotic variance is rather high, that is why the maximum likelihood method, which yields estimations with the smallest possible asymptotic variance, is often preferred.

3.1.2 Maximum likelihood method

Suppose we observe a realization $x = \{x_1, x_2, ..., x_n\}$ of a random sample $X = \{X_1, X_2, ..., X_n\}$ with all random variables $X_1, X_2, ..., X_n$ independently and identically distributed. The density function $f(\cdot; \theta^*)$ of X_i belongs to a certain family of distributions $\{f(\cdot, \theta), \theta \in \Theta \subset R^m\}$. The values of the components of vector $\theta^* = \{\theta_1^*, \theta_2^*, ..., \theta_m^*\}$ are unknown. We would like to estimate them. The idea of the maximum likelihood method is to find such $\hat{\theta}^*$ for which the probability to observe data $x = \{x_1, x_2, ..., x_n\}$ is maximal.

The sample $X_1, X_2, ..., X_n$ have the following joint distribution:

$$f(X, \theta) = \prod_{i=1}^{n} f(X_i; \theta).$$

Replacing X_i with observations x_i we get a function of θ:

$$L(\theta) = \prod_{i=1}^{n} f(x_i; \theta) \tag{3.2}$$

which is called likelihood function. The corresponding log-likelihood function is defined as

$$l(\theta) = \sum_{i=1}^{n} \log f(x_i; \theta). \tag{3.3}$$

The maximum likelihood estimation $\hat{\theta}_{ML}^*$ of parameter vector θ^* can be derived via maximization of the likelihood or the log-likelihood function:

$$\hat{\theta}_{ML}^* = \arg\max_{\theta \in \Theta} L(\theta) = \arg\max_{\theta \in \Theta} l(\theta). \tag{3.4}$$

Therefore we should find a solution of a system of m equations:

$$\frac{\partial l(\theta)}{\partial \theta_i} = 0, \quad i = 1, ..., m. \tag{3.5}$$

Given some regularity conditions (see, for example [4]) the system 3.5 has a solution with probability 1. This solution has the following properties [4, p. 81].

1. The ML estimation is strong consistent:

$$P\left(\lim_{n\to\infty}\hat{\theta}^*_{ML}(X_1, X_2, ..., X_n) = \theta^*\right) = 1.$$

2. The ML estimation is asymptotically normal distributed [Held, s. 81]:

$$\hat{\theta}^*_{ML} \sim^{approx} N(\theta^*, \left[J(\hat{\theta}^*_{ML})\right]^{-1}), \tag{3.6}$$

where $J(\theta)$ is expected Fisher information

$$J(\theta) = E(I(\theta)),$$
$$I(\theta) = -\frac{\partial^2 l(\theta)}{\partial\theta}.$$

Often it is very difficult or even impossible to solve 3.5 analytically. In this case, some iterative procedure, for example the Newton-Raphson-algorithm or the expectation-maximization algorithm, can be used to obtain the solution.

3.2 Linear structural model without replications

In this section, we first introduce the linear structural model. Then, the Deming regression approach is described, the derivation of the corresponding ML-estimations is given and the most important properties of this estimations are discussed. Further, some analytical formulas for calculation of confidence intervals for estimations are given.

3.2.1 Model formulation

The standard regression model with one explanatory variable, is given by

$$y = \beta_0 + \beta_1 \cdot x + \epsilon, \tag{3.7}$$

where the independent variable x is fixed and the error ϵ is normally distributed with mean zero and is uncorrelated with ξ. Given a set of independent observations,

$$(x_1, y_1), \ldots (x_n, y_n)$$

the unknown intercept β_0, and slope β_1 are estimated using least squares techniques.

The corresponding standard linear ME model assumes that the variables x and y are observed with measurement errors and the true measurements ξ and η of x and y are unobservable. Formally, we can express this as follows:

$$x = \xi + \delta \quad and \quad y = \eta + \epsilon, \tag{3.8}$$

where the measurement errors δ and ϵ are uncorrelated with each other and with the true measurements.

The true measurements ξ and η are related by

$$\eta = \beta_0 + \beta_1 \cdot \xi. \tag{3.9}$$

The unknown intercept β_0, and slope β_1 should be estimated using observable data

$$(x_1, y_1), \ldots (x_n, y_n).$$

The intercept β_0 is modeling the additive systematic bias between two measurement procedures and the slope β_1 is modeling the proportional systematic bias. The assumption about linear relationship between the true measurements should be verified using a special linearity test, which we do not consider here.

For the sample size n, the linear univariate ME model can be formulated as follows [1, p.14].

$$\eta_i = \beta_0 + \beta_1 \cdot \xi_i, \quad i = 1, 2, \ldots n.$$

$$x_i = \xi_i + \delta_i \quad and \quad y_i = \eta_i + \epsilon_i, \quad i = 1, 2, \ldots n.$$

The measurement errors are assumed to be uncorrelated and have a mean of zero:

$$\forall i \quad E(\delta_i) = E(\epsilon_i) = 0,$$

$$\forall i \quad var(\delta_i) = \sigma_\delta^2, \quad var(\epsilon_i) = \sigma_\epsilon^2,$$

$$\forall i \neq j \quad cov(\delta_i, \delta_j) = cov(\epsilon_i, \epsilon_j) = 0,$$

$$\forall i, j \quad cov(\delta_i, \epsilon_j) = 0.$$

The measurement errors are assumed to be normally distributed in most cases.

There are three separate models depending on the assumption about ξ:

- the **functional model** considers ξ_i as unknown fixed value,

- the **structural model** considers ξ_is as identically distributed random variables with $E(\xi_i) = \mu$ and $var(\xi_i) = \sigma^2$,

- the **ultrastructural model** assumes that ξ_i are independent random but not identically distributed variables, having possibly different means μ_i and variance σ^2. Obviously, the ultrastructural model is a generalization of the functional and structural models.

Model 3.7 differs substantially from 3.8-3.9. Trying to write 3.8-3.9 in terms of an ordinary regression model one obtains

$$y = \beta_0 + \beta_1 \cdot x + (\epsilon - \beta_1 \delta) \equiv \beta_0 + \beta_1 \cdot x + \zeta. \tag{3.10}$$

In the ME model the predictor x is a random variable which is correlated with the error term ζ: $cov(x, \zeta) = -\beta_1 \sigma_\delta^2$. This covariance is zero only if $\sigma_\delta^2 = 0$ which is equal to model 3.7. Using an ordinary regression model if σ_δ^2 is considerably higher than zero yields biased estimation of the regression parameters.

An important problem which arises in ME models is identifiability. Formally, if $\mathbf{Z}$ is a random vector whose distribution is from some family $\Im = \{F_\theta; \quad \boldsymbol{\theta} \in \Theta\}$, then the parameter θ_i, the ith component of the vector $\boldsymbol{\theta}$, is identifiable if and only if all its components are identifiable. The model is said to be identifiable if $\boldsymbol{\theta}$ is **identifiable** [1, p.5].

It is common to assume that all the random variables in the ME model are jointly normal. In this case, the structural model 3.9-3.8 is not identifiable. [1, p.5]

To make ME identifiable one should formulate some additional assumption about model parameters. It will often be assumed that the ratio of the error variances $\lambda \equiv \sigma_\epsilon^2/\sigma_\delta^2$, is known. This model is normally called **Deming regression**.

3.2.2 Deming regression approach

Deriving ML estimations of parameters of Deming regression

Equations 3.8-3.9 and the assumption that $\xi_i \sim N\left(\mu, \sigma^2\right)$ give :

$$E(x) = E(\xi) = \mu,$$

$$E(y) = E(\eta) = \beta_0 + \beta_1 \mu,$$

$$var(x) = \sigma_x^2 = var(\xi) + \sigma_\delta^2 = \sigma^2 + \sigma_\delta^2,$$

$$var(y) = \sigma_y^2 = var(\eta) + \sigma_\epsilon^2 = \beta_1^2 \sigma^2 + \sigma_\epsilon^2,$$

$$cov(x, y) = \sigma_{xy} = cov(\xi, \eta) = \beta_1 \sigma^2$$

The ML estimates for the corresponding parameters in the bivariate normal distribution are [1, p.15]:

$$\hat{\mu} = \bar{x},$$

$$\hat{\mu}_x = \bar{y},$$

$$\hat{\sigma}_x^2 = s_{xx} = \frac{1}{n} \sum (x_i - \bar{x})^2,$$

$$\hat{\sigma}_y^2 = s_{xx} = \frac{1}{n} \sum (y_i - \bar{y})^2,$$

$$\hat{\sigma}_{xy} = s_{xy} = \frac{1}{n} \sum (x_i - \bar{x})(y_i - \bar{y}).$$

Using the invariance properties of ML estimates gives us

$$\bar{x} = \hat{\mu}, \tag{3.11}$$

$$\bar{y} = \hat{\beta}_0 + \hat{\beta}_1 \hat{\mu}, \tag{3.12}$$

$$s_{xx} = \hat{\sigma}^2 + \hat{\sigma}_\delta^2, \tag{3.13}$$

$$s_{yy} = \hat{\beta}_1^2 \hat{\sigma}^2 + \hat{\sigma}_\epsilon^2, \tag{3.14}$$

$$s_{xy} = \hat{\beta}_1^2 \hat{\sigma}^2 \tag{3.15}$$

From $\hat{\sigma}_\delta^2 \geq 0$, $\hat{\sigma}_\epsilon^2 \geq 0$ and $\sigma^2 > 0$ we obtain the following set of five restrictions [1, p.15]:

$$s_{xx} \geq s_{xy}/\hat{\beta}_1, \tag{3.16}$$

$$s_{yy} \geq \hat{\beta}_1 s_{xy}, \tag{3.17}$$

$$s_{xx} \geq \hat{\sigma}_\delta^2, \tag{3.18}$$

$$s_{yy} \geq \hat{\sigma}_\epsilon^2, \tag{3.19}$$

$$sign\left(s_{xy}\right) = sign\left(\hat{\beta}_1\right) \tag{3.20}$$

Assuming that λ is known from 3.11-3.15 we get the quadratic equation [1, p.16]

$$\hat{\beta}_1^2 s_{xy} + \hat{\beta}_1 \left(\lambda s_{xx} - s_{yy}\right) - \lambda s_{xy} = 0. \tag{3.21}$$

with two roots

$$\hat{\beta}_1 = \frac{s_{yy} - \lambda s_{xx} \pm \sqrt{(s_{yy} - \lambda s_{xx})^2 + 4\lambda s_{xy}^2}}{2s_{xy}} \equiv \frac{U}{2s_{xy}}.$$

From 3.20 follows $U > 0$. Then the ML estimation for $\hat{\beta}_1$ is

$$\hat{\beta}_1 = \frac{s_{yy} - \lambda s_{xx} + \sqrt{(s_{yy} - \lambda s_{xx})^2 + 4\lambda s_{xy}^2}}{2s_{xy}} \tag{3.22}$$

The ML solutions for the remaining parameters can now be derived from 3.11-3.15:

$$\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}, \tag{3.23}$$

$$\hat{\sigma}_\delta^2 = \frac{s_{yy} - 2\hat{\beta}_1 s_{xy} + \hat{\beta}_1^2 s_{xx}}{\lambda + \hat{\beta}_1^2}, \tag{3.24}$$

$$\hat{\sigma}^2 = s_{xy}/\hat{\beta}_1. \tag{3.25}$$

Some important properties of the ML-estimations of Deming regression:

- the ML estimations of β_0 and β_1 for the functional model when λ is known are the same as those of the corresponding linear structural relationship [1, p.23],

- the estimation of regression coefficients via method of moments yields solutions which are equivalent to the ML-estimations. That is why the estimation 3.22 is rather robust against violations of the assumption of normal distribution of measurement errors. However, the estimation can be biased by outliers [3, p.212-213].

- Maximum likelihood solution for Deming regression with $\lambda = 1$ is equal to the orthogonal regression estimation. Orthogonal regression minimizes the sum of squares of the orthogonal distances from the data points to the regression line instead of the sum of squares of the vertical distances, as in standard regression. If $\lambda \neq 1$ we can rescale the data such that we get $\lambda = 1$ using the transformation $\dot{x} = \sqrt{\lambda}x$ [1, p.9]. The squared distances r_i^2 between data points and their predictions can be calculated using Pythagoras' theorem:

$$r_i^2 = \hat{\delta}_i^2 + \hat{\epsilon}_i^2 = \left(\dot{x}_i - \hat{\dot{\xi}}_i\right)^2 + (y_i - \hat{\eta}_i)^2 = \lambda \left(x_i - \hat{\xi}_i\right)^2 + (y_i - \hat{\eta}_i)^2. \tag{3.26}$$

Estimation of variance of the estimations

Different formulas for estimation of variance of the estimations of regression coefficients can be found in literature. Strike proposes following formulas (given in Looney [13]):

$$se\left(\hat{\beta}_1\right) = \sqrt{\frac{\hat{\beta}_1^2 [1 - r^2]}{(n-2)r^2}} \tag{3.27}$$

$$se\left(\hat{\beta}_0\right) = \sqrt{\frac{\left[se\left(\hat{\beta}_1\right)\right]^2 \sum x_i^2}{n}} \tag{3.28}$$

$$cov\left(\hat{\beta}_1, \hat{\beta}_0\right) = -\bar{x}\left[se\left(\hat{\beta}_1\right)\right]^2 \tag{3.29}$$

16

where

$$r^2 = \frac{\sum (x_i - \bar{x})}{\sqrt{\sum (x_i - \bar{x})^2 \sum (y_i - \bar{y})^2}}$$

Other formulas can be found at Kelly [11]. He calculates estimations of variances using the influence functions:

$$se\left(\hat{\beta}_1\right) = \sqrt{\frac{\hat{\beta}_1^{\,2} \left[1 - r^2\right]}{nr^2}} \tag{3.30}$$

$$se\left(\hat{\beta}_0\right) = \frac{1}{n} \left[s_{yy} - 2\hat{\beta}_1 s_{xy} + \hat{\beta}_1^2 s_{xx} + \frac{\sum x_i^2 \hat{\beta}_1^2}{n s_{xy}^2} \left(s_{xx} s_{yy} - s_{xy}^2 \right) \right] \tag{3.31}$$

$$cov\left(\hat{\beta}_1, \hat{\beta}_0\right) = -\bar{x} \left[se\left(\hat{\beta}_1\right) \right]^2 \tag{3.32}$$

In this work we evaluate performance of the both versions of formulas. The both analytical methods of estimation of variances assume normal distributed measurement errors and should not be used if this assumption is violated. In case of violated normality assumptions, resampling methods, discussed in section 3.6, yield more reliable results.

The confidence intervals for the estimations of the regression coefficients and of the bias at some medical decision point X_c can be calculated as follows:

$$\hat{\beta}_0 \pm t_{1-\frac{\alpha}{2};n-2}\, se\left(\hat{\beta}_0\right) \tag{3.33}$$

$$\hat{\beta}_1 \pm t_{1-\frac{\alpha}{2};n-2}\, se\left(\hat{\beta}_1\right) \tag{3.34}$$

$$[\beta_0 + (\beta_1 - 1) \cdot x_c] \pm t_{1-\frac{\alpha}{2};n-2} \cdot \sqrt{se\left(\hat{\beta}_0\right)^2 + se\left(\hat{\beta}_1\right)^2 x_c \left(x_c - 2\bar{x}\right)} \tag{3.35}$$

Alternatively one can use nonparametric or semiparametric methods of calculation of CIs (see section 3.6).

3.3 Linear structural model with replications

In this section, the linear structural model with replications is formulated and derivation of the corresponding ML estimations is briefly described. Then, averaged and Linnet's weighted Deming regression approaches are introduced. The last part of this section deals with Barnett's method of estimation of regression parameters in case the measurement errors can vary among items.

3.3.1 Model formulation

The proper design of a measurement comparison study should provide the following characteristics of the data, which are important for model formulation:

1. The variance of the true measurements of the reference method should be much higher than the variance of the measurement error.

2. The measurand stays unchanged within every replication of measurement procedure.

3. The measurement errors of a measurement procedure are independent and identically distributed for all replications at the same item.

4. The measurement errors of a measurement procedure at the different items are independent.

5. The measurement errors of the test- and comparative method are independent.

6. The comparative method should be free of matrix effect.

First we consider the situation of constant standard deviation of measurement errors. It is equivalent to say, that the variances of the measurement errors of both, test- and comparative methods, do not change within the whole measurement range. As usually we assume the measurement errors to be normally distributed and that neither the comparative method nor the test method contain random matrix effect. The last assumption should be made about the true measurements of the comparative method. We choose the structural model, which treats the true values as independent identically normal distributed random variables. The structural model is advantageous for the calculation of ML estimation for regression coefficients.

The suitable model formulation and derivation of the corresponding ML estimations can be found in Chan and Mak [6]. Assume that ξ and η are true comparative- and test measurement procedures which cannot be obtained without error, δ is the measurement error of the comparative method and ϵ is the measurement error of the test method. Then, the observed measurements are $x = \xi + \delta$ and $y = \eta + \epsilon$. For each (ξ_i, η_i) r repeated observations x_{ij} and y_{ij}, $j = 1, \ldots r$ are obtained. Then, the model can be formulated as follows:

$$\eta_i = \beta_0 + \beta_1 \xi_i \tag{3.36}$$

$$x_{ij} = \xi_i + \delta_{ij}, \quad y_{ij} = \eta_i + \epsilon_{ij} \tag{3.37}$$

$$\xi_i \sim N\left(\mu, \sigma^2\right) \tag{3.38}$$

$$\delta_{ij} \sim N\left(0, \sigma_\delta^2\right), \quad \epsilon_{ij} \sim N\left(0, \sigma_\epsilon^2\right) \tag{3.39}$$

$$\forall i \quad \forall l \neq k \quad cov(\delta_{il}, \delta_{ik}) = cov(\epsilon_{il}, \epsilon_{ik}) = 0, \tag{3.40}$$

$$\forall i \neq j \quad \forall l, k \quad cov(\delta_{il}, \delta_{jk}) = cov(\epsilon_{il}, \epsilon_{jk}) = cov(\delta_{il}, \epsilon_{jk}) = 0. \tag{3.41}$$

Now we can build a **likelihood function** for this model.

Let

$$\mathbf{x_i} = (x_{i1}, x_{i2}, \ldots x_{ir})^T,$$

$$\mathbf{y_i} = (y_{i1}, y_{i2}, \ldots y_{ir})^T,$$

$$\mathbf{z_i}^T = \left(\mathbf{x_i}^T, \mathbf{y_i}^T\right).$$

Then $\mathbf{z_i} \sim N\left(\mathbf{m}, \mathbf{V}\right)$, where

$$\mathbf{m}^T = \left(\mu \mathbf{1}_{r1}^T, (\beta_0 + \beta_1 \mu)\, \mathbf{1}_{r1}^T\right)$$

$$\mathbf{V} = \mathbf{\Sigma} + \sigma^2 \mathbf{b}\mathbf{b}^T, \quad \mathbf{\Sigma} = \left[\begin{array}{cc} \sigma_\delta^2 \mathbf{I}_r & 0 \\ 0 & \sigma_\epsilon^2 \mathbf{I}_r \end{array}\right], \quad \mathbf{b} = \left(\mathbf{1}_{r1}^T, \beta_1 \mathbf{1}_{r1}^T\right),$$

$\mathbf{1}_{r1}$ denotes the $r \times s$ matrix with all entries 1, and $\mathbf{I}_r$ denotes the $r \times r$ identity matrix. The log likelihood of model 3.36-3.41 is

$$ln\, L = const - \frac{1}{2} n\, ln\, |\mathbf{V}| - \frac{1}{2} \sum_n \mathbf{d}_i^T \mathbf{V}^{-1} \mathbf{d}_i, \tag{3.42}$$

where $\mathbf{d}_i = \mathbf{z}_i - \mathbf{m}$. Chan and Mak [6] have shown that

$$ln\, L = const - \frac{1}{2} n\, ln\sigma^2 - \frac{1}{2} rn\, ln\, \sigma_\delta^2 - \frac{1}{2} rn\, ln\, \sigma_\epsilon^2 - \frac{1}{2} n\, ln\, a - \frac{1}{2} \sum_n \left(h_i - c_i^2 a^{-1}\right), \tag{3.43}$$

with

$$h_i = \sigma_\delta^{-2} \sum_r (x_{ij} - \mu)^2 + \sigma_\epsilon^{-2} \sum_r (y_{ij} - \beta_0 - \beta_1 \mu)^2,$$

$$c_i = \sigma_\delta^{-2} \sum_r (x_{ij} - \mu) + \beta_1 \sigma_\epsilon^{-2} \sum_r (y_{ij} - \beta_0 - \beta_1 \mu),$$

$$a = \mathbf{b}^T \mathbf{\Sigma}^{-1} \mathbf{b} + \sigma^{-2} = r\sigma_\delta^{-2} + r\beta_1^2 \sigma_\epsilon^{-2} + \sigma^{-2}.$$

3.3.2 The maximum likelihood solution

Chan and Mak [6] have proved the following lemma.

Lemma 1. The maximum likelihood estimate $\hat{\beta}_0$ and $\hat{\mu}$ for β_0 and μ when β_1, σ^2, σ_δ^2 and σ_ϵ^2 are fixed, satisfy

$$\hat{\mu} = \bar{x}_{..}, \quad \hat{\beta}_0 + \beta_1 \hat{\mu} = \bar{y}_{..},$$

where

$$\bar{x}_{..} = \sum_n \sum_r x_{ij}/(nr) \quad \bar{y}_{..} = \sum_n \sum_r y_{ij}/(nr).$$

Thus to maximize ln L, it suffices to maximize 3.43 with μ and $\beta_0 + \beta_1 \mu$ replaced by $\bar{x}_{..}$ and $\bar{y}_{..}$, respectively.

Let L_1 denote the likelihood function L of 3.43 when μ and $\beta_0 + \beta_1 \mu$ are replaced by $\bar{x}_{..}$ and $\bar{y}_{..}$, respectively. By differentiating lnL_1 with respect to the parameters β_1, σ^2, σ_δ^2 and σ_ϵ^2 and equating to zero, we get

$$\beta_1^2 s_{xy} + \beta_1 \left(\lambda s_{xx} - s_{yy}\right) - \lambda s_{xy} = 0, \tag{3.44}$$

$$\sigma_\epsilon^2 + \beta_1^2 \sigma^2 = t_{yy}, \tag{3.45}$$

$$\sigma_\delta^2 + \sigma^2 = t_{xx}, \tag{3.46}$$

$$(2r - 1)\sigma_\epsilon^2 = r\left(t_{xx}\lambda + t_{yy}\right) - r\left(\beta_1 s_{xy} + \lambda s_{xx}\right), \tag{3.47}$$

where $\lambda = \sigma_\epsilon^2/\sigma_\delta^2$. We are using following notation:

$$t_{xx} = \sum_{i=1}^{n}\sum_{j=1}^{r} \frac{(x_{ij} - \bar{x}_{..})^2}{nr}, \quad t_{yy} = \sum_{i=1}^{n}\sum_{j=1}^{r} \frac{(y_{ij} - \bar{y}_{..})^2}{nr} \tag{3.48}$$

$$w_{xx} = \sum_{i=1}^{n}\sum_{j=1}^{r} \frac{(x_{ij} - \bar{x}_{i.})^2}{rn}, \quad w_{yy} = \sum_{i=1}^{n}\sum_{j=1}^{r} \frac{(y_{ij} - \bar{y}_{i.})^2}{rn} \tag{3.49}$$

$$s_{xx} = \sum_{i=1}^{n} \frac{(\bar{x}_{i.} - \bar{x}_{..})^2}{n}, \quad s_{yy} = \sum_{i=1}^{n} \frac{(\bar{y}_{i.} - \bar{y}_{..})^2}{n} \quad s_{xy} = \sum_{i=1}^{n} \frac{(\bar{x}_{i.} - \bar{x}_{..})\,(\bar{y}_{i.} - \bar{y}_{..})}{n}. \tag{3.50}$$

Chan and Mak [6] have shown the following theorem.

Theorem 1. The maximum likelihood estimate $\hat{\beta}$ is a solution of

$$p_{nr}(\beta) = r^{-1}\left(k_0\beta^4 + k_1\beta^3 + k_2\beta^2 + k_4\right) = 0, \tag{3.51}$$

where

$$k_0 = (r-1)s_{xx}s_{xy}t_{xx}, \tag{3.52}$$

$$k_1 = rs_{xx}^2 w_{yy} - (r-1)s_{xy}^2 t_{xx} - (r-1)s_{xx}s_{yy}t_{xx} - rs_{xy}^2 w_{xx}, \tag{3.53}$$

$$k_2 = (3r-1)(s_{xy}s_{yy}w_{xx} - s_{xy}s_{xx}w_{yy}), \tag{3.54}$$

$$k_3 = rs_{xy}^2 w_{yy} + (r-1)s_{xy}^2 t_{yy} + (r-1)s_{xx}s_{yy}t_{yy} - rs_{yy}^2 w_{xx}, \tag{3.55}$$

$$k_4 = -(r-1)s_{xy}s_{yy}t_{yy} \tag{3.56}$$

if a real solution exists [6].

Other parameters can be calculated using equations 3.44 - 3.47.

Note. The polynomial p_{nr} can have 0, 2 or 4 real roots. The exact real roots can be calculated using the method of Ludovico Ferrari and formulas of Geralomo Cardano. The ML-estimation $(\hat{\beta}_1, \hat{\sigma}^2, \hat{\sigma}_\delta^2, \hat{\sigma}_\epsilon^2)$ should belong to the parameter space of the model. The parameter space is specified with the help of several inequalities. The full specification of the parameter space of the error-in-variables model with replications cannot be found in literature to our knowledge. Quite important are the following inequalities:

$$sign(\hat{\beta}) = sign(s_{xy}); \tag{3.57}$$

$$\sigma^2 > 0;\ \sigma_\epsilon^2 > 0;\ \sigma_\delta^2 > 0 \tag{3.58}$$

Our experience with this method of estimation have shown that if the data suits to the model assumptions, the polynomial has two real roots with different signs in most cases. Using the condition $sign(\hat{\beta}) = sign(s_{xy})$, we can get a unique solution. If the model assumptions are considerably violated (for example there is a strong matrix effect in the test method), the polynomial can have 4 roots. If there is more than one root which satisfies inequalities 3.57-3.58, we can choose the solution which yields the highest likelihood.

Calculations of CIs for estimations

Analytical formulas for calculation of variance of estimations can be found in Chan and Mak [6]. The corresponding confidence intervals (CIs) should be calculated using formulas 3.33-3.35. We use nonparametric methods for calculation of CIs, described in section 3.6.

3.3.3 Averaged Deming regression

The equation 3.44, which was derived directly from $\partial ln\,L_1/\partial \beta_1$ and $\partial ln\,L_1/\partial \sigma^2$ is the familiar equation in the linear structural relationship model when λ is known (see 3.21). Instead of using the full maximum likelihood procedure, one can first estimate λ by w_{yy}/w_{xx}, and finally obtain the estimate β^* of β by 3.44. However, since the variance estimates $r(r-1)^{-1}w_{xx}$ and $r(r-1)^{-1}w_{yy}$ are independent while the maximum likelihood estimates $\hat{\sigma}_\epsilon$ and $\hat{\sigma}_\delta$ are related through 3.45-3.47, the values of $\hat{\beta}_{ML}$ and β^* are not identical [6].

Simply formulated, this means that we can take averaged replications for each item and calculate the Deming regression fit using formulas 3.22-3.23 with λ estimated by w_{yy}/w_{xx}. This approach was proposed by Miller [3, p.210] for the analysis of replicated measurements with errors in both variables. The CLSI EP9 recommends his suggestion as a standard method for the analysis of method comparison studies. In case of normal distributed measurement errors, one can use formulas 3.27-3.29 or 3.30-3.32 for the estimation of the variance of regression parameters and bias at some decision point X_c.

In the following, we will call this method **averaged Deming regression**.

Averaged Deming regression does not yield the true ML estimation but it has several advantages over the ML estimation:

1. It is very simple to calculate.

2. It yields one solution which always lays in the parameter space of the Deming regression model.

3. It's estimations are in most cases very close to the true ML-estimations (see chapter 4).

4. The estimates of the averaged Deming regression are the MM estimates (one can derive them using the method of moments). That is why they are robust against violation of the normality assumptions [3, p. 212-213].

5. Taking the means of replications we reduce the variability of the data and gain thereby additional robustness.

Calculations of CIs for estimations

For calculation of CIs of estimations of averaged Deming regression, the corresponding formulas for simple Deming regression can be used (3.27-3.35). If model assumptions are violated, nonparametric methods for calculation of CIs, described in section 3.6, should be used.

3.3.4 Linnet's weighted Deming regression

Now we consider the situation with constant coefficients of variation (CV) of measurement errors. Similar to the weighted OLS regression, we can give some weight to each data point and calculate the ML estimates for the weighted data. A procedure to calculate the weighted Deming regression was discussed by Linnet ([7], [8]).

Linnet assumes that the variances of the are proportional to the squares of the average of the true measurement values ξ_i and η_i. Then

$$Var(\delta_i) = f_x^2 \left[\frac{\xi_i + \eta_i}{2} \right]^2, \tag{3.59}$$

$$Var(\epsilon_i) = f_y^2 \left[\frac{\xi_i + \eta_i}{2} \right]^2 \tag{3.60}$$

Under this assumption, $\lambda^* = f_x^2/f_y^2$ is constant.

Note: Linnet considers λ as the ratio of the variance of the measurement error of the comparative method X to the variance of the measurement error of the test method Y. Most books and papers designate λ as the ratio of variance of measurement error of Y to the variance of the measurement error of X. That is why we denote Linnet's version as λ^*. Then:

$$\lambda^* = \lambda^{-1} \tag{3.61}$$

A weighted modification of the slope estimation procedure is to minimize the sum of squares

$$S_w = \sum_{i=1}^{n} \left[w_i \left(x_i - \hat{\xi}_i \right)^2 + \lambda^* w_i \left(y_i - \hat{\eta}_i \right)^2 \right],$$

where the weights

$$w_i = \left[\frac{\xi_i + \eta_i}{2} \right]^{-2}$$

and

$$\hat{\eta}_i = \bar{y}_w + \beta_1 \left(\hat{\xi}_i - \bar{x}_w \right).$$

The estimation of the regression coefficients can be carried out using the following formulas:

$$\hat{\beta}_1 = \frac{(\lambda^* q_w - u_w) + \sqrt{(u_w - \lambda^* q_w)^2 + 4\lambda^* p_w^2}}{2\lambda^* p_w} \tag{3.62}$$

$$\hat{\beta}_0 = \bar{y}_w - \hat{\beta}_1 x_w. \tag{3.63}$$

Here

$$\bar{x}_w = \sum_i w_i x_i / \sum_i w_i, \qquad\qquad \bar{y}_w = \sum_i w_i y_i / \sum_i w_i,$$

$$u_w = \sum_i w_i \left(x_i - \bar{x}_w\right)^2, \qquad\qquad q_w = \sum_i w_i \left(y_i - \bar{y}_w\right)^2,$$

$$p_w = \sum_i w_i \left(x_i - \bar{x}_w\right)\left(y_i - \bar{y}_w\right).$$

In order to estimate λ^*, Linnet suggests to use duplicate measurements. Let the data contain n pairs of duplicated measurements $\left(\left(x_{i1}, x_{i2}\right), \left(y_{i1}, y_{i2}\right)\right)$. For the calculation of regression coefficients, the means of duplicates $\left(x_{i\cdot}, y_{i\cdot}\right)$ should be used. Then λ^* should be estimated as follows:

$$\hat{f}_x^2 = \frac{1}{2n} \sum_i \frac{\left(x_{i1} - x_{i2}\right)^2}{w_i^2}, \tag{3.64}$$

$$\hat{f}_y^2 = \frac{1}{2n} \sum_i \frac{\left(y_{i1} - y_{i2}\right)^2}{w_i^2}, \tag{3.65}$$

$$\hat{\lambda}^* = \hat{f}_x^2 / \hat{f}_y^2. \tag{3.66}$$

In case of more than two measurement replicates we can generalize formulas 3.64-3.65. Assume that we have r replications $(r \geq 2)$ of each measurement method. Then:

$$\hat{f}_x^2 = \frac{1}{n\, r\, (r-1)} \sum_{i=1}^n \sum_{k=1}^r \sum_{l:k<l} \frac{\left(x_{ik} - x_{il}\right)^2}{w_i^2}. \tag{3.67}$$

with a similar expression for $\hat{f}_y^2$.

Note that from 3.59:

$$\sigma_{\delta i}^2 = f_x^2 \, w_i^2 \iff f_x^2 = \frac{\sigma_{\delta i}^2}{w_i^2}.$$

Assume that w_i are the true weights of the data points. Then

$$E\left(\frac{1}{nr(r-1)} \sum_{i=1}^n \sum_{k=1}^r \sum_{l:k<l} \frac{\left(x_{ik} - x_{il}\right)^2}{w_i^2}\right) =$$

$$\frac{1}{nr(r-1)} \sum_{i=1}^n \sum_{k=1}^r \sum_{l:k<l} E\left(\frac{\left(\delta_{ik} - \delta_{il}\right)^2}{w_i^2}\right) =$$

$$\frac{1}{nr(r-1)} \frac{r(r-1)}{2} \sum_{i=1}^n E\left(\frac{2\sigma_{\delta i}^2}{w_i^2}\right) = \frac{1}{n} \sum_{i=1}^n E\left(\frac{\sigma_{\delta i}^2}{w_i^2}\right) = f_x^2.$$

In order to get unbiased estimations of regression coefficients, we should know the true weights of the data points w_i. We can use

$$\hat{w}_i = \left(0.5\left(x_{i\cdot} + y_{i\cdot}\right)\right)^2 \tag{3.68}$$

as estimations for the weights. Linnet has shown that if we use such weights, the estimation of the slope is biased [7]. He has suggested an iterative procedure for calculation of regression coefficients. This procedure estimates weights using estimations $\hat{\xi}_i$ and $\hat{\eta}_i$ of the true error free measurements for the i-th data point. The procedure should be carried out as follows:

23

1. Take as initial weights $\hat{w}_i = (0.5\,(x_{i\cdot} + y_{i\cdot}))^2$. Calculate $\hat{\lambda}^*$ using formulas 3.64, 3.65, 3.66 and regression coefficients $\hat{\beta}_1$ and $\hat{\beta}_0$ using formulas 3.62-3.63.

2. Update the estimations of the weights:

$$\hat{w}_i^{(1)} = \left[\frac{\hat{\xi}_i + \hat{\eta}_i}{2}\right]^{-2},$$

where

$$\hat{\xi}_i = x_{i\cdot} + \hat{\lambda}^*\hat{\beta}_1\,d_i\left(1 + \hat{\lambda}^*\beta_1^2\right)^{-1},$$
$$\hat{\eta}_i = y_{i\cdot} - d_i\left(1 + \hat{\lambda}^*\beta_1^2\right)^{-1},$$
$$d_i = y_{i\cdot} - \left(\hat{\beta}_0 + \hat{\beta}_1 x_{i\cdot}\right)$$

3. Update estimations $\hat{\lambda}^*$ using new weights and estimate regression coefficients.

4. Iterate the procedure to estimate better values of weights to improve the weights in each iteration

Note: Later, Linnet gives better estimations for the weights [8]:

$$\hat{w}_i^{(1)} = \left[\frac{\hat{\xi}_i + \hat{\lambda}^*\hat{\eta}_i}{1 + \hat{\lambda}^*}\right]^{-2},$$

We will use these estimations for the calculation of weighted Deming regression fit.

Note: Linnets algorithm converges correctly only if all measurements of the both measurement procedures are positive.

calculation of CIs for estimates

For the estimation of variances of regression coefficients, Linnet suggests jackknife method [7],[8] (look at section 3.6). In order to prove performance of this method for estimation of variances of regression coefficients he has conducted a series of simulations which confirmed good performance of jackknife estimations. Another nonprametric methods, described in section 3.6, can also be used for calculation of CIs of the parameter estimates.

3.3.5 Barnett's approach for inhomogeneous error variances

This approach was developed for the investigation of the relationship between protein content of the urine and the administered dose of a particular drug [14]. The corresponding data contained replicated measurements with inhomogeneous variances. The model Barnett uses for his approach suits to the method comparison data with heteroscedastic errors where the CV is not constant. Barnett considers the following model:

Assume that we have two variables X and Y. Both of them are measured with error. There are n items, for each item n_i replicated measurements of X and Y are available. Denote ξ_i the true measurement of X and η_i the true measurement of Y. Let δ_{ij} be a measurement error of X for item i by j-th replication and ϵ_{ij} be a measurement error of Y for item i by j-th replication. Then the observed measurements x_{ij} and y_{ij} can be written in the following form:

$$x_{ij} = \xi_i + \delta_{ij},$$

$$y_{ij} = \eta_i + \epsilon_{ij},$$

$$\eta_i = \beta_0 + \beta_1 \xi_i.$$

Measurement errors are assumed to be uncorrelated and normal distributed:

$$\delta_{ij} \sim N\left(0, \sigma_i^2\right),$$

$$\epsilon ij \sim N\left(0, \lambda \sigma_i^2\right)$$

Thus it is a functional model (the true measurements of X are fixed constants). The variances are different for each item. The ratio between variances of the measurement errors of X and Y (λ) are constant.

The log-likelihood is

$$L\left(x, y | \beta_0, \beta_1, \lambda, \sigma, \xi\right) = const - \sum_{i=1}^{n} n_i log\left(\sigma_i^2 \sqrt{\lambda}\right) -$$

$$- \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n_i} \left\{(x_{ij} - \xi_i)^2 \sigma_i^{-2}\right\} - \frac{1}{2} \sum_{i=1}^{n} \sum_{j=1}^{n_i} \left\{(y_{ij} - \beta_0 - \beta_1 \xi_i)^2 \left(\lambda \sigma_i^2\right)^{-1}\right\} \quad (3.69)$$

Barnett derives the following formulas for calculation of ML estimates:

$$\hat{\sigma}_i^2 = \frac{\sum_j (x_{ij} - x_{i\cdot})^2 + \hat{\lambda}^{-1} \sum_j (y_{ij} - y_{i\cdot})^2}{2n_i} + \frac{\left(y_{i\cdot} - \hat{\beta}_0 - \hat{\beta}_1 x_{i\cdot}\right)^2}{2\hat{\lambda}\hat{\Delta}}, i = 1,...p, \quad (3.70)$$

$$\hat{\lambda} = n^{-1} \sum_i (\hat{\sigma}_i^2)^{-1} \left(\sum_j (y_{ij} - y_{i\cdot})^2 + \frac{n_i}{\hat{\Delta}^2}\left(y_{i\cdot} - \hat{\beta}_0 - \hat{\beta}_1 x_{i\cdot}\right)^2\right) \quad (3.71)$$

$$\hat{\Delta} = 1 + \frac{\hat{\beta}_1^2}{\hat{\lambda}}. \quad (3.72)$$

Then, denoting

$$\hat{N} = \sum_{i=1}^{p} \frac{n_i}{\hat{\sigma}_i^2}; \quad \xi_i = \frac{x_{i\cdot}}{\hat{\sigma}_i^2} \quad \eta_i = \frac{y_{i\cdot}}{\hat{\sigma}_i^2} \quad (3.73)$$

$$\xi_{\cdot\cdot} = \sum_i \xi_i; \quad \eta_{\cdot\cdot} = \sum_i \eta_i \quad (3.74)$$

one gets ML estimators for regression coefficients:

$$\hat{\beta}_0 = \frac{\eta_{\cdot\cdot} - \hat{\beta}_1 \xi_{\cdot\cdot}}{\hat{N}} \quad (3.75)$$

$$\hat{\beta}_1 = \Theta \pm \sqrt{\Theta^2 + \hat{\lambda}}, \quad \Theta = \frac{s_{\eta\eta} - \hat{\lambda} s_{\xi\xi}}{2s_{\eta\xi}}, \quad (3.76)$$

The formula 3.75 gives two roots. The root with the same sign as $s_{\eta\xi}$ should be chosen.
Here

$$s_{\eta\xi} = \sum_i \frac{\hat{\sigma}_i^2}{n_i} \xi_i \eta_i - \frac{\xi_{..}\eta_{..}}{\hat{N}}, \tag{3.77}$$

$$s_{\xi\xi} = \sum_i \frac{\hat{\sigma}_i^2}{n_i} \xi_i^2 - \frac{\xi_{..}^2}{\hat{N}}, \tag{3.78}$$

$$s_{\eta\eta} = \sum_i \frac{\hat{\sigma}_i^2}{n_i} \eta^2 - \frac{\eta_{..}^2}{\hat{N}}. \tag{3.79}$$

Barnett notes, that the solution is analogous to the solution of the problem for the unreplicated case. In fact, the ML estimates of the regression coefficients are calculated using averaged replications of X and Y which are weighted with $\hat{\sigma}_i$. The quantities $\frac{n_i}{\hat{\sigma}_i}$ can be considered as sample size for the different groups of observations.

It is not feasible to obtain closed-form expressions for the ML estimations $\hat{\beta}_0$, $\hat{\beta}_1$, $\hat{\lambda}$, $\hat{\sigma}_i^2$ by the observations (x_{ij}, y_{ij}) alone. That is why Barnett suggests an iterative procedure for the computation of ML estimates:

1. First, the starting values for the estimates $\hat{\sigma}_i^0$ and $\hat{k}^0$ should be calculated:

$$\hat{\sigma}_i^0 = n_i^{-1} \sum_i (x_{ij} - x_{i.})^2, \hat{\lambda}^0 = n^{-1} \sum_i \sum_j \frac{(y_{ij} - y_{i.})^2}{\hat{\sigma}_i^0} \tag{3.80}$$

2. Then the starting estimates $\hat{\beta}_0$ and $\hat{\beta}_1$ can be obtained directly from 3.75 and 3.76.

3. These can be developed iteratively by using equations 3.70, 3.71, 3.76, 3.75 in that order until the estimates become stable in value.

Barnett admits that his algorithm is not ideal. Often it converges quite slowly.

Our experience with Barnett's algorithm have shown that the algorithm sometimes needs more than 100 iterations until it converges, especially in cases of data with only 2 measurement replications. Simulation studies (see chapter IV, simulation 1.2) have shown that the algorithm gives biased estimates. The variance of estimates is relatively high. Centering of data reduces the bias of Barnett's approach considerably, but the variance of estimates stays much higher than the variances of the other methods described in this work.

Unfortunately, no more information about this approach was found in the literature. So there are no references which could help to validate our implementation of Barnett's algorithm or support the results of our simulations.

calculation of CIs for estimates

In his paper [15] Barnett gives calculation of the inverse Fisher matrix of estimates. Thus, it is possible to calculate analytical CIs for the estimates. All nonparametric methods described

in section 3.6 may be also used for calculation of CIs. The performance of different methods of calculation of CIs has not been analyzed for Barnett's method.

3.4 Modeling matrix effect in test method

This section deals with modeling of the matrix effect in the test method. First, the idea of the matrix effect is briefly explained. Then, the linear structural ME model with error in equation is described, which can be used for modeling of a matrix effect in test method. After this, MM parameter estimates for this model are given.

3.4.1 Matrix effect

In chemical analysis, matrix refers to the components of a sample other than the analyte of interest. The matrix can have a considerable effect on the measurement results; such effect is called **matrix effect**.

Roughly speaking, absolute matrix effect is the difference between the response of equally concentrated analytes in solvent and in matrix extracts (urine or blood, for example). The matrix may cause signal enhancement or signal suppression that negatively affect the measurement quality.

A relative matrix effect is the variation of absolute matrix effects between several lots of the same matrix.

3.4.2 Model formulation

Discussion about modeling of the matrix effect in method comparison studies can be found at Carstensen [16] and Dunn et.al. [10]. In both sources, the matrix effect is considered a random effect which affects the true measurements. We consider the model from the paper of Dunn and Roberts:

$$x_{ij} = \xi_i + m_i + \delta_{ij},$$

$$y_{ij} = \beta_0 + \beta_1 \xi_i + c_i + \epsilon_{ij}.$$

Here

$i = 1, 2, ..., n$ - denote the items,

$j = 1, ...n_i$ - denote the measurement replications.

δ_{ij} and ϵ_{ij} are normal distributed and uncorrelated measurement errors,

$m_i \sim N(0, \sigma_m^2)$ is the matrix effect of method X,

$c_i \sim N(0, \sigma_c^2)$ is the matrix effect of method Y.

Matrix effects are independent of measurement errors and of each other. Dunn et.al. mention that

this model is under-identified. That is why we need any assumption about the variances of matrix effects. There are two possibilities:

1. $\sigma_m^2 = \sigma_c^2$,

2. $\sigma_m^2 = 0$.

One should take the restriction which better suits to the data. Since CLSI EP9 requires that the comparative method should not contain any matrix effect in the following we consider the second restriction.

Dunn et. al. use SEM (structural equation models) for estimation of this model. Our experience have shown that the computations using SEM were rather slow. That is why we use more simple methods of estimation.

The estimates of the regression coefficients for the model with matrix effect only in the test method were derived by Fuller [2, p.106-110]. He considers the following model:

$$x_{ij} = \xi_i + \delta_{ij}, \tag{3.81}$$

$$y_{ij} = \eta_i + \epsilon_{ij} \tag{3.82}$$

$$\eta_i = \beta_0 + \beta_1 \xi_i + c_i. \tag{3.83}$$

This model is called **the model with an error in the equation.** Fuller gives ML estimates of the regression coefficients of this model assuming that the variances of the measurement errors σ_δ^2 and σ_ϵ^2 are known. He recommends this model specially for laboratory experiments with replicated measurements, where variables x and y do not have the perfect linear relation to each other. He suggests to use the replications for the estimation of variances of measurement errors. The calculation of regression coefficients should be conducted using means of replications. Then

$$x_{i\cdot} = \xi_i + \bar{\delta}_i, \tag{3.84}$$

$$y_{i\cdot} = \eta_i + \bar{\epsilon}_i, \tag{3.85}$$

$$\eta_i = \beta_0 + \beta_1 \xi_i + c_i. \tag{3.86}$$

The estimates of $\beta = (\beta_0, \beta_1)$ in case the measurement errors are uncorrelated can be calculated using following formula [2, p. 110]:

$$\tilde{\beta} = (M_{xx} - S_{\delta\delta})^{-1} M_{xy}, \tag{3.87}$$

where

$$M_{xx} = \sum_{i=1}^{n} (1, x_i)^T (1, x_i) = \begin{pmatrix} 1 & \bar{x} \\ \bar{x} & \bar{x^2} \end{pmatrix}, \quad M_{xy} = \sum_{i=1}^{n} (1, x_i)^T y_i = \begin{pmatrix} \bar{y} \\ \bar{xy} \end{pmatrix}, \quad S_{\delta\delta} = \begin{pmatrix} 0 & 0 \\ 0 & \hat{\sigma}_\delta^2 \end{pmatrix}.$$

$\hat{\sigma}_\delta^2$ is an unbiased estimate of the variance of measurement error of X.

Note: As in case of the averaged Deming regression, the $\tilde{\beta}$ is not a full ML estimation.

Oman et.al [9] use Fuller's results for the analysis of the relationship between the two methods of measurement of glomerular filtration MCC and ECC. They extend his result for unbalanced design (the number of replications m_i for each item can be different) with correlated errors using the method of moments. It is easy to adapt their formulas for the case with uncorrelated errors. Denoting

$$S_{xxw} = \sum_{i=1}^{n} \sum_{j=1}^{m_i} (x_{ij} - x_{i\cdot})^2, \quad S_{yyw} = \sum_{i=1}^{n} \sum_{j=1}^{m_i} (y_{ij} - y_{i\cdot})^2,$$

$$S_{xxb} = \sum_{i=1}^{n} m_i (x_{i\cdot} - x_{\cdot\cdot})^2, \quad S_{yyb} = \sum_{i=1}^{n} m_i (y_{i\cdot} - y_{\cdot\cdot})^2,$$

$$S_{xyb} = \sum_{i=1}^{n} m_i (x_{i\cdot} - x_{\cdot\cdot}) (y_{i\cdot} - y_{\cdot\cdot})$$

$$N = \sum_{i=1}^{n} m_i, \quad t = N^{-2} \sum_{i=1}^{n} m_i^2, \quad f = (n-1)/(N-1),$$

we can use following MM estimates for the model with matrix effect in the test method [9]:

$$\hat{\sigma}_\delta^2 = \frac{1}{N-n} S_{xxw}, \quad \hat{\sigma}_\epsilon^2 = \frac{1}{N-n} S_{yyw}, \tag{3.88}$$

$$\hat{\sigma}^2 = \frac{1}{N(1-t)} \{S_{xxb} - f S_{xxw}\} \tag{3.89}$$

$$\hat{\beta}_1 = \frac{S_{xyb}}{S_{xxb} - f S_{xxw}}, \tag{3.90}$$

$$\hat{\sigma}_c^2 = \frac{1}{N(1-t)} \left\{ S_{yyb} - f S_{yyw} - \frac{S_{xyb}^2}{S_{xxb} - f S_{xxw}} \right\} \tag{3.91}$$

$$\hat{\mu} = x_{\cdot\cdot}, \quad \hat{\beta}_0 = y_{\cdot\cdot} - \hat{\beta}_1 x_{\cdot\cdot} \tag{3.92}$$

Equations 3.89, 3.91 may give negative variances, indicating that the model is not appropriate for the data at hand. Oman et.al. interpret negative variance estimates as diagnostics. In particular, $\hat{\sigma}^2 \leq 0$ suggests that $\xi_i \equiv 0$ and thus that β_0 and β_1 are not identified.

Calculation of CIs for estimates

Deriving asymptotic standard errors using the delta method was shown to be very complicated, that is why Oman et.al. [9] suggest using parametric or nonparametric bootstrap, described in section 3.6. One may also use other nonparametric methods of calculation if CIs for estimates considered in section 3.6.

3.5 Diagnostic plots for checking the model assumptions

It is important to choose the correct model for the data. In this section we consider some diagnostic plots which can support this decision. Examples of diagnostic plots can be found in the practical part of the master thesis, where they are applied to the example data (see chapter 5).

3.5.1 Checking exchangeability of replications

Replicate measurements by a method on a given item are exchangeable if the distribution of data is invariant under permutation of replicate numbers within method and item [16, p. 50]. This is typically the case if the replicate measurements are made independently. Exchangeability of replications cannot be determined on the data alone; it is a function of study design [16, p. 64]. The design of method comparison data should provide exchangeable replications.

If there are doubts about the exchangeability of measurement replications we can make a random permutation of the replicates within method and item and see if the plot of the two methods against each other look the same [16, p. 51]. Results of regression analysis for original data and data with permuted replications should be similar.

If the replications were conducted during a particular period of time, the following recommendation of Oman et. al.[9] can be useful. In order to check whether the assumption about exchangeability of replications is violated they suggest to plot within-individual paired differences $d_{i(jk)} = x_{ij} - x_{ik}$ for the points in time of the measurements $t_{ij} > t_{ik}$ vs. $t_{i(jk)} = t_{ij} - t_{ik}$. If the measurements are exchangeable, the graph should pick up a monotone trend.

3.5.2 Checking assumption about normality and homoscedasticity of measurement errors before model was fitted

To check this assumptions we can plot

1. absolute differences of averaged replications $(x_{i.} - y_{i.})$ by two methods against the average levels $(x_{i.} + y_{i.})/2$,

2. relative differences of averaged replications $\frac{(x_{i.} - y_{i.})}{(x_{i.} + y_{i.})/2}$ by two methods against the average levels $(x_{i.} + y_{i.})/2$.

These are the Bland-Altman plots, which belong to the standard empirical analysis of the method comparison data [14]. If the first scatter plot has trumpet shape and the points of the second plot disperse randomly around zero, then the measurement errors are heteroscedastic with constant CV.

Additionally, we can use the differences between replications within each item to check the model assumptions about variances of measurement errors. If the data contain duplicated measurements, the following plots, suggested by Linnet [7] can be useful:

3. differences between replicates $(x_{i1} - x_{i2})$ vs. $(x_{i1} + x_{i2})/2$,

4. differences between replicates $(y_{i1} - y_{i2})$ vs. $(y_{i1} + y_{i2})/2$.

If the data contain more than two replications for each item, we can use plots suggested by Oman et.al. [9]:

5. absolute values of the within-individual deviations $x_{ij} - x_{i.}$ vs. individual means $x_{i.}$,

6. absolute values of the within-individual deviations $y_{ij} - y_{i\cdot}$ vs. individual means $y_{i\cdot\cdot}$.

If the differences get larger with increasing values of the measurements, then the data have proportional errors.

3.5.3 Residual plots

After the regression model was fitted, we can check the model assumptions using residual plots. Generally, in regression analysis residual plots can help to prove the following model assumptions [see 17, pp. 64-71]:

- assumption about normality and homoscedasticity of error terms,

- assumptions about independence of error terms,

- linearity assumption.

In case of OLS regression, the residual plot is a scatter plot with model residuals on the y-axis and (in most cases) fitted values of the response variable on the x-axis. A residual is defined as $y_i - \hat{y}_i = y_i - (\hat{\beta}_0 + \hat{\beta}_0 x_i)$.

Linear regression models with errors in both variables yield three kinds of residuals:

1. horizontal residuals $x_i - \hat{x}_i$,

2. vertical residuals $y_i - \hat{y}_i$,

3. optimized residuals (the distances between data points and their predictions which were minimized by estimation of regression parameters).

Residual plot for the ML-approach, averaged- and weighted Deming regression

For averaged and weighted Deming regression, we can use Linnet's suggestion [7] and plot the optimized residuals vs. the mean of predicted values of the test- and comparative method. Assume, $\hat{\xi}_i$ is the predicted value of $x_{i\cdot}$ and $\hat{\eta}_i$ is the predicted value of $y_{i\cdot}$. We can calculate the predicted values using the following formulas [9]:

$$\hat{\xi}_i = x_{i\cdot} + \hat{\lambda}^* \hat{\beta}_1 d_i \left(1 + \hat{\lambda}^* \beta_1^2\right)^{-1},$$
$$\hat{\eta}_i = y_{i\cdot} - d_i \left(1 + \hat{\lambda}^* \beta_1^2\right)^{-1},$$
$$d_i = y_{i\cdot} - \left(\hat{\beta}_0 + \hat{\beta}_1 x_{i\cdot}\right)$$

Here $\lambda^* = \lambda^{-1}$. The x-axis of the residual plot contains the mean of the predicted values $\left(\hat{\xi}_i + \hat{\eta}_i\right)/2$.

The y-axis contains the optimized residuals. For Deming regression with $\hat{\lambda} = 1$ we consider as

optimized residuals the square roots of euclidean distances between data points and their predictions. The sign of the residual indicates if the data point lays above or below the regression line. If $\hat{\lambda} \neq 1$ the distance should be slightly changed (see 3.26):

$$r_i = sign \cdot \left[\left(x_{i\cdot} - \hat{\xi}_i \right)^2 + \hat{\lambda}^* \left(y_{i\cdot} - \hat{\eta}_i \right)^2 \right]^{1/2}. \tag{3.93}$$

For the weighted Deming regression the residuals should be additionally weighted [7]:

$$r_i^w = sign \cdot \left[\hat{w}_i \left(x_{i\cdot} - \hat{\xi}_i \right)^2 + \hat{w}_i \hat{\lambda}^* \left(y_{i\cdot} - \hat{\eta}_i \right)^2 \right]^{1/2} \tag{3.94}$$

$$\tag{3.95}$$

The *sign* of r_i is identical with the sign of the vertical distance d_i from the regression line.
If the residuals disperse randomly around the zero line we can conclude that the model assumptions hold.

Residual plot for the model with matrix effect in the test method

Oman et.al. [9] suggest to plot the predicted values $\hat{\xi}_i$ vs. $\hat{r}_i = y_{i\cdot} - \hat{\beta}_0 - \hat{\beta}_1 y_{i\cdot}$. They have shown that if the model assumptions hold, ξ_i and r_i should be independent. $\hat{\xi}_i$ can be calculated using the following formulas:

$$\hat{\xi}_i = \hat{\theta}^T \hat{W}_i^{-1} z_i \left(\hat{\theta}^T \hat{W}_i \hat{\theta} \right)^{-1}, \tag{3.96}$$

where

$$z_i = \begin{pmatrix} x_{i\cdot} \\ y_{i\cdot} - \hat{\beta}_0 \end{pmatrix}, \quad \hat{\theta} = \begin{pmatrix} 1 \\ \hat{\beta}_1 \end{pmatrix}, \quad \hat{W}_i = \begin{pmatrix} m_i^{-1} \hat{\sigma}_\delta^2 & 0 \\ 0 & \hat{\sigma}_c^2 + m_i^{-1} \hat{\sigma}_\epsilon^2 \end{pmatrix}.$$

3.6 Nonparametric and semiparametric approaches for the calculation of confidence intervals

In section 2.3.1 we have considered analytical intervals for regression parameters for the averaged Deming regression. This confidence intervals should only be used if model assumptions about homogeneous variances of the measurement errors hold. Otherwise one should use nonparametric methods of estimates. For other regression methods which are considered in this master thesis (Linnet's weighted Deming regression, Barnett's method and MM-estimation for the model with matrix effect in test method) we will use only nonparametric methods for calculation of confidence intervals.

Now we consider some methods of nonparametric and semiparametric estimation of the confidence intervals (CIs). For general description of this methods we use recommendations of Efron and Tibshirani [5]. Specific recommendations for method comparison data were found at Linnet [7], [8] and Oman et.al. [9]. It is interesting to consider suggestions of Oman et.al.[9] in more detail, because to our knowledge it is the only source talking about the bootstrap for bivariate regression model with errors in both variables. Additionally, R.Tibshirani has discussed the paper with the authors, that is why the suggested bootstrap methods should agree with his opinion about it.

There are some possibilities to calculate CIs for estimates of regression parameters. First, we can use the fact, that if we have only one predictor, the regression coefficients have t distribution with n-2 degrees of freedom [17, p.]. So the (1-α)%-confidence interval can be calculated using formula

$$\hat{\theta} \pm t_{n-2;1-\frac{\alpha}{2}} se(\hat{\theta}), \tag{3.97}$$

where $t_{n-2;1-\frac{\alpha}{2}}$ denotes $1 - \frac{\alpha}{2}$ quantile of t distribution with n-2 degrees of freedom.

For the estimation of $se(\theta)$ we can use resampling methods (jackknife or bootstrap).

3.6.1 Estimation of variances of regression parameters

The idea of resampling estimation of the variance of some statistic is as follows. Suppose that z is a data vector with assumed joint distribution H, let $\hat{\theta} = t(z)$ be some statistic and consider the functional H, $\sigma(H) = var_H(\hat{\theta})$. The resampling estimator of $\sigma(H)$ is $\sigma(\hat{H})$, where $\hat{H}$ is an appropriate estimator of H. We can approximate $\sigma(\hat{H})$ by generating B samples $\{z_1^*, z_2^*, ... z_B^*\}$ from $\hat{H}$ and computing the sample variance of estimates $\theta_b^* = t(z_b^*)$ [9].

Jackknife method

Suppose, we have data $z = \{z_1, z_2, ... z_n\}$, where $z_i = (x_{i1}, x_{i2}, ... x_{ir}, y_{i1}, y_{i2}, ..., y_{ir})$. The jack-knife focuses on the samples that leave out one observation at a time:

$$z_{(i)} = (z_1, z_2, ..., z_{i-1}, z_{i+1}, ..., z_n). \tag{3.98}$$

We consider the version of variance estimation in terms of pseudo-values ([7]; [5, p. 149]):

$$\tilde{\theta}_i = n\hat{\theta} - (n-1)\hat{\theta}_{(i)}, \tag{3.99}$$

were $\hat{\theta}$ is the global estimation of θ and $\hat{\theta}_{(i)}$ is estimation of θ using sample $z_{(i)}$. Then

$$\hat{se}_{jack} = \left\{ \frac{\sum_{i=1}^{n} \left(\tilde{\theta}_i - \tilde{\theta} \right)^2}{(n-1)n} \right\}^{1/2}, \tag{3.100}$$

where $\tilde{\theta} = n^{-1} \sum_{i=1}^{n} \tilde{\theta}_i$.

Note: Jackknife approach may fail if the statistic $\hat{\theta}$ is not smooth. The idea of smoothness is that small changes in the data set cause only small changes in the statistic. A simple example of a non-smooth statistic is the median [5, p. 148].

The jackknife approach is recommended for the estimation of the variance of regression coefficients and bias at some medical decision point X_c for the averaged Deming regression (if the normality assumption is violated) and Linnet's weighted Deming regression [7], [3].

Bootstrap methods

In this case $\sigma(\hat{H})$ is approximated by using Monte Carlo sampling, generating a large number B of bootstrap samples $\{z_1^*, z_2^*, ... z_B^*\}$ from $\hat{H}$. Each z_b^* is a vector of n drawings (with replacement) either from the empirical distribution of z (the nonparametric bootstrap) or from a parametric estimate of distribution of z based on the learning sample z. We consider three possibilities to draw bootstrap samples in case of replicated data of the method comparison study. The main source of information about this problem is the paper of Oman et.al [9] (see section 3.4).

First nonparametric approach In the first nonparametric approach each bootstrap sample is computed by sampling n items with replacement and using all observations of each sampled individual. In case of an unbalanced design (different number of replications for each item) this method can overestimate $\sigma(\hat{H})$ because the number of observations of the learning sample is not preserved.

Second nonparametric approach the second nonparametric approach is called bootstrapping the residuals. This method is described for the case of OLS regression by Efron and Tibshirani [5, p. 113]. After fitting the linear model we get estimates of regression coefficients $\hat{\beta}_0$ and $\hat{\beta}_1$ and residuals $\hat{r}_i = y_i - \hat{\beta}_0 - \hat{\beta}_1 x_i$. We can generate new samples using the following data:

$$(x_i^*, y_i^*) = (x_i, \hat{\beta}_0 + \hat{\beta}_1 x_i + \hat{r}_i^*),$$

where $\hat{r}_i^*$ is randomly sampled from the set $\{\hat{r}_1, ..., \hat{r}_n\}$. The advantage of this method is that the study design stays unchanged by resampling. The downside is the sensitivity to the model assumptions. We did not find a reference how to conduct this kind of bootstrap in case of Deming regression, with errors in both variables. In case of unreplicated measurements it could be

conducted as following:

$$(x_i^*, y_i^*) = (\hat{\xi}_i, \hat{\eta}_i) + (\hat{\delta}_i^*, \hat{\epsilon}_i^*),$$

where $\hat{\xi}_i$ and $\hat{\eta}_i$ are the predictions of the true measurements of methods X and Y for the i-th item. The $\hat{\delta}_i^*$ and $\hat{\epsilon}_i^*$ are randomly sampled from the sets $\left\{ x_i - \hat{\xi}_i \,|\, i = 1, 2, \dots, n \right\}$ and $\{y_i - \hat{\eta}_i \,|\, i = 1, 2, \dots, n\}$ respectively. However it is still unclear how should be drawn the samples in replicated case.

Oman et.al. [9] also abolish the idea of residual resampling for the structural model with error in equation.

Parametric bootstap In this case samples are drawn from the estimated distribution of the data. This approach is used by Oman et.al. [9] for estimation of variances of the regression coefficients. It is relatively unproblematic to conduct parametric bootstrap for the averaged Deming regression or for the model with matrix effect in the test method. For the averaged Deming regression we can draw every bootstrap sample as follows:

1. Generate $\{\xi_1, ..., \xi_n\}$ from the normal distribution with mean $\hat{\mu}$ and variance $\hat{\sigma}^2$.

2. Generate measurement errors $\{(\delta_{ij}, \epsilon_{ij}) \,|\, i = 1, \dots n, j = 1, \dots r\}$, where

$$\begin{pmatrix} \delta_{ij} \\ \epsilon_{ij} \end{pmatrix} \sim N \left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \hat{\sigma}_\delta^2, 0 \\ 0, \hat{\sigma}_\epsilon^2 \end{pmatrix} \right).$$

Then the bootstrap sample contains pairs

$$(x_{ij}^*, y_{ij}^*) = \left(\xi_i, \hat{\beta}_0 + \hat{\beta}_1 \xi_i \right) + (\delta_{ij}, \epsilon_{ij}). \tag{3.101}$$

Analogously one can conduct parametric bootstrap in case of the structural model with error in equation (see [9]).

The sampling for Linnet's weighted Deming regression is problematic, because Linnet's model definition is somewhat unclear.

The only bootstrap method, which can be applied to all models considered in this master thesis, is the nonparametric bootstrap. Thus in the following we take a closer look only at this approach.

3.6.2 Calculation of bootstrap confidence intervals

The confidence interval 3.97 assumes an equal-tailed distribution of $\hat{\theta}$. Sometimes the distribution of $\hat{\theta}$ is skewed. Then the quantile or BC_a method could be more appropriate for estimation of the confidence intervals.

The quantile (percentile) CI

Assume that we generate B independent bootstrap data sets $\{z_1^*, z_2^*, \ldots z_B^*\}$ and compute the bootstrap replications $\theta_b^* = t(z_b^*)$, b = 1, 2, $\ldots$, B. The approximate $(1-\alpha)$-% quantile interval is given by [5, p. 170]

$$\left[\hat{\theta}_{lo}, \hat{\theta}_{up}\right] \approx \left[\theta_b^{*(\alpha/2)}, \theta_b^{*(1-\alpha/2)}\right], \tag{3.102}$$

were $\theta_b^{*(\alpha/2)}$ and $\theta_b^{*(1-\alpha/2)}$ denote $100 \cdot \alpha/2$ % and $100 \cdot (1-\alpha/2)$ % percentiles of the empirical distribution of $\theta_b^* = t(z_b^*)$. Quantile intervals achieve better balance in the left and right sides than the simple CIs (3.97). But they can still undercover overall. This is the consequence of non-parametric inference: the quantile interval has no knowledge of the underlying normal distribution and uses the empirical distribution in its place [5, p. 175]. The BC_a CIs should achieve better coverage than the percentile CI.

The BC_a method

The BC_a CIs are a substantial improved version of the percentile method. The abbreviation is standing for "bias-corrected and accelerated". Efron and Tibshirani have shown that BC_a CIs are more accurate than percentile CIs (5, p.187).

Let $\theta^{*(\alpha)}$ indicate the $100 \cdot \alpha$th percentile of B bootstrap replications $\{\theta_1^*, \theta_2^*, \ldots, \theta_B^*\}$. The BC_a interval is then:

$$\left[\hat{\theta}_{lo}, \hat{\theta}_{up}\right] \approx \left[\theta_b^{*(\alpha_1)}, \theta_b^{*(1-\alpha_2)}\right], \tag{3.103}$$

where

$$\alpha_1 = \Phi\left(\hat{z}_0 + \frac{\hat{z}_0 + z^{(\alpha/2)}}{1 - \hat{a}(\hat{z}_0 + z^{(\alpha/2)})}\right), \tag{3.104}$$

$$\alpha_2 = \Phi\left(\hat{z}_0 + \frac{\hat{z}_0 + z^{(1-\alpha/2)}}{1 - \hat{a}(\hat{z}_0 + z^{(1-\alpha/2)})}\right) \tag{3.105}$$

Here $\Phi(\cdot)$ is the standard normal cumulative distribution function and $z^{(\alpha)}$ is the 100αth percentile point of a standard normal distribution. The parameter $\hat{z}_0$ can be estimated as follows [5, p.186]:

$$\hat{z}_0 = \Phi^{-1}\left(\frac{\#\left\{\theta_b^* < \hat{\theta}\right\}}{B}\right). \tag{3.106}$$

$\hat{z}_0$ measures the median bias of θ^*, that is the discrepancy between the median θ^* and $\hat{\theta}$ in normal units.

There are various ways to compute acceleration $\hat{a}$. The easiest is given in terms of the jackknife values of a statistic $\hat{\theta} = t(z)$. Let $\hat{\theta}_{(i)} = t(z_{(i)})$, and define $\hat{\theta}_{(\cdot)} = \sum_{i=1}^{n} \hat{\theta}_{(i)}/n$ [5, p.186]. Then

$$\hat{a} = \frac{\sum_{i=1}^{n}\left(\hat{\theta}_{(\cdot)} - \hat{\theta}_{(i)}\right)^3}{6\left\{\sum_{i=1}^{n}\left(\hat{\theta}_{(\cdot)} - \hat{\theta}_{(i)}\right)^2\right\}^{3/2}}. \tag{3.107}$$

The quantity $\hat{a}$ is called the acceleration because it refers to the rate of change of the standard error $\hat{\theta}$ with respect to the true parameter θ, measured on the normalized scale.

3.7   Conclusions

In this part of the master thesis we have considered theoretical foundations and properties of the following methods of estimation, which can be used for analysis of method comparison data with replicated measurements:

1. ML estimation,

2. averaged Deming regression,

3. Linnet's weighted Deming regression,

4. Barnett's approach,

5. MM-estimation in case the test method contains matrix effect.

ML-estimates and estimates of the averaged Deming regression can be applied to data with constant standard deviations of measurement errors (look at section 2.4). Such data are also called "homoscedastic data". These two approaches yield very similar results in most cases, but the estimates of averaged Deming regression are not full ML estimates. Another difference is the order of estimation of slope β_1 and ratio of variances of measurement errors λ. Using the ML-method we first estimate the slope. Then, we use this estimation for calculation of λ. Using averaged Deming regression, first λ is estimated. Calculation of the estimate of slope is based on the value of the estimated λ.

Linnet's weighted Deming regression is a weighted version of averaged Deming regression. This approach is developed for the analysis of data with constant coefficient of variation of measurement errors (look at section 2.4). Such data we call "heteroscedastic data with constant CV". The estimates of Linnet's weighted Deming regression are calculated using an iterative procedure.

Barnett's method is developed for data with heteroscedastic measurement errors when CVs are not necessarily constant. For calculation of corresponding estimates an iterative procedure is suggested.

MM-estimation in case test test method contains matrix effect will the following be called "MM-method". This approach may also be used for the analysis of data without matrix effect. In this case the estimation of the variance of the matrix effect can be negative, but the estimates of regression coefficients should be close to the ML-estimates.

Averaged- and Linnet's weighted Deming regressions are recommended by CLSI EP9 for the analysis of method comparison data. That is why it is especially important to investigate reliability of

these methods. This analysis is conducted via Monte-Carlo simulation studies in chapter 5 of the master thesis.

The next important problem is the calculation of confidence intervals (CIs) for regression coefficients and systematical bias between two measurement procedures at some medical decision points (MDP). For averaged Deming regression, analytical formulas for calculation of the CIs are given. This formulas should only be used if model assumptions (normality and homoscedasticity of measurement errors) hold. If the assumptions are violated, resampling methods are recommended, especially the jackknife method. For the other regression methods we will only use resampling methods to estimate the variance. It is interesting to investigate the performance of the different methods of calculation of CIs of parameter estimates. This analysis is conducted via Monte-Carlo simulation studies in chapter 5 of the master thesis.

In the theoretical part some diagnostic plots are considered, which could be used to choose the appropriate regression model for the data. Examples of such plots are shown in chapter V of the master thesis.

Chapter 4

Simulation studies

In the theoretical part we have considered five methods for the estimation of regression coefficients for linear ME models (in brackets the short names are given, which are used in plots and tables for the identification of the methods):

1. maximum likelihood estimation (ML);

2. estimation via. averaged Deming Regression (ave.Deming, ADEM);

3. estimation via. Linnet's weighted Deming Regression (WDeming, WDEM);

4. estimation of method of moments (MM);

5. estimation via. Barnett's approach (Barnett).

Each of these methods has particular model assumptions:

- ML-estimation and averaged Deming regression assume homoscedastic measurement errors.

- Method of moments (MM) assumes homoscedastic measurement errors and allows for a matrix effect in the test method.

- Linnet's weighted Deming regression assumes heteroscedastic measurement errors with constant CV.

- Barnett's method allows for different variance of measurement errors within each item.

All methods assume the ratio of variances of measurement errors λ to be constant over the whole measurement range. In the following we will evaluate method performance

- in case data are consistent with all model assumptions,

- in case model assumptions are violated.

The first group of simulation studies, which are described in this part of the master thesis, were conducted to investigate the performance of estimates of regression coefficients for different data situations (simulations 1.1 - 1.8 [??]). The main attention was focused on the averaged Deming regression and Linnet's weighted Deming regression, which are recommended by CLSI EP-9, for the analysis of method comparison data. For example, it is interesting to see how close the estimates of the averaged Deming regression are to the full ML estimates. These simulations are also a good validation tool for the algorithms.

Furthermore we have considered different methods for calculation of confidence intervals for estimates of regression coefficients and bias at some medical decision point (in brackets the short names are given, which are used in plots and tables for the identification of the methods):

- analytical confidence intervals based on normal distribution assumption (availably only for averaged Deming regression) (an.1),

- analytical confidence intervals based on influence function theory combined with normal distribution assumption (availably only for averaged Deming regression) (an.2),

- Jackknife confidence intervals (jack),

- nonparametric bootstrap quantile confidence intervals (boot.q),

- nonparametric bootstrap BC_a confidence intervals (boot.BC_a).

The second group of simulation studies has explored the coverage of these confidence intervals combined with different methods of estimates of regression coefficients (simulations 2.1-2.3 [4.2]). The most important question to be answered by these simulations, is the coverage of the analytical confidence intervals of the second type (an.2) in case of homoscedastic data and the coverage of jackknife intervals for the averaged and Linnet's weighted Deming regression within different simulation scenarios. These methods are suggested by CLSI EP9 [18] for the analysis of method comparison data, that is why it has a high practical relevance.

4.1 Simulation of data

In this master thesis, 11 simulation studies were conducted. Table 4.1 contains the overview of the simulations 1.1 - 1.8. The parameters of the simulations are described in section 4.2. Table 4.2 contains parameters of simulations 2.1 - 2.3.

For each simulation study, the data sets were simulated using the following algorithm:

1. First, the true measurements of the comparative method were simulated for n = 100 items
$$\{\xi_i \,|\, \forall\, i = 1, 2, ..., 100 \quad \xi_i \sim N(\mu = 200(300), \sigma = 25)\}.$$

2. Matrix effects in the comparative method (simulation 1.5) and in the test method (simulation 1.4, simulation 1.5) were simulated as follows:

$$\tilde{\xi}_i = \xi_i + m_i, m_i \sim N(0, \sigma_m = s_{mx}),$$

$$\tilde{\eta}_i = a + b\xi_i + c_i, c_i \sim N(0, \sigma_c = s_{my}).$$

3. For data without matrix effect

$$\tilde{\xi}_i = \xi_i$$

$$\tilde{\eta}_i = a + b\xi_i$$

4. For each item i, r measurement replications of the methods X and Y were simulated as follows:

assume $m = min\,\{\xi_i \,|\, i = 1, 2, ..., 100\}$ then

$$X_{ij} = \tilde{\xi}_i + \delta_{ij}$$

$$Y_{ij} = \tilde{\eta}_i + \epsilon_{ij}, \quad i = 1, 2, \,...\,100; \quad j = 1, 2...r.$$

For simulations 1.1-1.5 and 2.1, homoscedastic errors were simulated as random values from the following distributions:

$$\delta_{ij} \sim N(0, \sigma_\delta = s_{0x}), \quad \epsilon_{ij} \sim N(0, \sigma_\delta = s_{0y})$$

For simulations 1.6-1.8 and 2.2, heteroscedastic errors with constant CV were simulated as follows::

$$\delta_{ij} \sim N(0, \sigma_\delta = s_{0x} + CV_x \,|\xi_i - m|), \quad \epsilon_{ij} \sim (0, \sigma_\epsilon = s_{0y} + CV_y \,|\xi_i - m|)$$

For simulation 2.3, errors with outliers were simulated as follows:

$$\delta_{ij} = \delta_{ij1} + \delta_{ij2}, \quad \delta_{ij1} \sim N(0, \sigma_\delta = s_{0x}), \quad \delta_{ij2} \sim Laplace(location = 0, scale = s_x)$$

$$s_x = \frac{1}{\sqrt{2}}(c_{0x} + c_{1x} \,|\xi_i - m|);$$

$$\epsilon_{ij} = \epsilon_{ij1} + \epsilon_{ij2}, \quad \epsilon_{ij1} \sim N(0, \sigma_\epsilon = s_{0y}), \quad \epsilon_{ij2} \sim Laplace(location = 0, scale = s_y)$$

$$s_y = \frac{1}{\sqrt{2}}(c_{0y} + c_{1y}|\xi_i - m|);$$

Note: Laplace or double-exponential distribution has heavier tails than the normal distribution. The Laplace density function is

$$f(x) = \frac{1}{2b}exp\left(-\frac{|x - a|}{b}\right),$$

where a is a location parameter and b a scale parameter. If the random variable Y has Laplace distribution, then $E(Y) = a$ and $Var(Y) = 2b^2$.

4.2 Investigation of performance of estimates

In order to explore performance of estimates of regression coefficients we conduct the following simulations:

Simulation	Model assumptions	Scope: to investigate the dependence between performance of estimates and
Simulation 1.1	homoscedastic data without matrix effect	number of replications and variance of measurement errors
Simulation 1.2	homoscedastic data without matrix effect	true slope b and intercept a
Simulation 1.3	homoscedastic data without matrix effect	true error ratio λ
Simulation 1.4	homoscedastic data with matrix effect in test method	strength of matrix effect
Simulation 1.5	homoscedastic data with equal matrix effect in both methods	strength of matrix effect
Simulation 1.6	heteroscedastic data without matrix effect	strength of heteroscedasticity
Simulation 1.7	heteroscedastic data without matrix effect	true slope b and intercept a
Simulation 1.8	heteroscedastic data without matrix effect	true error ratio λ

Table 4.1: Scopes and model assumptions of the simulations studies 1.1-1.8.

The performance of estimates was measured using **mean squared error** MSE. Let $\hat{\theta}$ be an estimation of parameter θ. Then

$$MSE(\hat{\theta}) = E\left[(\hat{\theta} - \theta)^2\right].$$

In order to estimate the MSE we simulate 2000 data sets with the same simulation parameters. Based on each data set we calculate estimates of regression parameters. Let $\hat{\theta}_i$ be an estimation of the parameter θ using the i-th data set. Then

$$\hat{MSE}(\hat{\theta}) = 2000^{-1} \Sigma_{i=1}^{2000} (\hat{\theta}_i - \theta)^2.$$

Furthermore we consider plots which show the median and the 2.5%- and 97.5%-quantile of the

distribution of estimates. Using these plots we can get a picture of bias and variance of the estimators.

If the median of the distribution of the estimators is at the true value of the simulated data, we assume the estimation is unbiased, if the distribution is symmetric, because the mean of a symmetric distribution is equal to the median. If the distribution is skewed the median and mean can lead to the different conclusions.

Simulation 1.1

Scope: Investigation of the dependence between the number of replications and the strength of variability of measurement errors and the performance of the estimates.

Model assumptions: Data with homoscedastic measurement errors without matrix effect.

Parameters of simulation

Parameter	fixed	varying
number of replications (nrepl)		2, 3, 5, 10
intercept a	0	
slope b	1	
λ	1	
$k = \frac{sd(\,measurement\ error\ of\ X\,)}{sd(\,true\ measurements\ of\ X\,)}$		0.02, 0.05, 0.1, 0.2, 0.3, 0.4
$m.x = \frac{sd(\,Matrix\ effect\ of\ X\,)}{sd(\,Measurement\ error\ of\ X\,)}$	0	
$m.y = \frac{sd(\,Matrix\ effect\ of\ Y\,)}{sd(\,Measurement\ error\ of\ Y\,)}$	0	
CV.x	0	
CV.y	0	

Note. The strength of the variability of the measurement errors is controled via

k = sd(measurement error of X)/sd(true measurements of X).

Expectations

1. We expect higher variance of the estimates with rising variance of the measurement errors.

2. We assume that the variance of estimates decreases with increasing number of measurement replications.

3. The homoscedastic data suit perfectly to the model assumptions of the averaged Deming regression and ML-method. That is why we expect these methods to have the best performance. The MM method should have higher variance. It is not clear whether this method yields biased estimates in case of no matrix effect in test method. The model assumptions of the Linnet's weighted Deming regression are violated here. We expect poorer results from this method. The model assumptions of Barnett's approach include the case with constant

errors. Yet this algorithm does not use this information. That is why we expect less efficient estimates than the estimates of ML-Method and averaged Deming regression.

Simulation results

1. As expected, the variances of the estimates tend to get smaller with increasing number of replications (see pictures 4.1-4.2).

2. As expected, the variances of the estimates tend to get bigger with increasing variance of the measurement errors (see figures 4.3-4.4).

3. Estimates of all methods except Barnett's approach are unbiased and have nearly the same variance.

4. Linnet's algorithm needs no more than 3 iterations until convergence, if the treshhold is fixed by 10^{-7} (look at the fig. 4.5).

5. The distribution of the estimates of the variance of the matrix effect using the MM-Method is symmetrical around zero (fig. 4.6). Thus, if we apply the MM-Method to the data without matrix effect in the test method, we can get reasonable estimates of regression coefficients and very small estimation of the variance of the matrix effect, which sometimes is negative. We can consider it as an indicator of absence of a matrix effect.
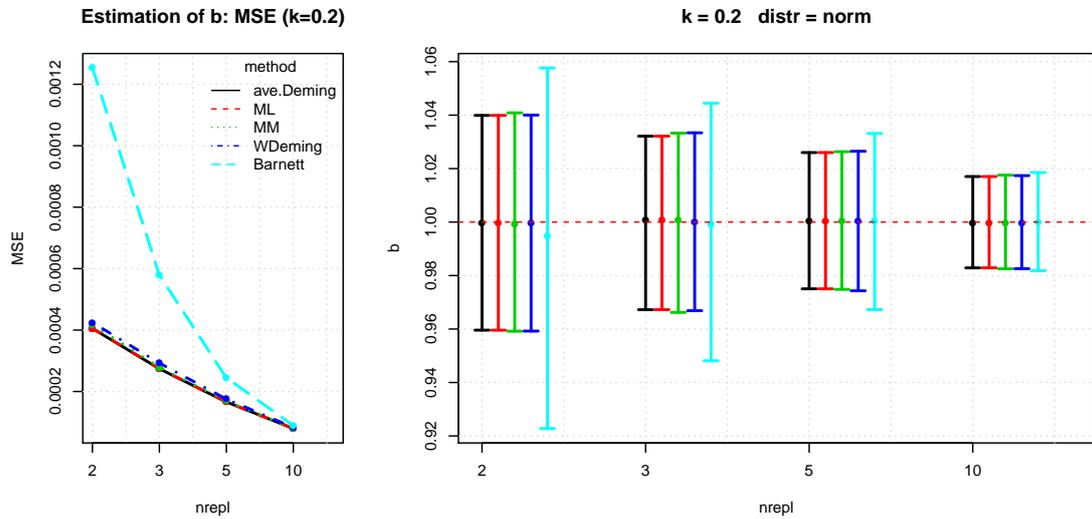
Figure 4.1: **Simulation 1.1.** Estimation of slope vs. number of measurement replications: MSE (left) and plot with median, 2.5% and 97.5%-quantiles of the distribution of estimates. k is fixed at 0.2.
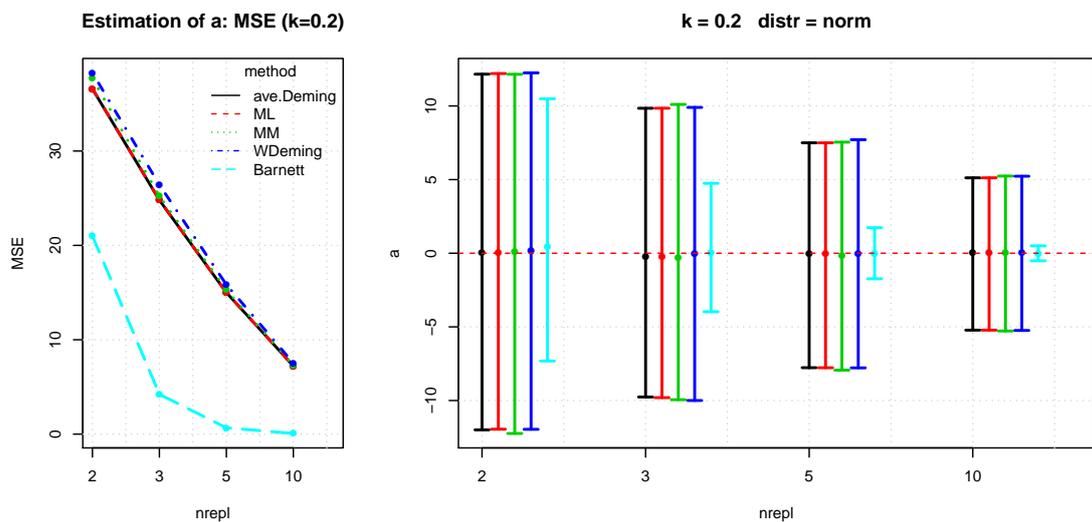
Figure 4.2: **Simulation 1.1.** Estimation of intercept vs. number of measurement replications: MSE (left) and plot with median, 2.5% and 97.5%-quantiles of the distribution of estimates. k is fixed at 0.2.
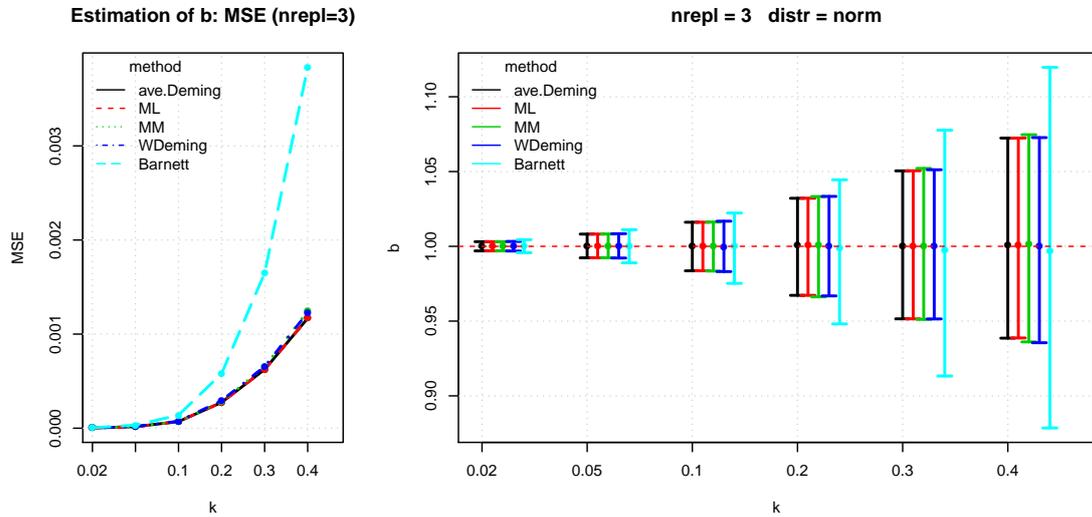
Figure 4.3: **Simulation 1.1.** Estimation of slope vs. strength of the variance of measurement errors: MSE (left) and plot with median, 2.5% and 97.5%-quantiles of the distribution of estimates. Number of replications is fixed at 3.
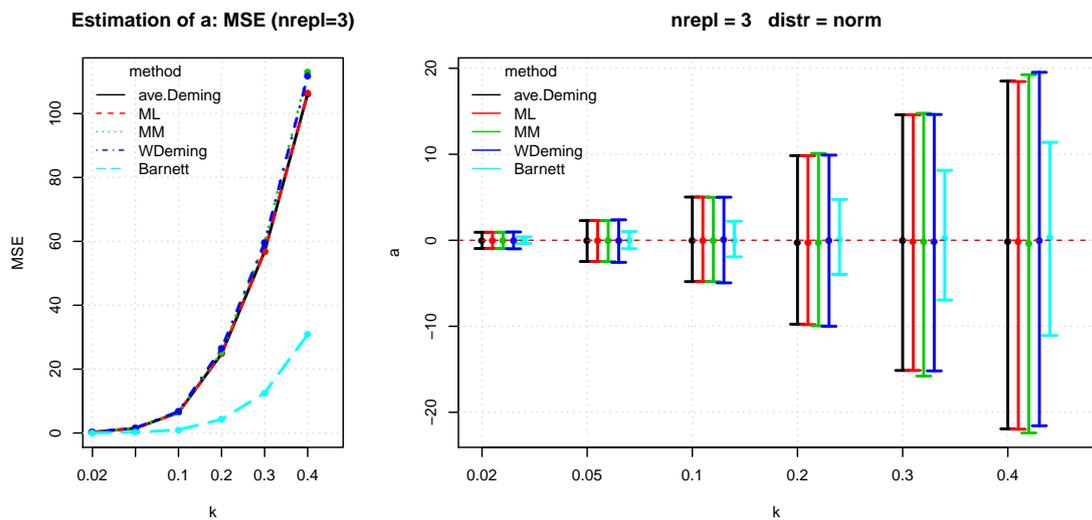
Figure 4.4: **Simulation 1.1.** Estimation of intercept vs. strength of the variance of measurement errors: MSE (left) and plot with median, 2.5% and 97.5%-quantiles of the distribution of estimates. Number of replications is fixed at 3.
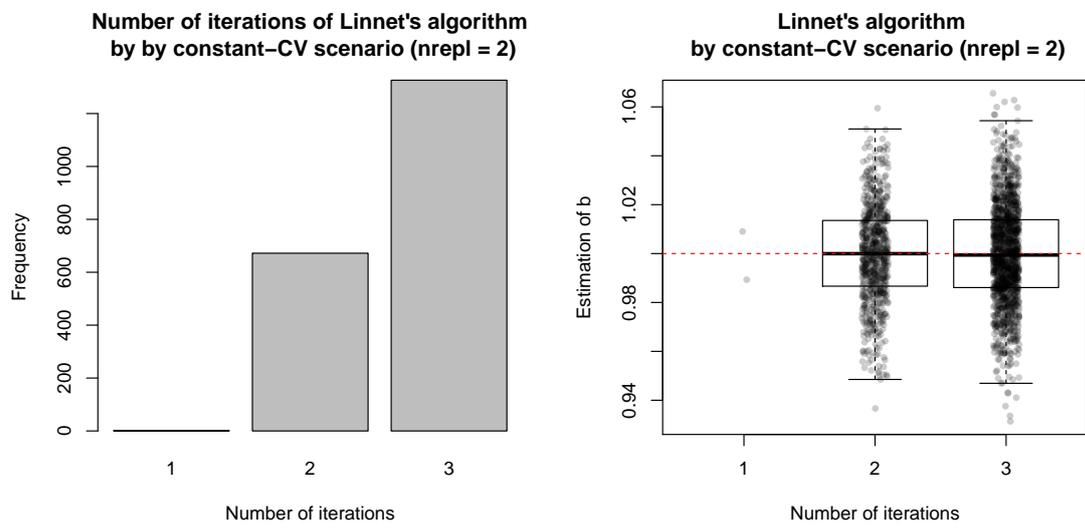
Figure 4.5: **Simulation 1.1.** Left: Number of iterations of the Linnet's weighted Deming algorithm which was applied to homoscedastic data. The algorithm converges mostly after 3 or 4 iterations. Right: Number of iterations vs. corresponding estimates of slope.
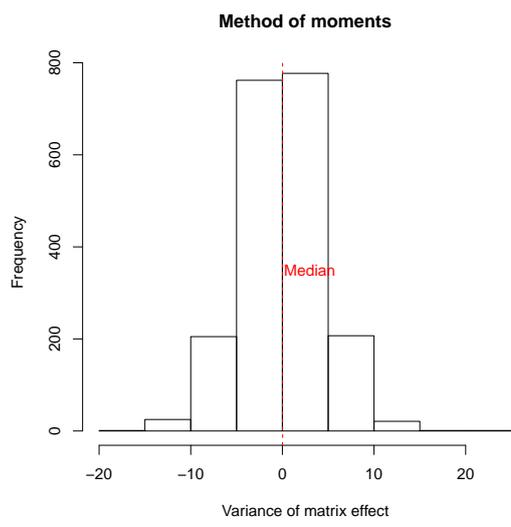
Figure 4.6: **Simulation 1.1.** Distribution of estimates of variance of the matrix effect calculated via MM-Method in case of data without matrix effect.

Simulation 1.2

Scope: Investigation of the dependence between the true slope or true intercept of the data and the performance of the estimates.

Model assumptions: Homoscedastic data without matrix effect.

Parameters of simulation

Parameter	fixed	varying
number of replications (nrepl)	2	
intercept a		0, 0.25, 2.5, 5, 25
slope b		0.5, 0.8, 1, 1.2, 2
λ	1	
$k = \frac{sd(\,measurement\ error\ of\ X\,)}{sd(\,true\ measurements\ of\ X\,)}$	0.2	
$m.x = \frac{sd(\,Matrix\ effect\ of\ X\,)}{sd(\,Measurement\ error\ of\ X\,)}$	0	
$m.y = \frac{sd(\,Matrix\ effect\ of\ Y\,)}{sd(\,Measurement\ error\ of\ Y\,)}$	0	
CV.x	0	
CV.y	0	

Expectations

1. The true slope and the true intercept should not affect the bias of estimates of regression parameters.

2. The variance of estimates is probably dependent on the value of the slope (for example, we can see a functional dependency between the slope and the variance of the estimates in formulas for analytical calculation of variances 2.27-2.28).

3. The homoscedastic data suit perfectly to the model assumptions of the averaged Deming regression and ML-method. That is why we expect that the estimates of this methods will have the lowest variances. MM estimates may have higher variances. The model assumptions of the Linnet's weighted Deming regression are violated here. We expect poorer results of this method. The model assumptions of Barnett's approach include the case with constant errors. Yet this algorithm does not use this information. That is why we expect less efficient estimates than the estimates of ML-Method and averaged Deming regression.

Simulation results

1. Estimates of all methods except Barnett's approach are unbiased independent of the values of the true slope or/and intercept (fig. 4.7, top).

2. Estimates of Barnett's approach are biased. The bias depends on the value of the true intercept: the bigger the true intercept the bigger the bias (fig. 1.7, top). The bias is

negative. This drawback of the actual implementation of Barnett's approach makes it useless for the analysis of method comparison data. That is why we do not consider this method in following simulations.

3. As expected, the variance of the estimates of the regression coefficients is affected by the true slope and is not affected by the true intercept (fig. 4.7).

4. The averaged Deming regression and the ML-estimates have the same variance.

5. The variances of the estimates of the MM-Method and of Linnet's weighted Deming regression are slightly higher than the variances of the ML estimates and the estimates of the averaged Deming regression.

Note. The estimates of Barnett's method can be improved by centering the data as following (see fig. 4.8):

1. First center all measurements as
$x^c_{ij} \equiv x_{ij} - x_{i.}, \; y^c_{ij} \equiv y_{ij} - x_{i.},$
i = 1,2, ... , n , j = 1,2, ... ,r.

2. Calculate the estimate of the slope $\hat{\beta}_1$ for the centered data.

3. Estimate the intercept using the original data as following: $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$.

The variance of the estimates of the improved Barnett's method is still considerably higher than the variance of the other estimates, but the bias becomes smaller.
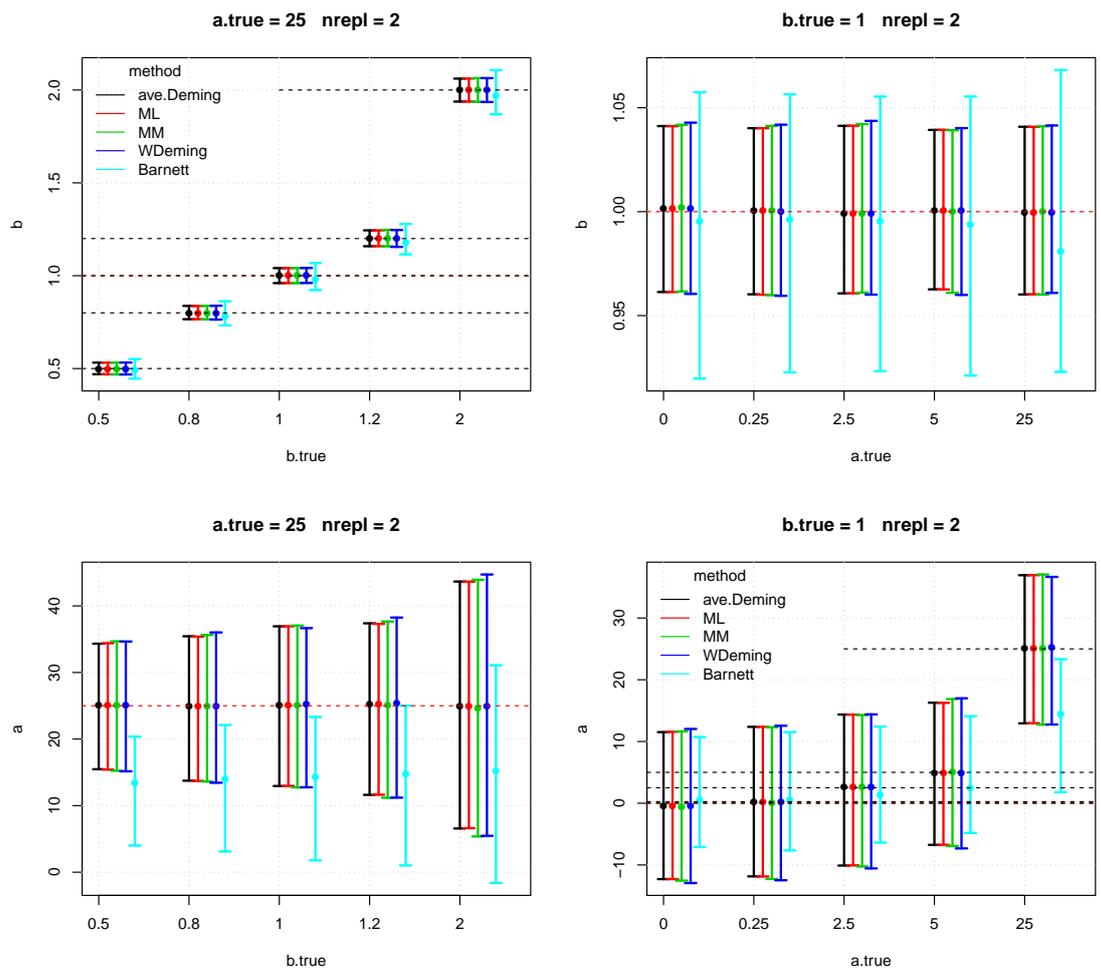
Figure 4.7: **Simulation 1.2.** Distribution of estimates of slope (top) and intercept (bottom) depending on the true values of slope (left) and intercept (right). Dotted horizontal lines show true values of parameters.
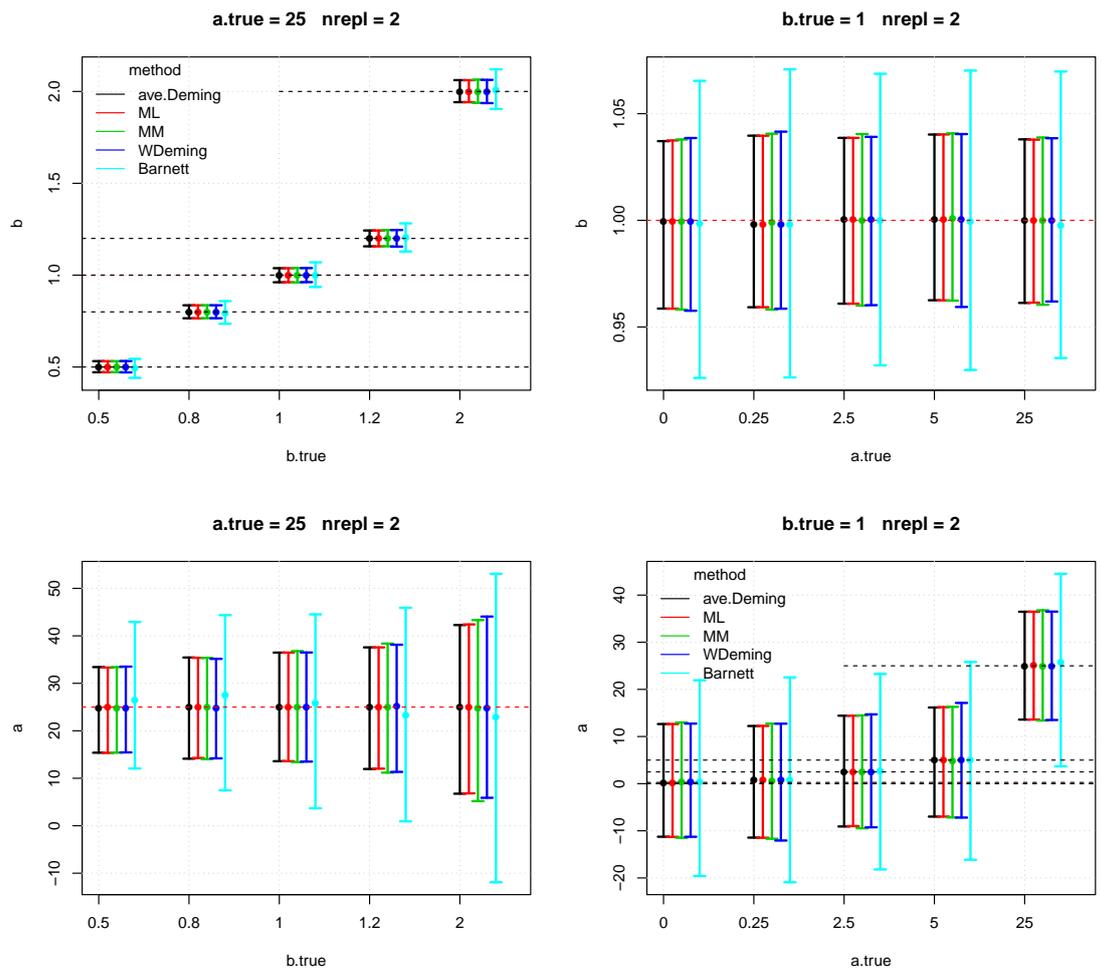
Figure 4.8: **Simulation 1.2a.** Results of simulation by using improved Barnett's procedure. Distribution of estimates of slope (top) and intercept (bottom) depending on the true values of slope (left) and intercept (right). Dotted horizontal lines depict true values of parameters.

Simulation 1.3

Scope: Investigation of the dependence between the true error ratio λ of the data and the performance of the estimates.

Model assumptions: Homoscedastic data without matrix effect.

Parameters of simulation

Parameter	fixed	varying
number of replications (nrepl)		2, 3
intercept a	0	
slope b	1	
λ	0.5, 0.8, 1, 1.2, 1.5, 2	
$k = \frac{sd(\,measurement\ error\ of\ X\,)}{sd(\,true\ measurements\ of\ X\,)}$	0.2	
$m.x = \frac{sd(\,Matrix\ effect\ of\ X\,)}{sd(\,Measurement\ error\ of\ X\,)}$	0	
$m.y = \frac{sd(\,Matrix\ effect\ of\ Y\,)}{sd(\,Measurement\ error\ of\ Y\,)}$	0	
CV.x	0	
CV.y	0	

Expectations

1. The true λ should not affect the bias of estimates of regression parameters.

2. The variance of the estimates of parameters should depend on the value of λ (the estimation of b is dependent on λ, and the variance of estimates is affected by estimation of b, look at formulas 2.27-2.28).

3. The homoscedastic data suit perfectly to the model assumptions of the averaged Deming regression and ML-method. We expect that this methods will have the lowest variances. The MM method may have higher variance. The model assumptions of the Linnet's weighted Deming regression are violated here. We expect poorer results of this method.

Simulation results

1. The results for 2 and 3 replications are similar (not shown).

2. The results of all methods are rather similar. The variances of the estimates of the MM-method and of Linnet's weighted Deming regression are slightly larger than the variances of ML-estimates and the estimates of the averaged Deming regression.

3. The true does not affect the bias of estimates of all methods: they stay unbiased (fig. 4.9-4.10).

4. The variances of the estimates tend to get higher if λ grows. The closeness to $\lambda = 1$ is irrelevant (fig. 4.9-4.10).

5. The distribution of the estimates of λ is slightly skewed with median at the true value (fig. 4.11).
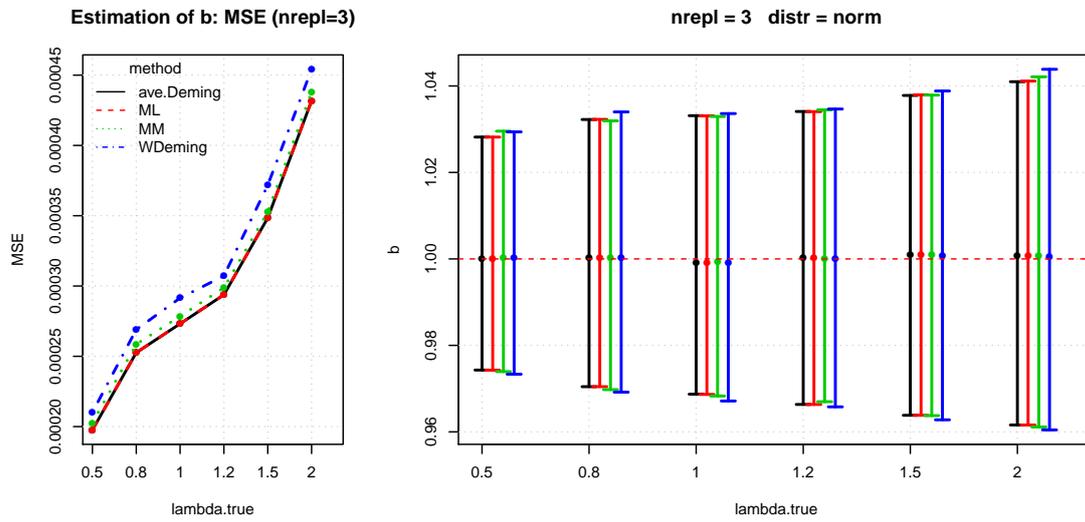
Figure 4.9: **Simulation 1.3.** Estimation of slope vs. true values of λ. MSE (left) and distribution of the estimates (median, 2.5%-quantile and 97.5%-quantile).
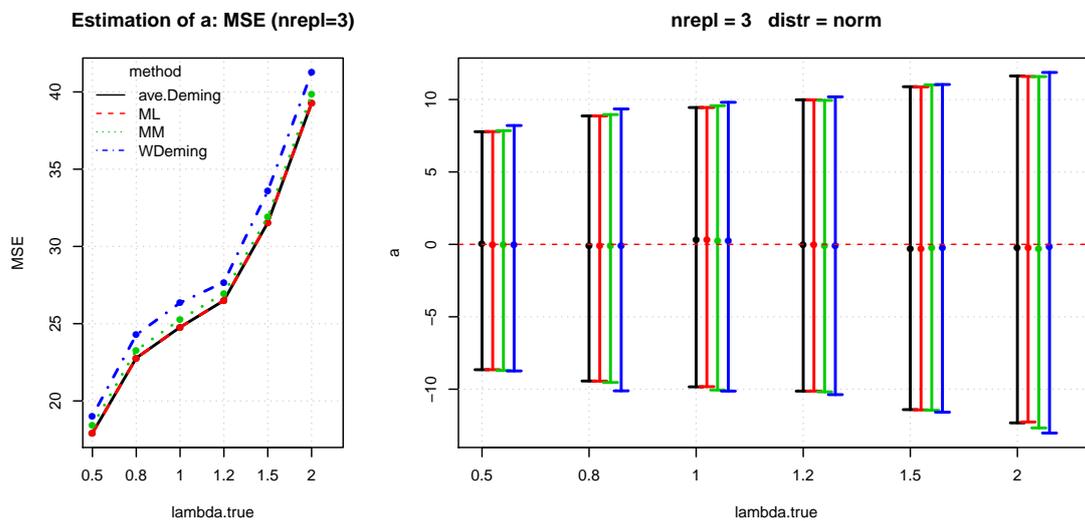
Figure 4.10: **Simulation 1.3.** Estimation of intercept vs. true values of λ. MSE (left) and distribution of the estimates (median, 2.5%-quantile and 97.5%-quantile).

nrepl = 3   distr = norm
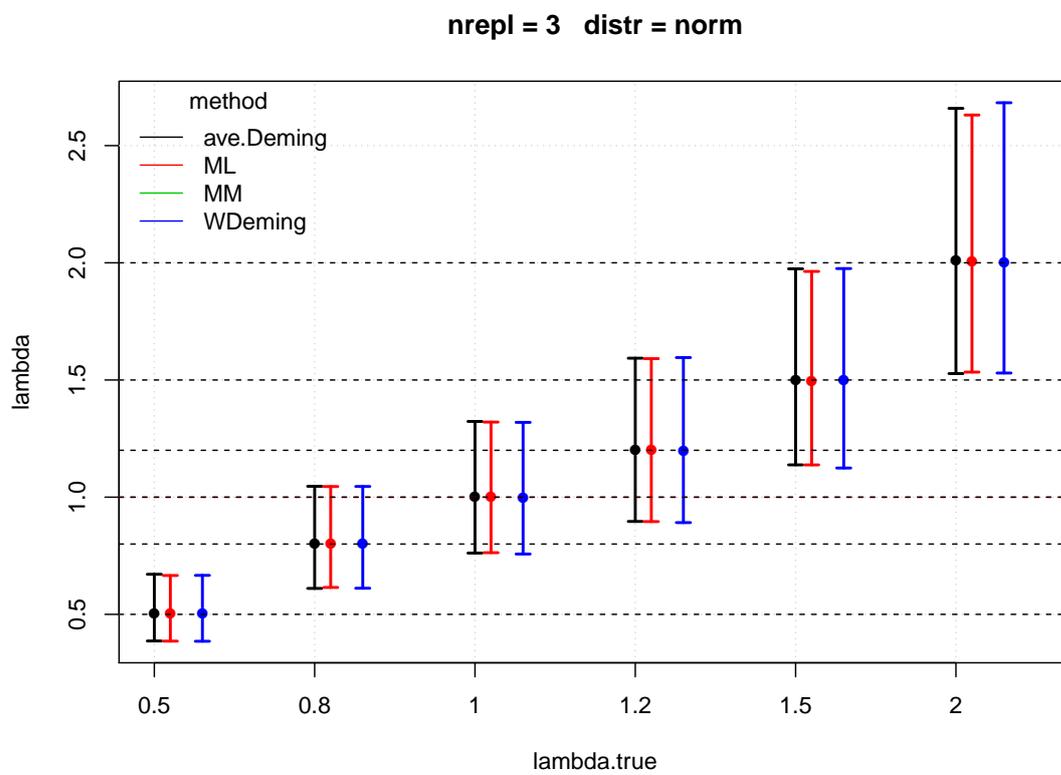
Figure 4.11: **Simulation 1.3.** Estimation of lambda vs. true value of lambda (dotted horizontal lines). Distribution of the estimates (median, 2.5%-quantile and 97.5%-quantile).

Simulation 1.4.

Scope: Investigation of the influence of a matrix effect in the test method on the performance of estimates.

Model assumptions: Homoscedastic data with matrix effect in the test method Y.

Parameters of simulation

Parameter	fixed	varying
number of replications (nrepl)		2, 3, 5, 10
intercept a	0	
slope b	1	
λ	1	
$k = \frac{sd(\,measurement\ error\ of\ X\,)}{sd(\,true\ measurements\ of\ X\,)}$	0.1	
$m.x = \frac{sd(\,Matrix\ effect\ of\ X\,)}{sd(\,Measurement\ error\ of\ X\,)}$	0	
$m.y = \frac{sd(\,Matrix\ effect\ of\ Y\,)}{sd(\,Measurement\ error\ of\ Y\,)}$		0.5, 1, 2, 5, 10
CV.x	0	
CV.y	0	

Note. The strength of matrix effect is controlled via

m.y = sd(Matrix effect of Y) / sd(Measurement error of Y).

Figure 4.12 shows examples of the corresponding simulated data. We can speak about strong matrix effect if m.y = 5 and m.y = 10. The corresponding scatter plots contain clear clusters of replications.

Expectations

1. The homoscedastic data with a matrix effect in the test method Y suit perfectly to the model assumptions of the MM-method. That is why we expect the best results from this method.

2. A weak matrix effect should not considerably affect the estimates of the ML-method, averaged Deming regression or weighted Deming Regression. We expect biased estimates of these methods if the matrix effect is strong.

3. The matrix effect introduces additional variability to the data. That is why we expect a higher variance of the estimates if the matrix effect gets larger.

57

Simulation results

1. The simulation shows that the MM-Method is the only method which yields unbiased estimates in case of a strong matrix effect in the test method (fig. 4.13-4.16).

2. The number of replications seem does not to affect the variance of estimates (fig. 4.15-4.16).

3. All methods yield proper results if the matrix effect is small (the variance of the matrix effect is lower or equal to the variance of the measurement errors).

4. If the matrix effect is strong, the ML-estimates and the estimates of the averaged and Linnet's weighted Deming Regression are considerably biased. The bias of the estimates of slope is negative, the bias of the estimates of intercept is positive. The larger the matrix effect the larger the bias (fig. 1.13-1.14).

5. This is the first simulation scenario which produces very different results of the ML- and averaged Deming regression estimates. The difference increases with the strength of the matrix effect. In case of a strong matrix effect (m.y = 10), the ML estimates are less biased and have smaller variance than the estimates of the averaged Deming regression(fig. 4.16).

6. The MM-Method yields reasonable estimates of the variance of the matrix effect. The distribution of estimates is slightly skewed. The median of the estimates is close to the true parameters.
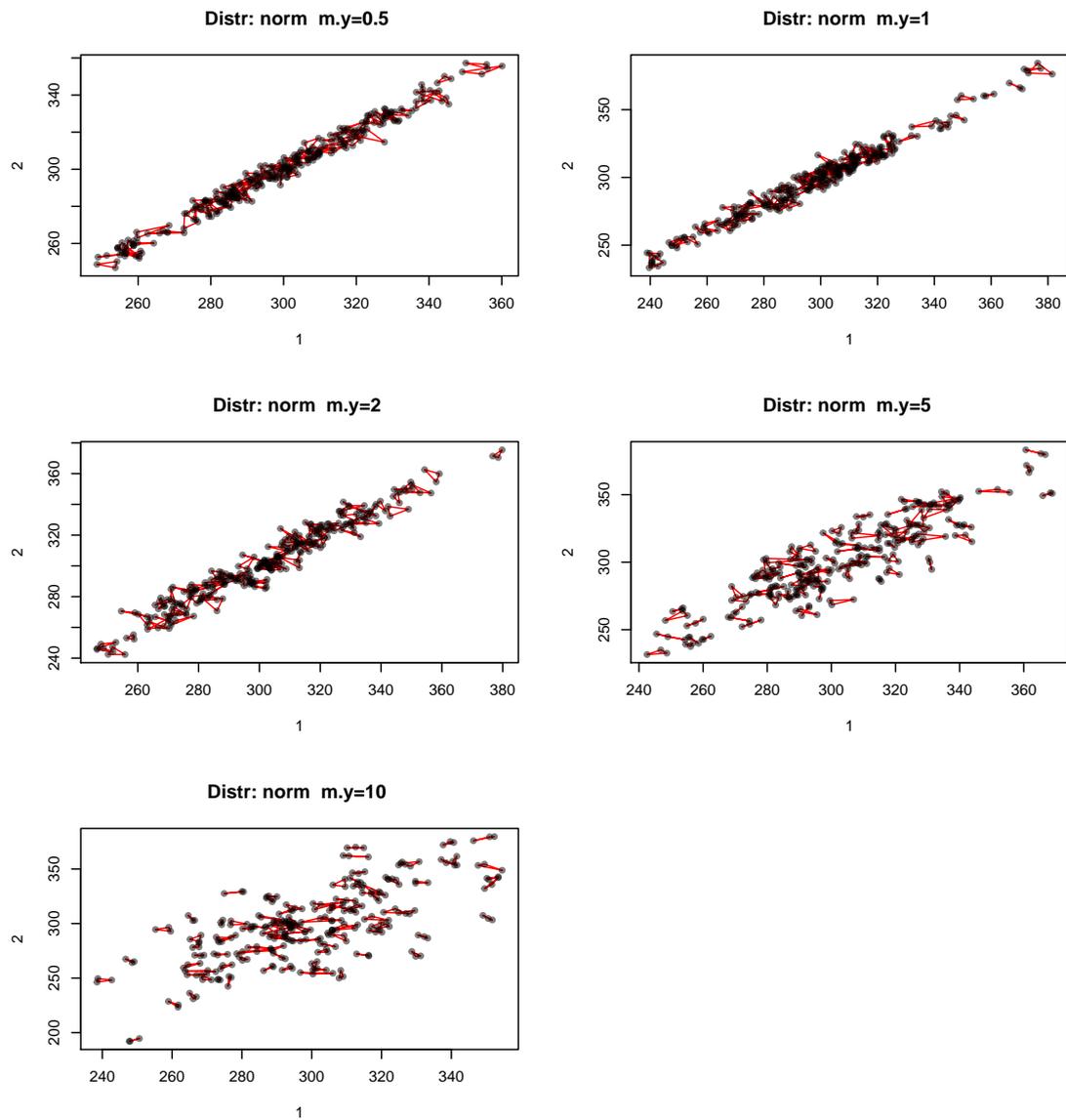
Figure 4.12: **Simulation 1.4.** Examples of simulated data with a matrix effect in the test method. Number of replications nrepl = 3.
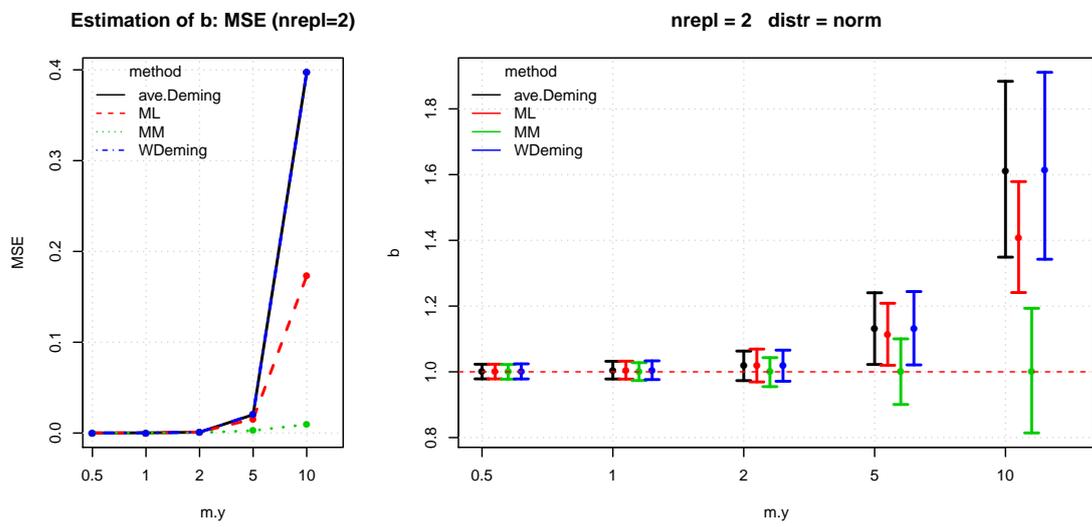
Figure 4.13: **Simulation 1.4.** Estimation of slope vs. strength of matrix effect in test method: MSE (left) and plot with median, 2.5% and 97.5%-quantiles of the distribution of estimates. k is fixed at 0.1, number of replications nrepl = 2.
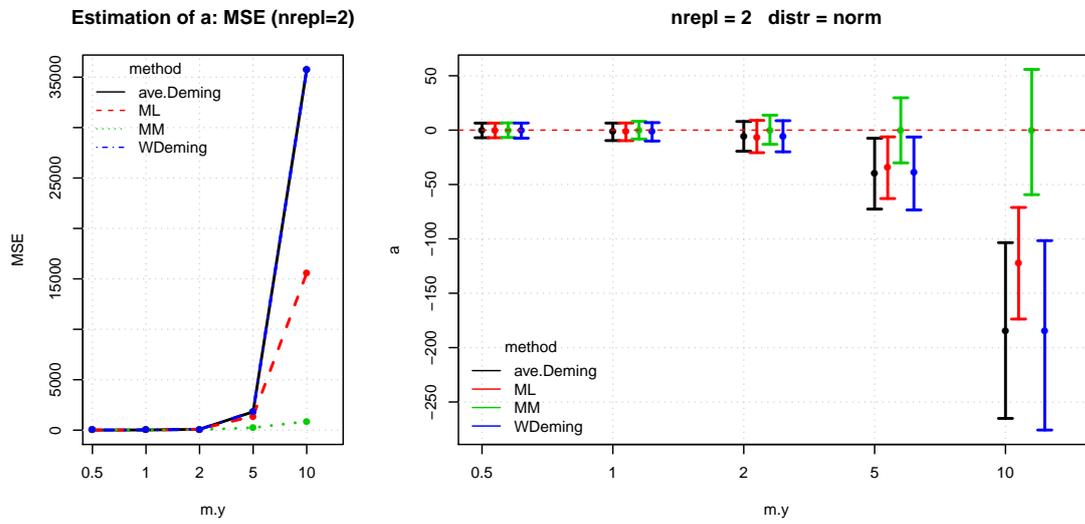
Figure 4.14: **Simulation 1.4.** Estimation of intercept vs. strength of matrix effect in test method: MSE (left) and plot with median, 2.5% and 97.5%-quantiles of the distribution of estimates. k is fixed at 0.1, number of replications nrepl = 2.
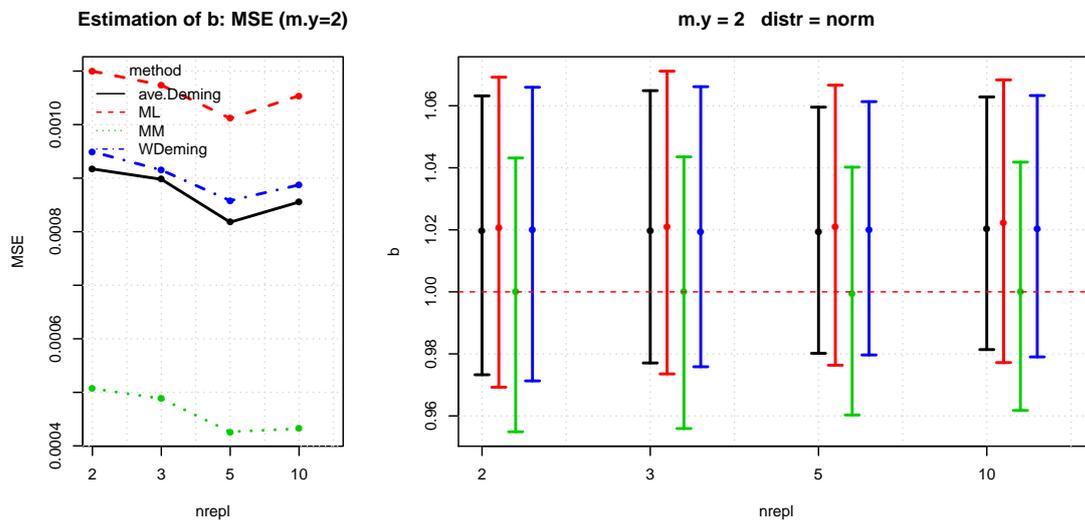
Figure 4.15: **Simulation 1.4.** Estimation of slope vs. number of replications: MSE (left) and plot with median, 2.5% and 97.5%-quantiles of the distribution of estimates. k is fixed at 0.1, m.y = 2.
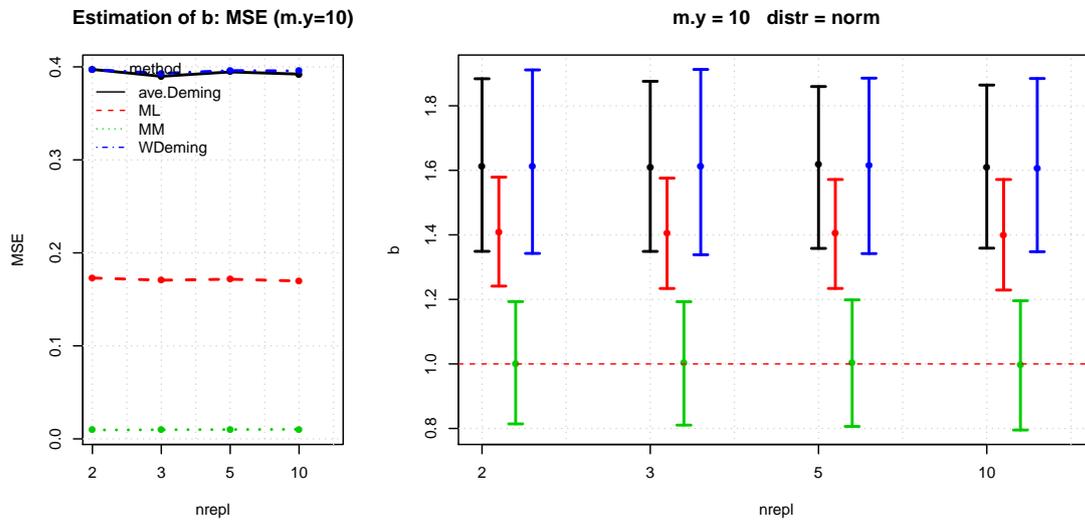
Figure 4.16: **Simulation 1.4.** Estimation of slope vs. number of replications: MSE (left) and plot with median, 2.5% and 97.5%-quantiles of the distribution of estimates. K is fixed at 0.1, m.y = 10.
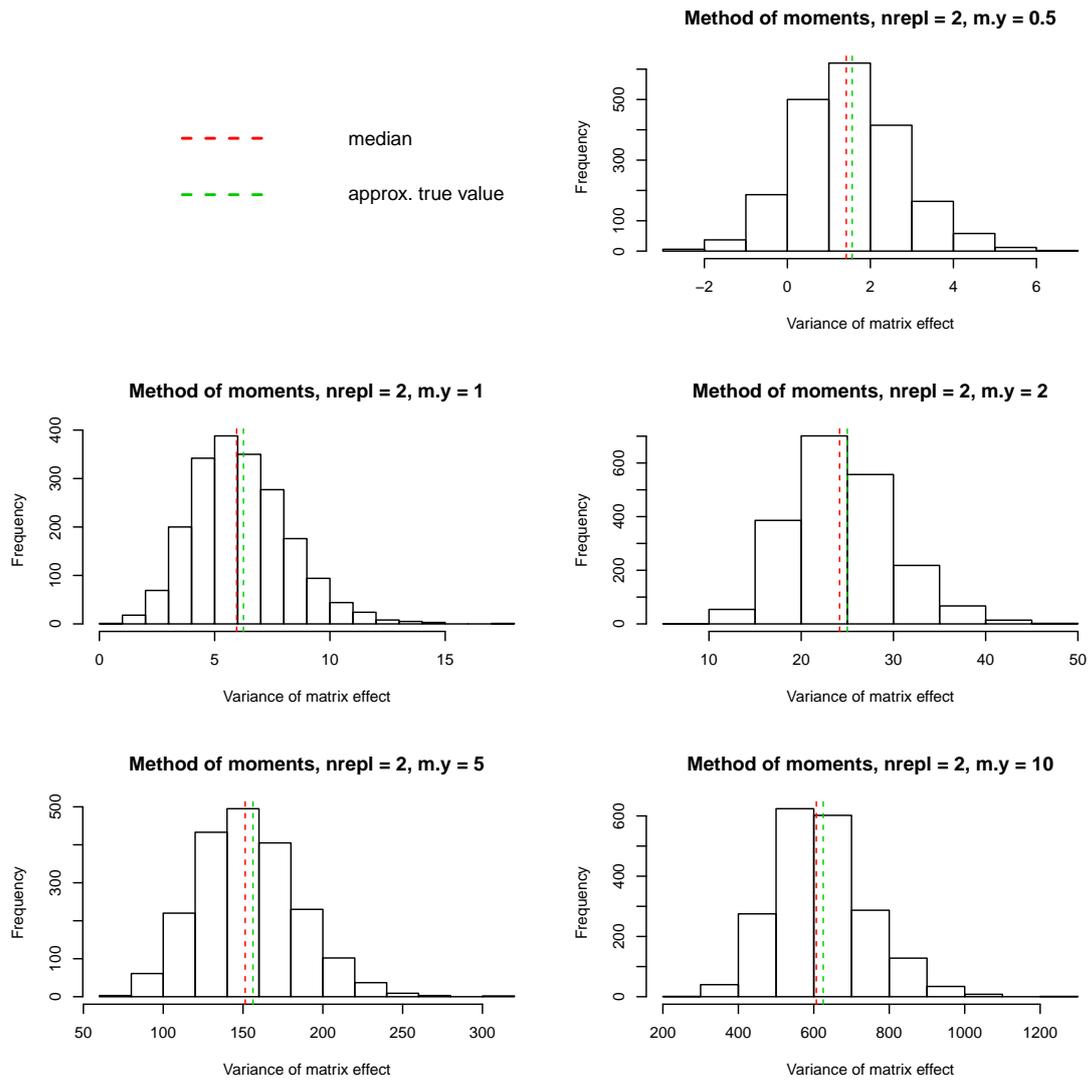
Figure 4.17: **Simulation 1.4.** Estimates of the variance of matrix effect. The true value of the variance of matrix effect is a random number which depends of the distribution of sampled true X-measurements. The approximate value is $(25 \cdot k \cdot m.y)^2$. k is fixed by 0.1.

Simulation 1.5.

Scope: Investigation of the influence of the matrix effect in both methods on performance of the estimates.

Model assumptions: Homoscedastic data with a matrix effect in both methods X and Y.

Parameter	fixed	varying
number of replications (nrepl)		2, 3, 5, 10
intercept a	0	
slope b	1	
λ	1	
$k = \frac{sd(\,measurement\ error\ of\ X\,)}{sd(\,true\ measurements\ of\ X\,)}$	0.1	
$m.x = \frac{sd(\,Matrix\ effect\ of\ X\,)}{sd(\,Measurement\ error\ of\ X\,)}$		0.5, 1, 2, 5, 10
$m.y = \frac{sd(\,Matrix\ effect\ of\ Y\,)}{sd(\,Measurement\ error\ of\ Y\,)}$		0.5, 1, 2, 5, 10, m.y = m.x
CV.x	0	
CV.y	0	

Note. Fig. 4.17 shows example data with equal matrix effect in both methods.

Expectations

In case of a matrix effect in both methods the model assumptions of all methods are violated. It is not clear, which method will yield better results.

Simulation results

1. In case of a strong matrix effect in both methods the estimates of the MM-Method are considerably biased (fig. 4.20).

2. The MM-estimation of the variance of the matrix effect in test method is overestimated (fig. 4.22).

3. The estimates of the ML-method, averaged Deming regression and weighted Deming regression are unbiased (fig. 4.19-4.20).

4. The variance of the estimates does not seem to be affected by the number of replications (fig. 4.19).

5. The variances of estimates tend to increase if the matrix effect gets stronger. The variances of the estimates of averaged and weighted Deming regressions grow much faster than the the variances of the ML-estimates. Hence for m.y = m.x = 10 the difference is rather high (fig. 4.20).

6. The ML-estimates of λ are unstable and have considerably larger variance than the estimates of other methods in case the variance of the matrix effect is close to the variance of measurement errors (fig. 4.21).
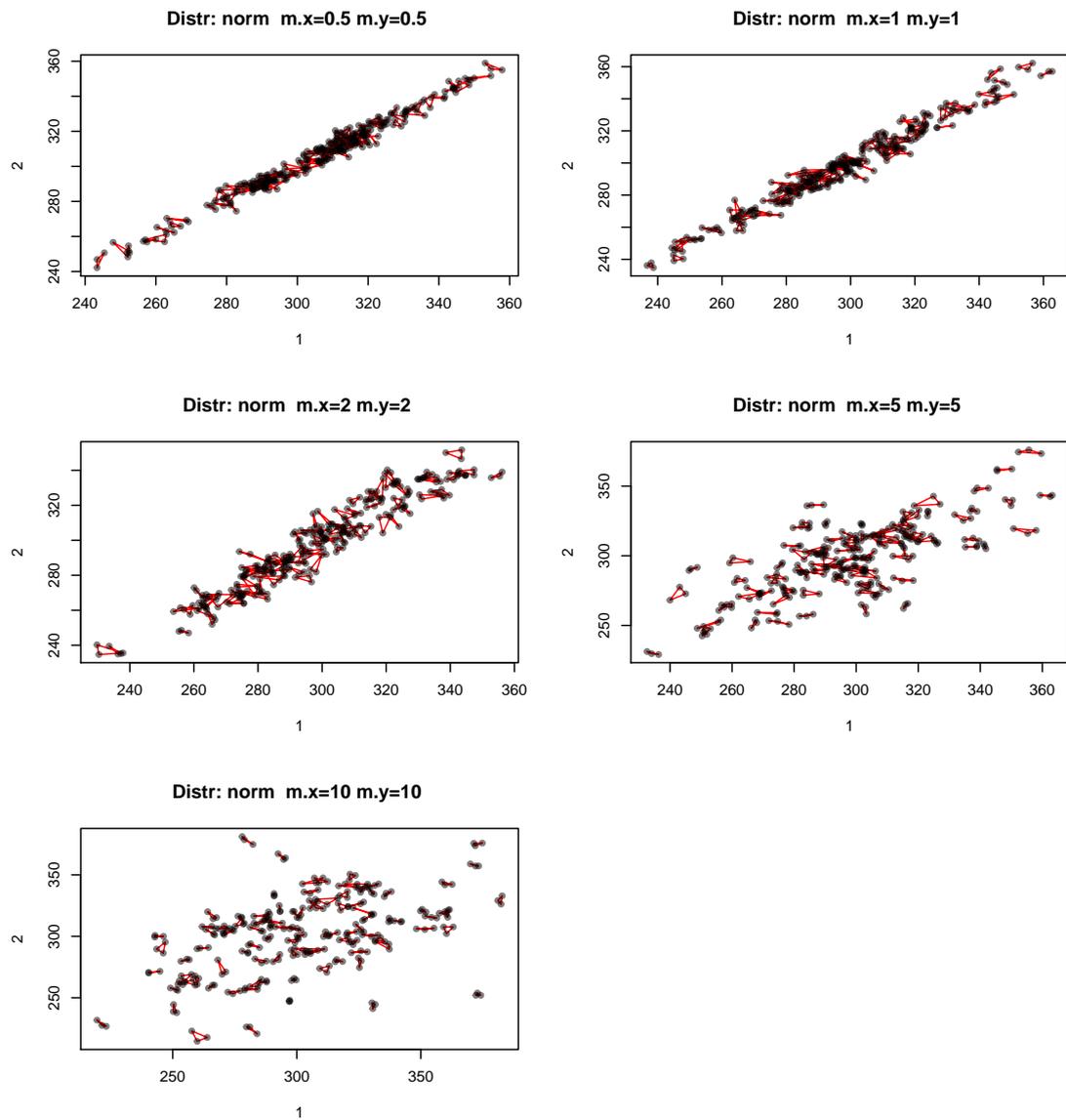
Figure 4.18: **Simulation 1.5.** Examples of simulated data with matrix effect in both methods. Number of replications nrepl = 3.
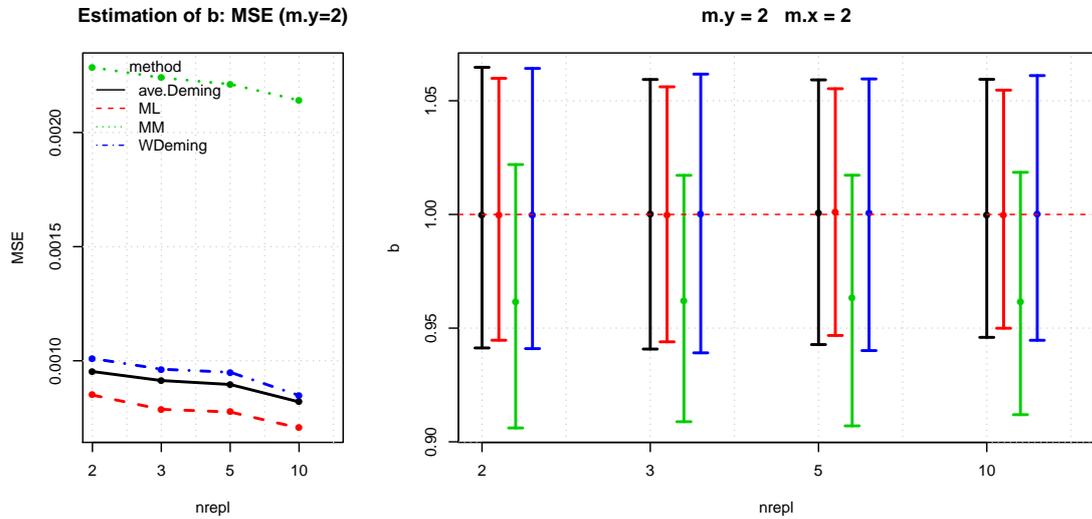
Figure 4.19: **Simulation 1.5.** Estimation of slope vs. number of replications: MSE (left) and plot with median, 2.5% and 97.5%-quantiles of the distribution of estimates. k is fixed at 0.1, m.x = m.y = 2.
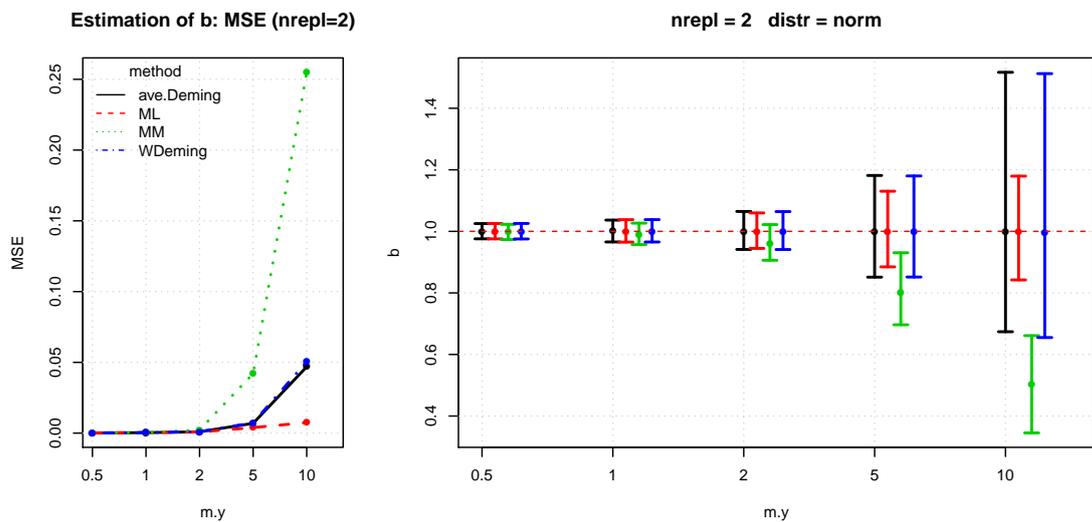
Figure 4.20: **Simulation 1.5.** Estimation of slope vs. strength of matrix effect in both methods: MSE (left) and plot with median, 2.5% and 97.5%-quantiles of the distribution of estimates. k is fixed at 0.1, number of replications nrepl = 2.
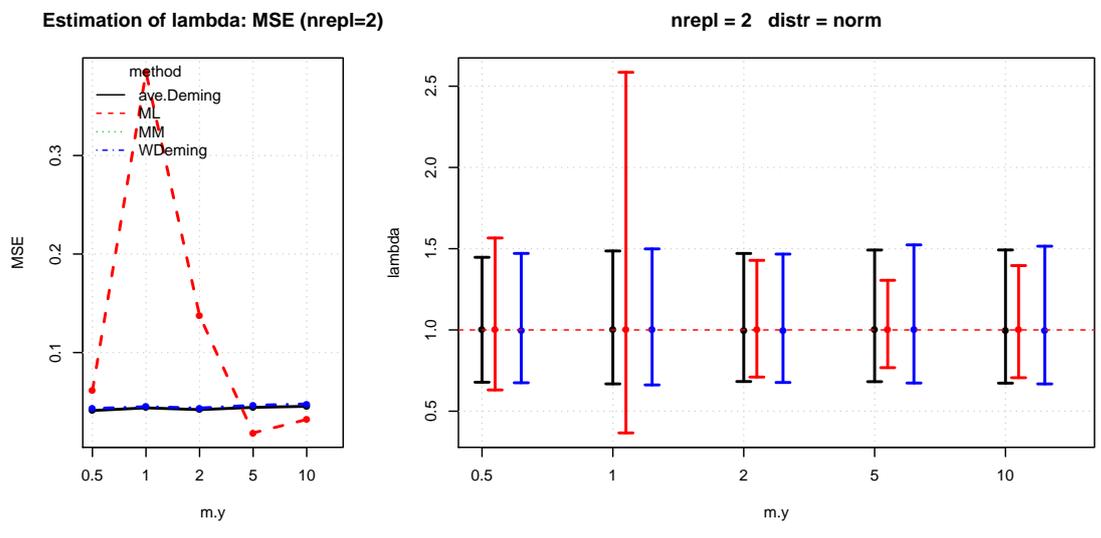
Figure 4.21: **Simulation 1.5.** Estimation of lambda vs. strength of matrix effect in both methods: MSE (left) and plot with median, 2.5% and 97.5%-quantiles of the distribution of estimates. k is fixed at 0.1, number of replications nrepl = 2.
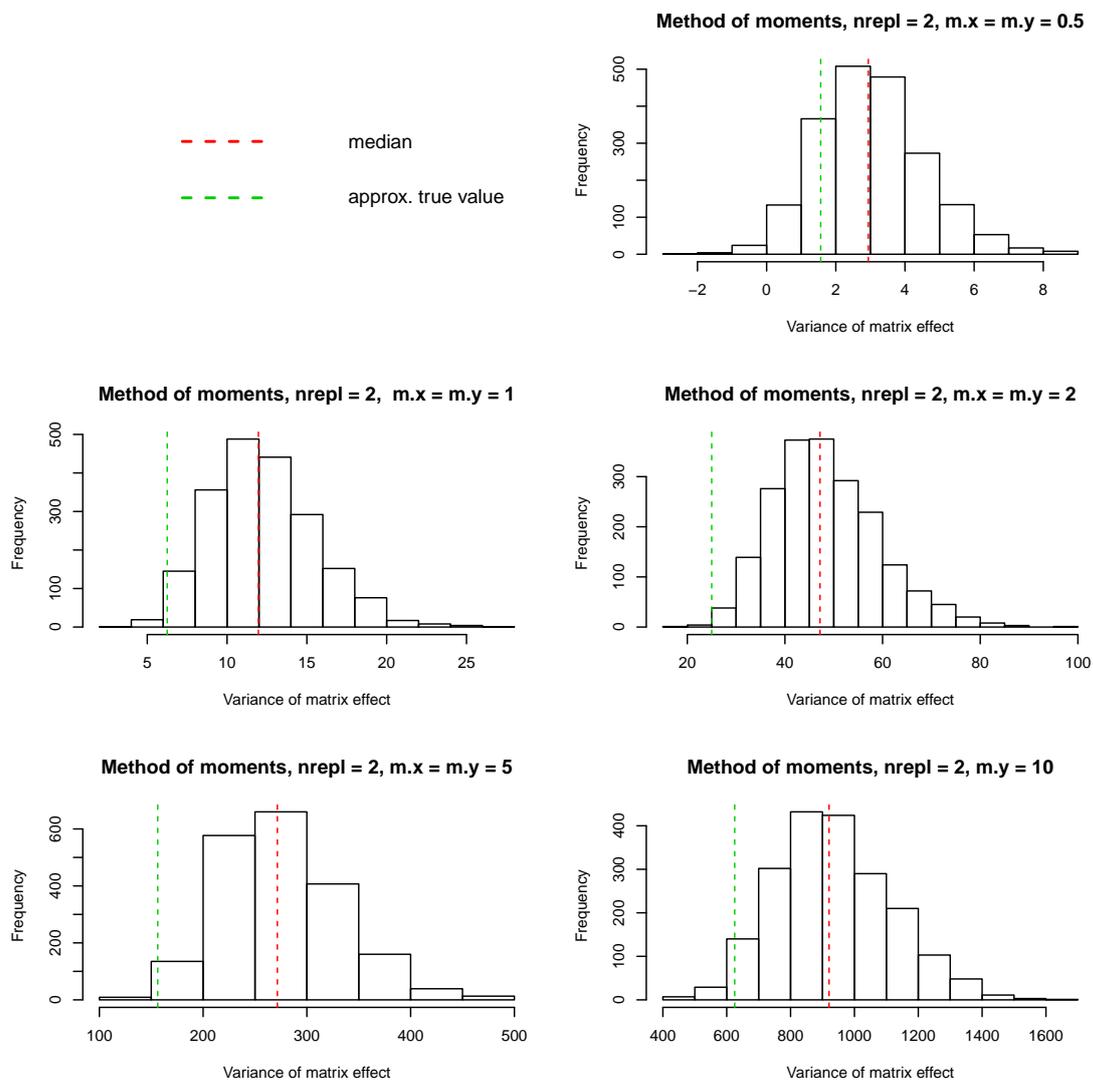
Figure 4.22: **Simulation 1.5.** MM-estimates of the variance of matrix effect in the test method in case of eqal matrix effect in both measurement methods X and Y. The true values are random numbers which depend of the distribution of sampled true X-measurements. The approximate value is $(25 \cdot k \cdot m.y)^2$. k is fixed at 0.1.

Simulation 1.6

Scope: Investigation of the dependence between the strength of heteroscedasticity and the performance of the estimates.

Model assumptions: Heteroscedastic data with constant CV without matrix effect.

Fixed parameters of simulation: Intercept a = 0, slope b = 1. Error ratio $\lambda = 1$.

Parameter	fixed	varying
number of replications (nrepl)		2, 3, 5
intercept a	0	
slope b	1	
λ	1	
$k = \frac{sd(\,measurement\ error\ of\ X\,)}{sd(\,true\ measurements\ of\ X\,)}$	0.1	
$m.x = \frac{sd(\,Matrix\ effect\ of\ X\,)}{sd(\,Measurement\ error\ of\ X\,)}$	0	
$m.y = \frac{sd(\,Matrix\ effect\ of\ Y\,)}{sd(\,Measurement\ error\ of\ Y\,)}$	0	
CV.x		0.05, 0.1, 0.15
CV.y		0.05, 0.1, 0.15; CV.y = CV.x

Note. Fig. 4.23 shows example data with equal matrix effect in both methods.

Expectations

1. The simulated data suits perfectly to the model assumptions of Linnet's weighted Deming regression. That is why we expect the best results from this method.

2. The ML- and MM-estimates and the estimates of the averaged Deming regression should not be biased by heteroscedastic data, but we expect higher variances of their estimates than for the estimates of the Linnet's weighted Deming regression.

Simulation results

1. As we have expected, the estimates of Linnet's weighted Deming regression, have the smallest variance (fig. 4.24-4.25). The number of iterations of Linnet's algorithm tends to get smaller with increasing number of replications and in most cases 3 or 4 iterations are needed (fig. 4.25).

2. The estimates of the ML-Method and averaged Deming regression are very similar. The estimates of the MM-method have slightly higher variance (fig. 4.24-4.25).
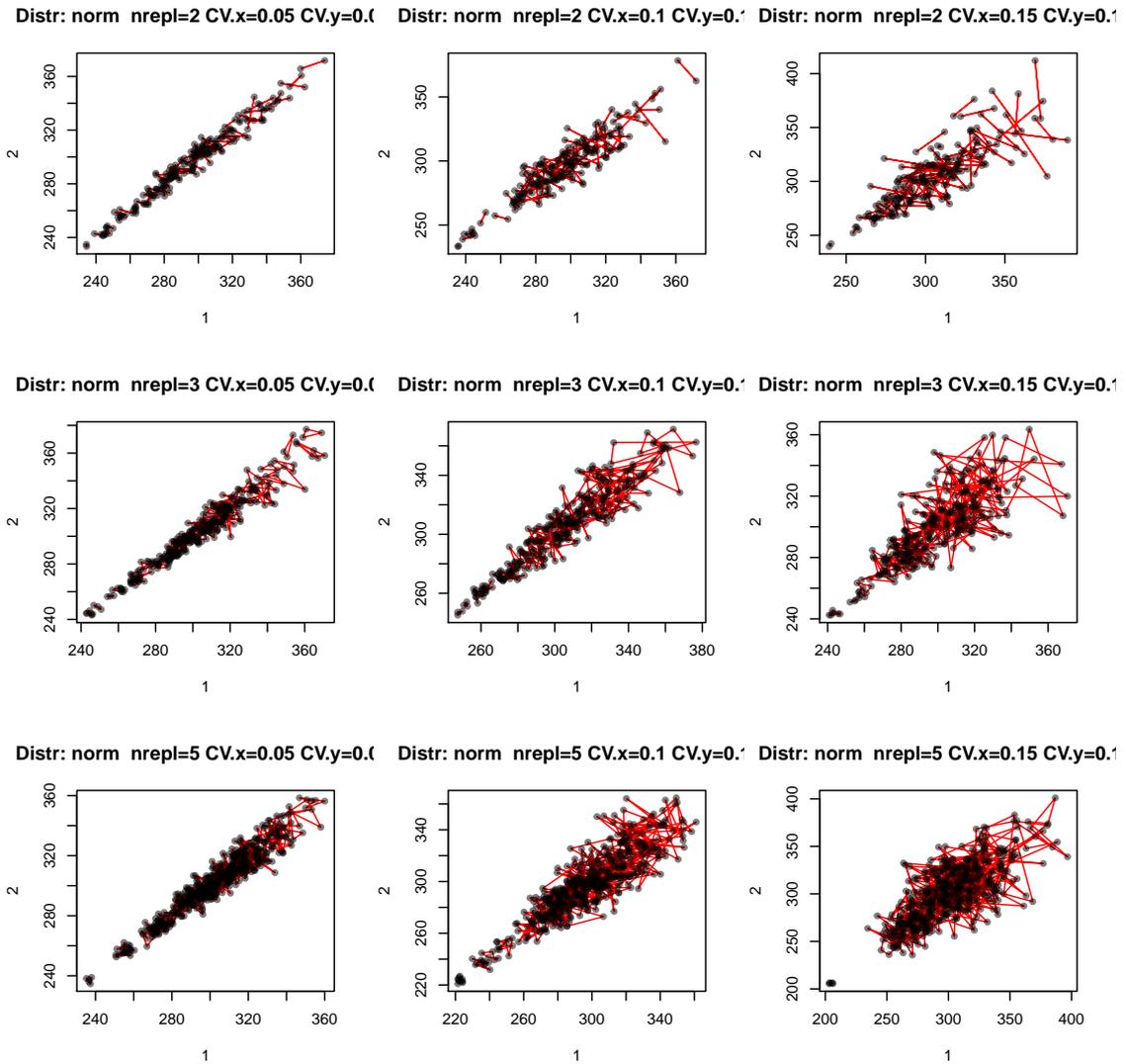
Figure 4.23: **Simulation 1.6.** Examples of simulated data with heteroscedastic measurement errors and constant CV.
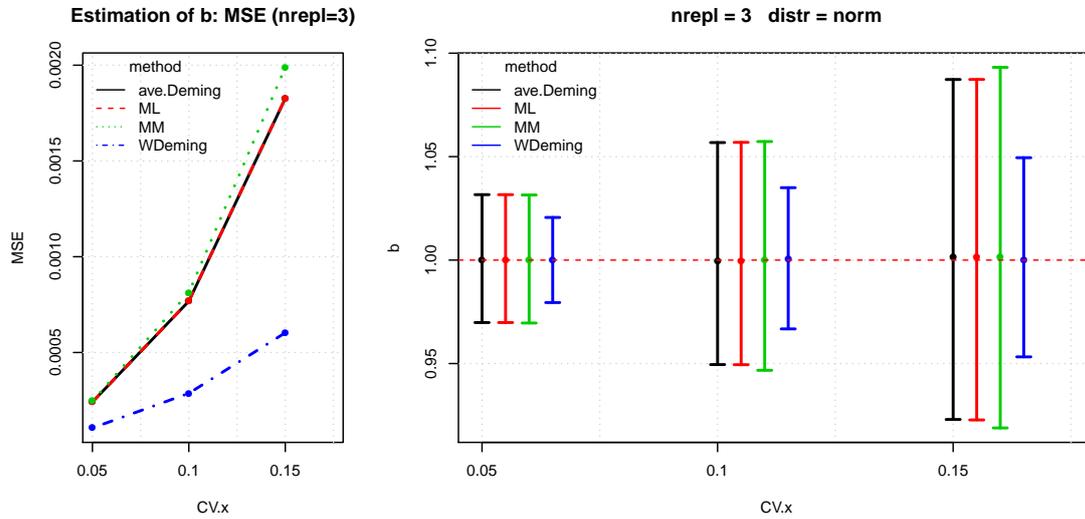
Figure 4.24: **Simulation 1.6.** Estimation of slope vs. CV of measurement errors. MSE (left) and distribution of the estimates (median, 2.5%-quantile and 97.5%-quantile). Please note: CV.x = CV.y.
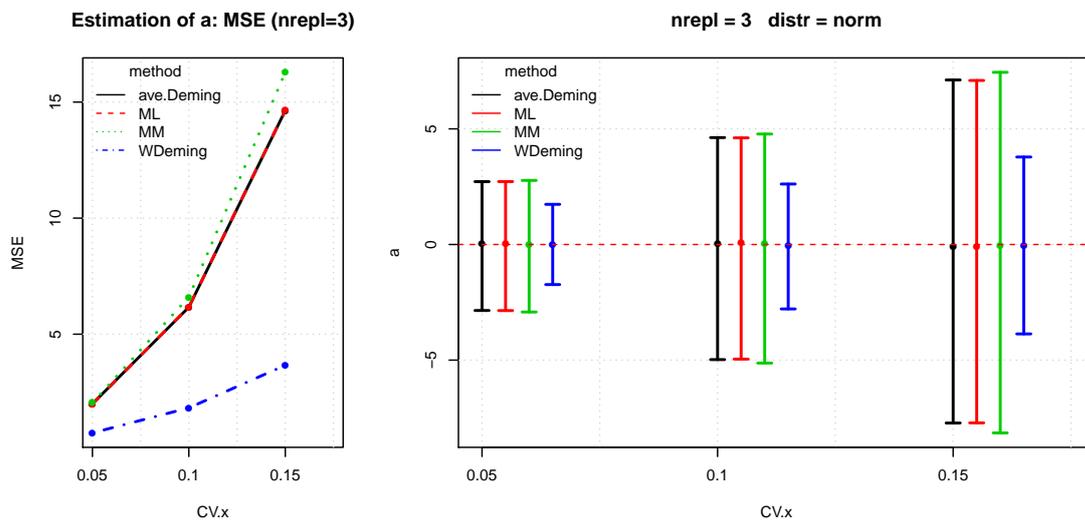
Figure 4.25: **Simulation 1.6.** Estimation of intercept vs. CV of measurement errors. MSE (left) and distribution of the estimates (median, 2.5%-quantile and 97.5%-quantile). Please note: CV.x = CV.y.
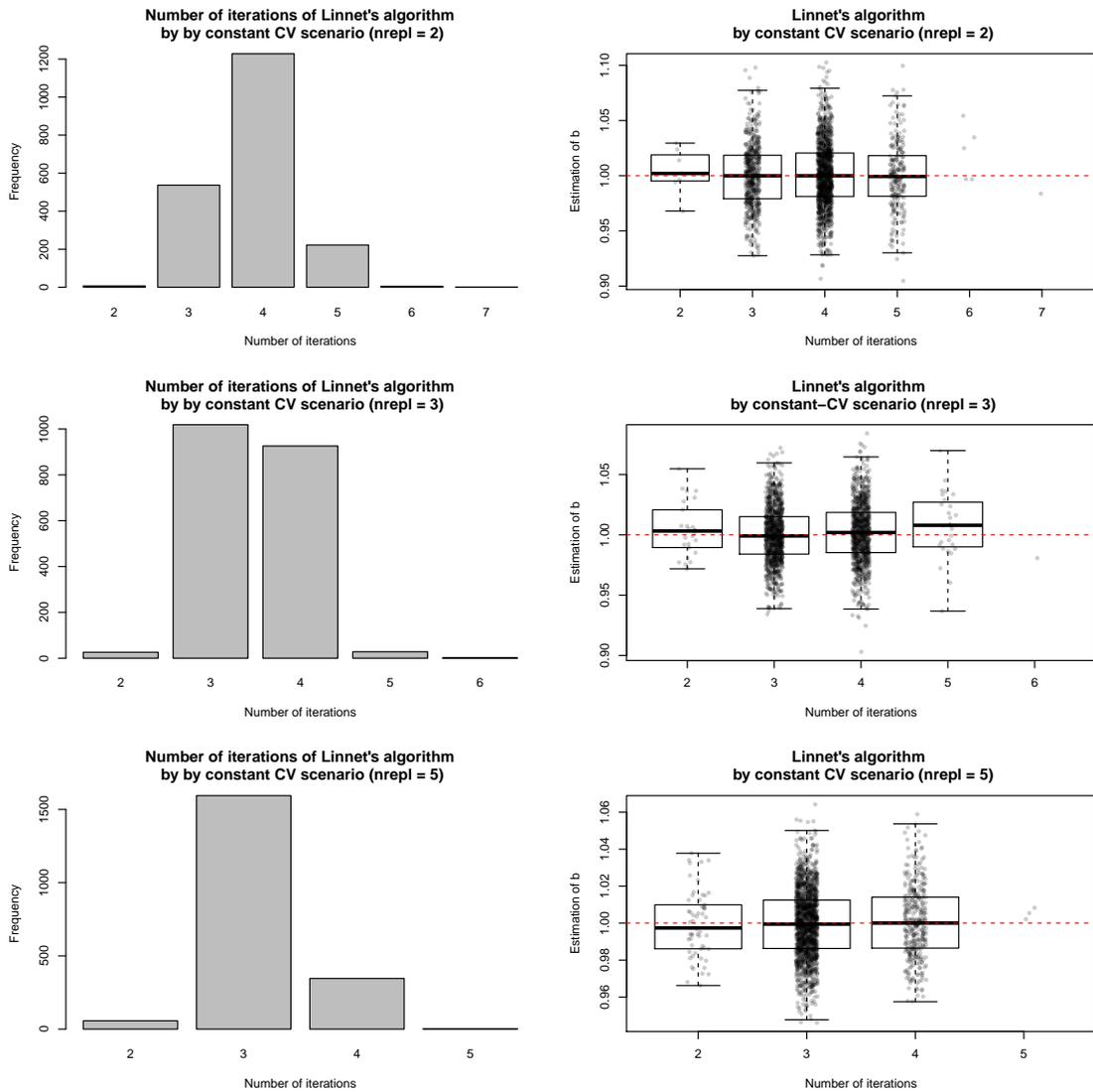
Figure 4.26: **Simulation 1.6.** Left: Number of iterations of Linnet's algorithm applied to heteroscedastic data (CV = 0.15). Right: Number of iterations vs. corresponding estimates of slope.

Simulation 1.7

Scope: Investigation of the dependence between the true slope and intercept and the performance of the estimates in case of heteroscedastic data with constant CV.

Model assumptions: Heteroscedastic data with constant CV without matrix effect.

Parameter	fixed	varying
number of replications (nrepl)	2	
slope a		0, 2.5, 5, 25
intercept b		0.5, 0.8, 1, 1.2, 1.2, 2
λ	1	
$k = \frac{sd(\,measurement\ error\ of\ X\,)}{sd(\,true\ measurements\ of\ X\,)}$	-	
$m.x = \frac{sd(\,Matrix\ effect\ of\ X\,)}{sd(\,Measurement\ error\ of\ X\,)}$	0	
$m.y = \frac{sd(\,Matrix\ effect\ of\ Y\,)}{sd(\,Measurement\ error\ of\ Y\,)}$	0	
CV.x	0.1	
CV.y	0.1	

Expectations

1. The bias of the estimates should not be affected by the true values of slope and intercept.

2. The variance of estimates may increase with increasing slope. The intercept should not affect the variation of the estimates.

3. The model assumption of Linnet's weighted Deming regression suits perfectly to this simulation scenario. That is why we expect estimates of this method to have the smallest variance.

Simulation results

The results agree with expectations (see fig. 4.27.)
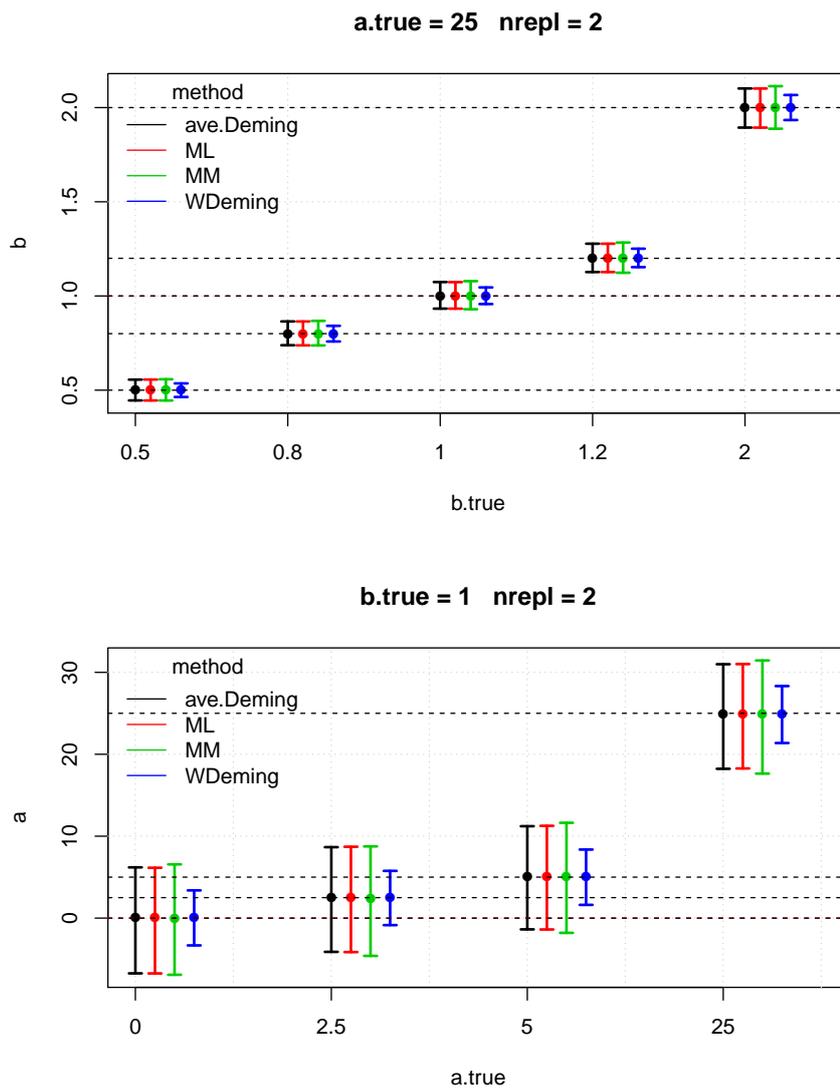
Figure 4.27: **Simulation 1.7.** Estimates of slope and intercept vs. the true values of these parameters.

Simulation 1.8

Scope: Investigation of the dependence between λ and the performance of the estimates in case of heteroscedastic data with constant CV.

Model assumptions: Heteroscedastic data with constant CV without matrix effect.

Parameter	fixed	varying
number of replications (nrepl)	2	
slope a	0	
intercept b	1	
λ		0.5, 1, 2
$k = \frac{sd(\,measurement\ error\ of\ X\,)}{sd(\,true\ measurements\ of\ X\,)}$	-	
$m.x = \frac{sd(\,Matrix\ effect\ of\ X\,)}{sd(\,Measurement\ error\ of\ X\,)}$	0	
$m.y = \frac{sd(\,Matrix\ effect\ of\ Y\,)}{sd(\,Measurement\ error\ of\ Y\,)}$	0	
CV.x	0.1	
CV.y	$\sqrt{\lambda}0.1$	

Note. We simulate heteroscedatic variances as

$\sigma_x = \sigma_{0x} + CV_x\,\xi_i$, where ξ_i is true measurement of X, $\sigma_0 = 1$.

$\sigma_y = \sqrt{\lambda}\,(\sigma_{0x} + CV_x\,\xi_i)$. Then

$\sigma_{0y} = \sqrt{\lambda}\,\sigma_{0x}$, $CV_y = \sqrt{\lambda}\,CV_x$

Thus

$\lambda = 0.5 \quad \Leftrightarrow \quad CV_y = \sqrt{0.5}\,CV_x = 0.7071\cdot, 0.1 = 0.07071,$

$\lambda = 1 \quad \Leftrightarrow \quad CV_y = \sqrt{1}\,CV_x = 0.1,$

$\lambda = 2 \quad \Leftrightarrow \quad CV_y = \sqrt{2}\,CV_x = 0.14142.$

Expectations

1. The true λ of the data should not affect the bias of the estimates of all methods.

2. The variance of estimates is expected to increase with increasing λ.

Simulation results

The results agree with our expectations:

1. The variances of estimates of all methods tend to increase with increasing λ (fig. 4.28-4.29).

2. The distributions of estimates of λ via. ML-method, averaged Deming regression and weighted Deming regression are slightly skewed with the median at the true value of λ (fig. 1.30). The estimates of λ calculated using Linnet's weighted Deming regression have the best accuracy, but the difference compared to the other methods is very small.
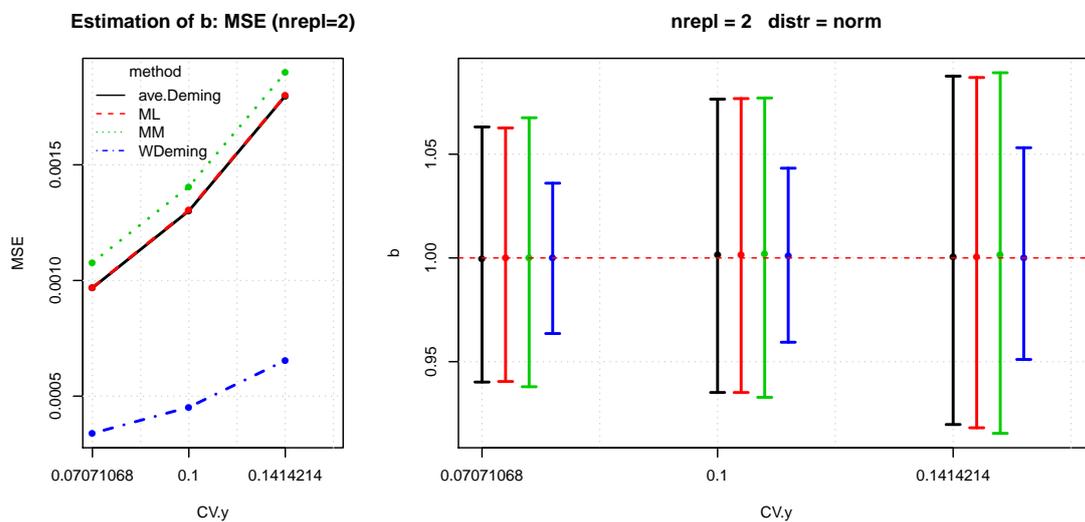
Figure 4.28: **Simulation 1.8.** Estimation of slope vs. CV of test method Y (CV of reference method x is fixed at 0.1. The CV.y values of 0.07071068, 0.1 and 0.1414214 correspond to the values of true$\lambda = 0.5, 1, 2$). MSE (left) and distribution of the estimates (median, 2.5%-quantile and 97.5%-quantile).
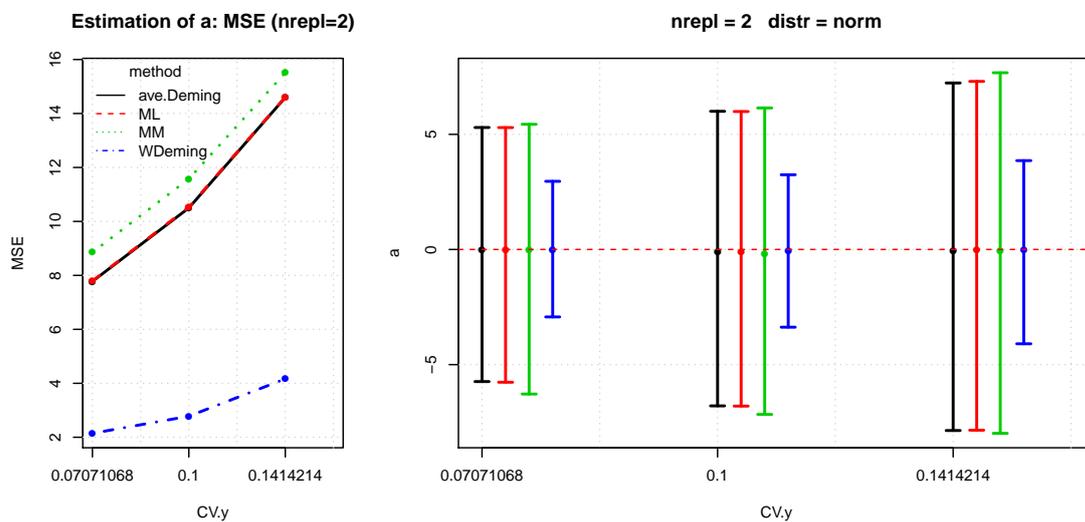
Figure 4.29: **Simulation 1.8.** Estimation of the intercept vs. the CV of the test method Y (CV of reference method x is fixed at 0.1. The CV.y values of 0.07071068, 0.1 and 0.1414214 correspond to the values of trueλ = 0.5, 1, 2). MSE (left) and distribution of the estimates (median, 2.5%-quantile and 97.5%-quantile).
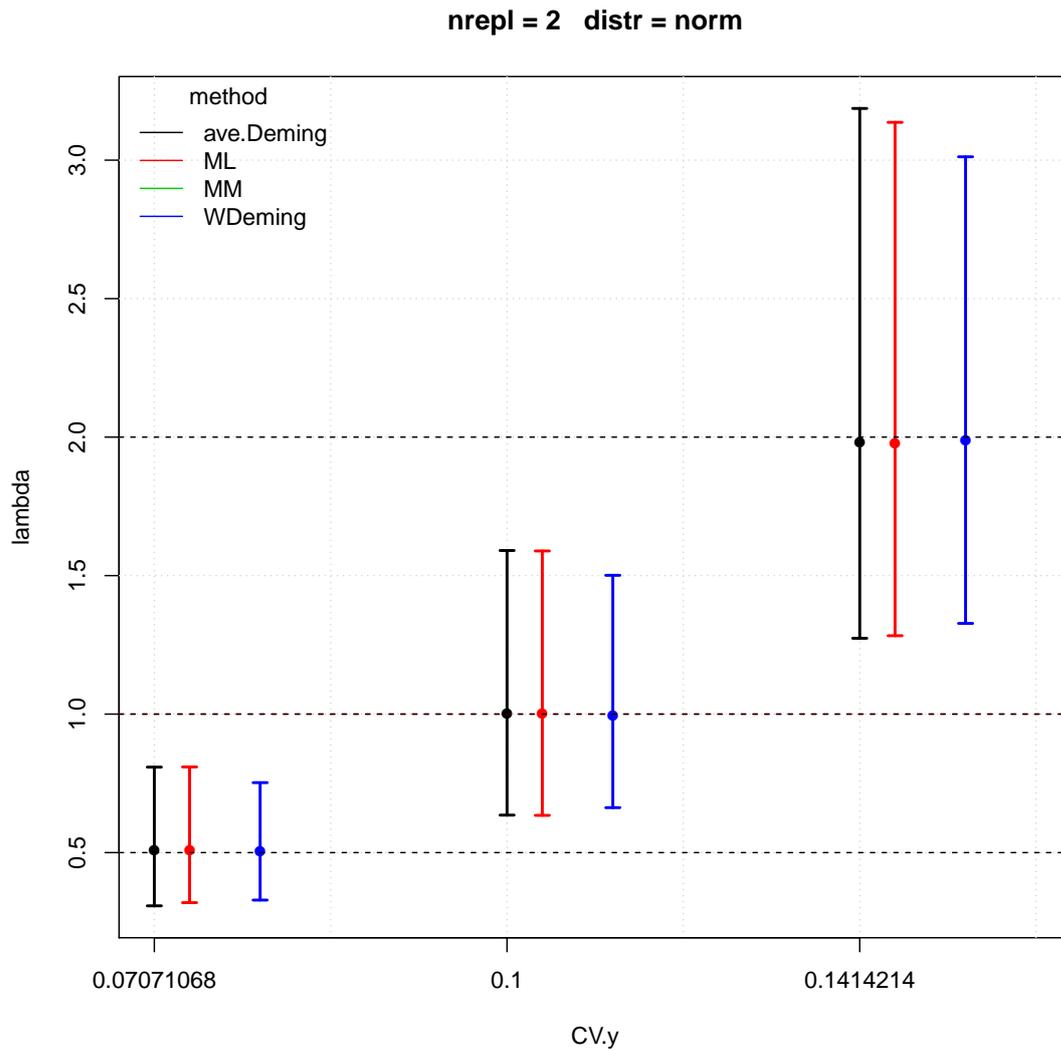
Figure 4.30: **Simulation 1.8.** Estimation of λ vs. true values of λ (CV of reference method x is fixed by 0.1. The CV.y values of 0.07071068, 0.1 and 0.1414214 correspond to the values of trueλ = 0.5, 1, 2). Distribution of the estimates (median, 2.5%-quantile and 97.5%-quantile).

4.2.1 Conclusions

ML-estimation

Our simulations show that

- The ML-method yields the best results for data with homoscedastic variances of errors of both methods X and Y. In this case the results are unbiased independent from the true slope, intercept and ratio of variances of measurement errors λ.

- in case of a weak matrix effect in the test method (the variance of the matrix effect is less than or equal to the variance of the measurement errors), this method yields negligible biased estimates of the slope and intercept.

- in case of a strong matrix effect in the test method, the ML estimates are considerably biased. The slope is overestimated and the intercept is underestimated. The stronger the matrix effect, the higher the bias and variance of estimates.

- in case of strong but equal matrix effect in both methods X and Y and $\lambda = 1$, the ML-estimates of regression coefficients are unbiased and have the smallest variance of all methods. The interaction between matrix effect and λ should be investigated in more detail.

- in case of heteroscedastic data, the ML-estimates of the regression coefficients are unbiased. The distribution of estimates of λ are slightly skewed with the median at the true value of λ. The variance of estimates of λ is higher than the variance of estimates of the Linnet's weighted Deming regression.

Averaged Deming regression

Our simulations suggest that

- the estimates of the averaged Deming regression are very close to the ML-estimates both in homoscedastic and heteroscedastic cases.

- in case of data with strong matrix effect, ML-estimation yields better results than the averaged Deming regression.

MM-method

For this method our simulations allow the following conclusions:

- This method yields reasonable estimates of regression parameters in case of homoscedastic data without matrix effect.

- In case of homoscedastic data with a matrix effect in the test measurement procedure the MM-method is the only approach that yields unbiased estimates.

- In case of data without a matrix effect the MM-esimations are rather similar to the ML estimates. The variance of estimates is slightly higher than the variance of the ML-estimates.

- The distribution of estimates of the variance of the matrix effect is slightly skewed, with median close to the true parameter. If there is no matrix effect in the test method the distribution of estimates is symmetric with median close to zero. In that situation the variance is estimated negative by the MM-method in 50% of the simulations. We can use this fact as an indicator of the absence of a matrix effect in the test method.

- In case of equal matrix effects in both methods the MM-estimates of regression parameters are strongly biased. The variance of the matrix effect is strongly overestimated. Please note: the model that we have used to derive the MM-estimates assumes a matrix effect only in the test method Y.

Linnet's weighted Deming regression

Our simulations show that

- the method yields unbiased estimates of regression coefficients in case of homoscedastic and heteroscedastic (with constant CV) data without matrix effect. For the convergence the algorithm needs 3-4 iterations in the most cases.

- in case of heteroscedastic data with constant CV, this method yields estimates with smallest variance.

- in case of homoscedasic data, the variance of estimates of Linnet's method is slightly higher than the variance of the ML-estimates and of the estimates of the averaged Deming regression.

- in case of a matrix effect, Linnet's weighted Deming regression shows almost the same behavior as the averaged Deming regression.

- in all simulation scenarios, the estimates of λ are slightly skewed with median close to the true parameter.

Barnett's Method

Barnett's algorithm converges very slowly in our simulations and yields unstable and biased estimates of regression coefficients, depending on the true slope of the data. Barnett's approach can be improved by centering of the data, however, the variance of the estimates stays high. The idea to have a proper solution for the model with such few assumptions is very attractive, but the current implementation of this method is not appropriate for use in method comparison studies.

4.3 Investigation of performance of confidence intervals

In order to investigate the performance of confidence intervals we conduct the following simulations:

- **Simulation 2.1.** Evaluation of performance of CIs in case of homoscedastic data.

- **Simulation 2.2.** Evaluation of performance of CIs in case of heteroscedastic data with constant CV.

- **Simulation 2.3.** Evaluation of performance of CIs in case of heteroscedastic data with Laplace distributed errors.

Table 4.2 shows the values of simulation parameters that are used in simulations 2.1 - 2.3. In all simulations we consider data with 2 and 3 measurement replications with slope b = 1, intercept a = 0 and $\lambda = 1$. The errors of both measurement methods X and Y have the same distribution.

Simulation	s_{0x}	s_{0y}	CV_x	CV_y	c_{0x}	c_{0x}	c_{1x}	c_{1y}
Simulation 2.1	5	5	0	0	0	0	0	0
Simulation 2.2	2	2	0.1	0.1	0	0	0	0
Simulation 2.3	2	2	0.05	0.05	10	10	0.01	0.01

Table 4.2: Overview of simulation parameters for the distribution of measurement errors (for the meaning of parameters see section "Simulation of data" 4.1).

To compare the performance of the confidence intervals, we estimate their coverage probability and the mean of their width. We estimate coverage parameters using the following algorithm:

1. Simulate 1000 data sets with the same combination of simulation parameters.

2. For each data set calculate confidence intervals for regression parameters of interest and proof whether the confidence intervals cover the true parameters.

Assume, C_k is the confidence interval for some parameter θ calculated using the i-th simulated data set. Then the coverage probability of C is

$$coverage(C) = \frac{\# \{\theta \in C_k \, k = 1, 2, ...1000\}}{1000}.$$

The desired coverage probability should be close to the confidence level, which was fixed at 95%. If the coverage probability of two CIs is the same, the CI with smaller average width is preferred for superior.

Note 1:

Simulation 1.2 has shown that the actual implementation of Barnett's approach yields biased

estimates of regression coefficients if the true intercept $a \neq 0$. That is why we do not consider this method in the following simulations.

Note 2:

We consider CIs for the estimates of

1. slope and intercept (for all methods),

2. λ (for all methods except MM-method),

3. the variance of the matrix effect (only for MM-method),

4. the bias between two measurement methods at different medical decision points. We consider 10%-, 20%-, ... 80%-, 90%-quantiles of the distribution of the simulated true concentrations ξ_i as medical decision points of interest.

General expectations of all simulations

Generally, we expect a good performance of jackknife CIs. The coverage of the quantile bootstrap CIs may be less than 95% (section 2.6.1). The improved bootstrap BC_a confidence interval should perform better than bootstrap quantile CIs (section 2.6.2).

Simulation 2.1

Scope: Investigation of the coverage of confidence intervals in case of homoscedastic data without matrix effect, with normal distributed errors.

Note. Figure 4.31 shows an example of simulated data, tables 4.32-4.33 contain results of the simulation.

Expectations

1. The data suit perfectly to the model assumptions of the ML- and averaged Deming regression approach. That is why we expect the best results of these methods.

2. We expect very similar performance of ML estimates and estimates of the averaged Deming regression.

3. The CIs of the MM-estimates may be slightly bigger.

4. The model assumptions of the Linnet's weighted Deming regression are violated. We expect poorer results of this method.

5. The confidence intervals in case of 3 replications should be narrower than the corresponding CIs for the data with 2 replications.

Simulation results for data with 2 replications:

1. The analytical CIs for the averaged Deming regression show a coverage of 0.94-0.96 for regression coefficients, as well as for bias at all medical decision points.

2. All CIs for regression coefficients have satisfying results.

3. Almost all CIs for the bias have a coverage of 0.94-0.96. The exception are the bootstrap CIs for the bias at medical decision points in the upper measurement range (80% and 90% quantiles). They undercover slightly (0.93).

4. The jackknife CIs for the ML- and averaged Deming regression estimates of λ have a coverage of 0.93, the bootstrap CIs for this parameter show a better performance.

5. The CIs of the ML- and the averaged Deming regression estimates have almost equal coverage and length. The CIs of MM-estimates are sometimes slightly broader.

6. The length of CIs of the estimates of Linnet's weighted Deming regression is equal or slightly broader than the length of CIs of the estimates of the averaged Deming regression.

Simulation results for data with 3 replications:

1. Coverage of all CIs, except of the bootstrap quantile CI for the variance of the matrix effect is improved.

2. The length of CIs gets smaller. For example, the length of the jackknife CI for the estimation of the slope via averaged Deming regression is 13 % less than the corresponding CI for the data with 2 replications.

Simulation 2.2

Scope: Investigation of performance of the confidence intervals in case of heteroscedastic errors with constant CV.

Note 1: Figure 4.34 shows an example of simulated data, tables 4.35-4.36 contain the results of the simulation.

Note 2: In this simulation the calculation of BC_a confidence intervals for the bias failed because of a program error.

Expectations:

1. The data suit perfectly to the model assumptions of Linnet's weighted Deming regression. That is why we expect good results for this method.

2. The ML- and MM-estimates and the estimates of the averaged Deming regression are unbiased in case of heteroscedastic data with constant CV. That is why we expect satisfactory results of the nonparametric CIs for their estimates. The length of CIs should be considerably higher than the length of the corresponding CIs of the Linnet's weighted Deming regression.
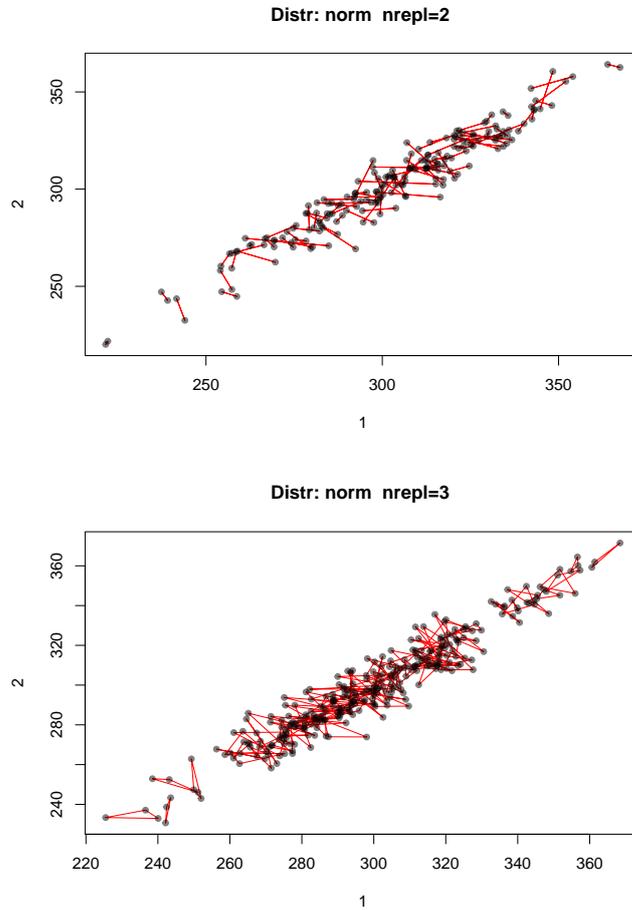
Figure 4.31: **Simulation 2.1.** Example of simulated data.

3. The analytical confidence intervals of the estimates of the averaged Deming regression should have poor coverage, since the model assumptions of this approach are violated.

Simulation results for data with 2 replications

1. As expected, the analytical CIs of the estimates of the averaged Deming regression fail.

2. All CIs of the estimates of the Linnet's weighted Deming regression have satisfactory coverage probability (0.94-0.97). An exception is the bootstrap BC_a CI for λ.

3. The coverage of jackknife and bootstrap quantile CIs for estimates of regression coefficients and bias via ML-, MM-method and averaged Deming regression are good (0.94-0.97). The CIs are considerably longer as the corresponding CIs of the estimates of Linnet's weighted Deming regression. For example, the jackknife CI for the estimates of slope using averaged Deming regression is 37% longer than the corresponding CI of the Linnet's weighted Deming regression.

4. All CIs for estimates of λ via ML-method and averaged Deming regression show a coverage of 0.93.

5. The bootstrap BC_a CI for estimation of λ of Linnet's weighted Deming regression has a coverage of 0.93. Against expectations, we do not observe better coverage of the bootstrap BC_a CIs compared to quantile bootstrap CIs.

Simulation results for data with 3 replications:

1. All confidence intervals get smaller. For example the length of the jackknife CI for the estimates of slope via Linnet's weighted Deming regression got 18 % smaller than the corresponding CI for the data with 2 measurement replications.

2. The coverage probability of bootstrap BC_a CI for the estimation of λ of the Linnet's weighted Deming regression is improved (0.94).

3. The coverage probability of jackknife CIs for estimation of λ via ML-estimation and averaged Deming regression was is improved (0.94).

4. The coverage probability of jackknife CIs for MM-estimation of the variance of matrix effect is also improved (0.94).

Simulation 2.3

Scope: Investigation of the performance of the CIs in case of slightly heteroscedastic Laplace distributed measurement errors.

Figure 4.37 shows example of simulated data. Figures 4.38-4.39 contain results of the simulation.

Expectations

These data do not suit to the assumptions of any model considered. Outliers may have a strong influence on estimates as well as on CIs. It is not clear which method and which type of CI will yield the best results.

Simulation results for data with 2 replications:

1. The coverage of the jackknife CIs of all methods achieves 0.94-0.95. The CIs of the Linnet's weighted Deming regression are approximately 25% narrower than the length of the CIs of the averaged Deming regression.

2. All bootstrap CIs do not achieve the desired coverage of 0.94-0.96.

3. The coverage of all CIs of all methods for λ are very poor.

4. The bootstrap quantile CI for the variance of matrix effect of MM-method have proper coverage of 0.95. The bootstrap BC_a CI do not achieve the confidence level.

5. The coverage of CIs of λ is very poor. Figure 4.40 shows an example of simulated data with distributions of the bootstrap samples of estimates of λ for the ML-method and the Deming regressions. We see that the estimates of λ are strongly biased.

Simulation results for data with 3 replications:

In this case the simulated data contain more outliers. The estimates of all methods are probably biased. That is why all results get worse. Now no method archives the coverage of 0.94-0.95.

4.3.1   Conclusions

The second series of simulations have shown the following:

1. Higher number of replications improves in most cases the performance of all CIs.

2. The results of the averaged Deming regression, ML- and MM- methods are very similar in all scenarios.

3. In case of homoscedastic data the CIs of the averaged Deming regression, and ML-method are slightly shorter than the CIs of MM-method and of Linnet's weighted regression.

4. In case of heteroscedastic data with constant CV and in case of data with Laplace distributed errors, Linnet's weighted Deming regression is considerably better than the other methods.

5. The best performance in all simulation scenarios achieve jackknife CIs.

6. The analytical CIs for the averaged Deming regression show proper performance in case of homoscedastic and normal distributed measurement errors. They fail in case of heteroscedastic or Laplace distributed errors.

7. Despite the expectations, the bootstrap BC_a CIs have the poorest performance of all non-parametric CIs.

8. Higher number of replications improves the estimation of CI for λ. In case of data with Laplace distributed measurement errors the estimation of λ of all methods is unacceptable.

Note. Despite of the good performance in our simulations of jackknife intervals for estimates of λ, it is reasonable to use bootstrap methods for calculation of CIs for this parameter. λ should be positive and jackknife CIs cannot guarantee it.

Coverage of CIs
homoscedatic errors, 2 replications

	a	b	lambda	Sc	Q10	Q20	Q30	Q40	Q50	Q60	Q70	Q80	Q90
ADEM:an.1	0.95	0.95			0.96	0.96	0.95	0.95	0.95	0.95	0.96	0.95	0.95
ADEM:an.2	0.95	0.95			0.96	0.96	0.95	0.95	0.96	0.96	0.96	0.96	0.95
ADEM:jack	0.96	0.96	0.93		0.95	0.95	0.95	0.95	0.95	0.96	0.96	0.95	0.95
ADEM:boot.q	0.95	0.95	0.94		0.95	0.95	0.95	0.94	0.95	0.95	0.94	0.93	0.93
ADEM:boot.bca	0.95	0.94	0.94		0.95	0.95	0.94	0.94	0.95	0.95	0.94	0.93	0.93
ML:jack	0.96	0.96	0.92		0.95	0.95	0.95	0.95	0.95	0.96	0.96	0.95	0.95
ML:boot.q	0.95	0.95	0.94		0.95	0.95	0.95	0.94	0.95	0.95	0.94	0.93	0.93
ML:boot.bca	0.95	0.94	0.94		0.95	0.95	0.94	0.94	0.95	0.94	0.94	0.93	0.93
MM:jack	0.95	0.96		0.94	0.95	0.96	0.95	0.95	0.96	0.96	0.96	0.95	0.95
MM:boot.q	0.95	0.95		0.93	0.95	0.95	0.95	0.94	0.95	0.95	0.94	0.93	0.93
MM:boot.bca	0.94	0.95		0.92	0.95	0.95	0.94	0.94	0.95	0.94	0.94	0.94	0.94
WDEM:jack	0.96	0.96	0.96		0.96	0.95	0.95	0.95	0.95	0.96	0.95	0.95	0.95
WDEM:boot.q	0.95	0.95	0.94		0.95	0.95	0.94	0.94	0.95	0.95	0.94	0.93	0.93
WDEM:boot.bca	0.94	0.94	0.94		0.95	0.94	0.94	0.94	0.95	0.94	0.94	0.93	0.93

Length of CIs
homoscedatic errors, 2 replications

	a	b	lambda	Sc	Q10	Q20	Q30	Q40	Q50	Q60	Q70	Q80	Q90
ADEM:an.1	24.28	0.08			3.20	2.58	2.23	2.04	1.98	2.04	2.23	2.58	3.21
ADEM:an.2	24.53	0.08			3.25	2.63	2.29	2.10	2.03	2.09	2.28	2.63	3.26
ADEM:jack	25.02	0.08	0.78		3.28	2.64	2.28	2.08	2.01	2.07	2.26	2.63	3.27
ADEM:boot.q	24.25	0.08	0.77		3.19	2.57	2.22	2.02	1.96	2.02	2.20	2.55	3.18
ADEM:boot.bca	24.27	0.08	0.77		3.19	2.57	2.22	2.02	1.96	2.02	2.20	2.55	3.18
ML:jack	25.03	0.08	0.78		3.28	2.64	2.28	2.08	2.01	2.07	2.26	2.63	3.27
ML:boot.q	24.26	0.08	0.77		3.19	2.57	2.22	2.02	1.96	2.02	2.20	2.55	3.18
ML:boot.bca	24.29	0.08	0.77		3.19	2.57	2.22	2.02	1.96	2.02	2.20	2.55	3.18
MM:jack	25.48	0.08		17.49	3.32	2.66	2.29	2.08	2.01	2.08	2.27	2.65	3.31
MM:boot.q	24.73	0.08		16.81	3.23	2.59	2.23	2.03	1.96	2.02	2.21	2.57	3.22
MM:boot.bca	24.74	0.08		17.06	3.23	2.59	2.23	2.03	1.96	2.02	2.21	2.57	3.22
WDEM:jack	25.70	0.09	0.86		3.34	2.67	2.29	2.08	2.01	2.08	2.28	2.66	3.34
WDEM:boot.q	24.68	0.08	0.85		3.21	2.57	2.22	2.02	1.96	2.02	2.22	2.59	3.24
WDEM:boot.bca	24.71	0.08	0.85		3.21	2.58	2.22	2.03	1.96	2.03	2.22	2.59	3.25

Figure 4.32: **Simulation 2.1.** Coverage (top) and corresponding length (bottom) of CIs. Please note: Sc denotes the variance of the matrix effect.
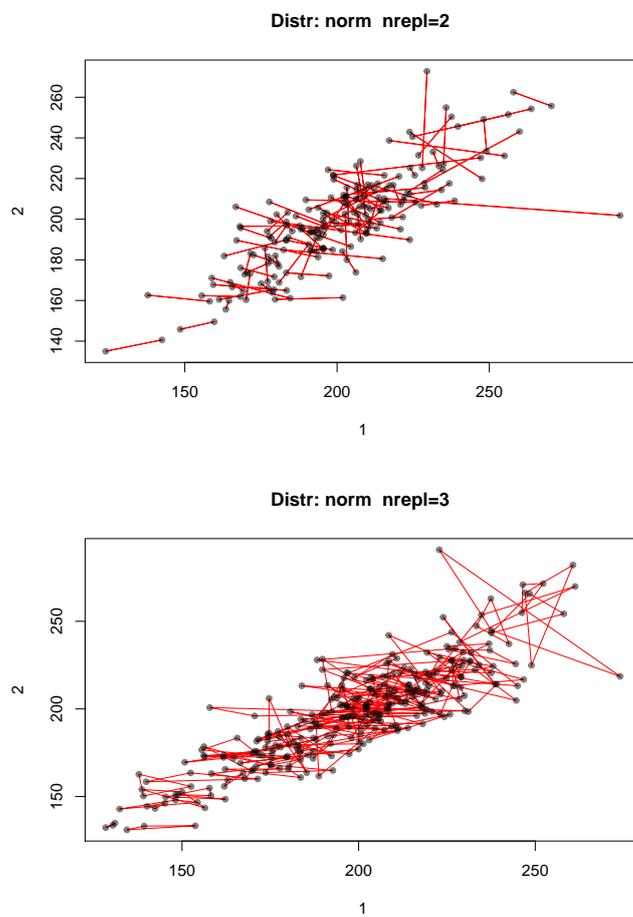
Coverage of CIs
homoscedatic errors, 3 replications

	a	b	lambda	Sc	Q10	Q20	Q30	Q40	Q50	Q60	Q70	Q80	Q90
ADEM:an.1	0.95	0.95			0.95	0.96	0.95	0.96	0.95	0.95	0.94	0.95	0.96
ADEM:an.2	0.95	0.95			0.95	0.96	0.96	0.96	0.95	0.95	0.94	0.95	0.96
ADEM:jack	0.95	0.95	0.95		0.95	0.96	0.95	0.96	0.95	0.95	0.95	0.95	0.96
ADEM:boot.q	0.95	0.95	0.96		0.94	0.95	0.94	0.95	0.94	0.95	0.94	0.95	0.95
ADEM:boot.bca	0.94	0.94	0.96		0.94	0.95	0.95	0.95	0.95	0.95	0.95	0.95	0.95
ML:jack	0.95	0.95	0.95		0.95	0.96	0.95	0.96	0.95	0.95	0.95	0.95	0.96
ML:boot.q	0.95	0.95	0.96		0.94	0.95	0.94	0.95	0.94	0.95	0.94	0.95	0.95
ML:boot.bca	0.94	0.94	0.96		0.94	0.95	0.95	0.95	0.95	0.95	0.95	0.95	0.95
MM:jack	0.95	0.95		0.95	0.96	0.96	0.96	0.96	0.95	0.95	0.95	0.96	0.96
MM:boot.q	0.95	0.94		0.93	0.94	0.95	0.94	0.95	0.94	0.95	0.95	0.95	0.96
MM:boot.bca	0.94	0.95		0.94	0.95	0.95	0.94	0.95	0.95	0.95	0.95	0.95	0.96
WDEM:jack	0.96	0.96	0.95		0.95	0.95	0.95	0.96	0.95	0.96	0.96	0.96	0.97
WDEM:boot.q	0.95	0.96	0.95		0.94	0.95	0.94	0.95	0.94	0.95	0.95	0.96	0.95
WDEM:boot.bca	0.94	0.95	0.96		0.95	0.95	0.95	0.95	0.95	0.95	0.95	0.96	0.96

Length of CIs
homoscedatic errors, 3 replications

	a	b	lambda	Sc	Q10	Q20	Q30	Q40	Q50	Q60	Q70	Q80	Q90
ADEM:an.1	19.46	0.06			2.58	2.08	1.80	1.65	1.60	1.65	1.80	2.09	2.58
ADEM:an.2	19.66	0.07			2.62	2.12	1.84	1.68	1.63	1.68	1.83	2.13	2.62
ADEM:jack	19.97	0.07	0.56		2.64	2.13	1.84	1.68	1.62	1.67	1.83	2.13	2.63
ADEM:boot.q	19.37	0.06	0.55		2.57	2.07	1.79	1.63	1.58	1.63	1.78	2.06	2.55
ADEM:boot.bca	19.37	0.06	0.55		2.58	2.07	1.79	1.63	1.58	1.63	1.78	2.07	2.55
ML:jack	19.97	0.07	0.56		2.64	2.13	1.84	1.68	1.62	1.67	1.83	2.13	2.63
ML:boot.q	19.37	0.06	0.55		2.57	2.07	1.79	1.63	1.58	1.63	1.78	2.06	2.55
ML:boot.bca	19.37	0.06	0.55		2.58	2.07	1.79	1.63	1.58	1.63	1.78	2.07	2.55
MM:jack	20.16	0.07		10.41	2.66	2.14	1.84	1.68	1.62	1.67	1.83	2.13	2.64
MM:boot.q	19.54	0.06		9.96	2.59	2.08	1.80	1.64	1.58	1.63	1.78	2.07	2.56
MM:boot.bca	19.56	0.06		10.27	2.59	2.08	1.80	1.64	1.59	1.63	1.78	2.07	2.56
WDEM:jack	20.45	0.07	0.57		2.68	2.15	1.85	1.68	1.62	1.68	1.84	2.16	2.68
WDEM:boot.q	19.65	0.07	0.56		2.58	2.07	1.79	1.63	1.58	1.64	1.79	2.10	2.60
WDEM:boot.bca	19.69	0.07	0.56		2.59	2.08	1.79	1.63	1.59	1.64	1.80	2.10	2.60

Figure 4.33: **Simulation 2.1.** Coverage (top) and corresponding length (bottom) of CIs. Please note: Sc denotes the variance of the matrix effect.
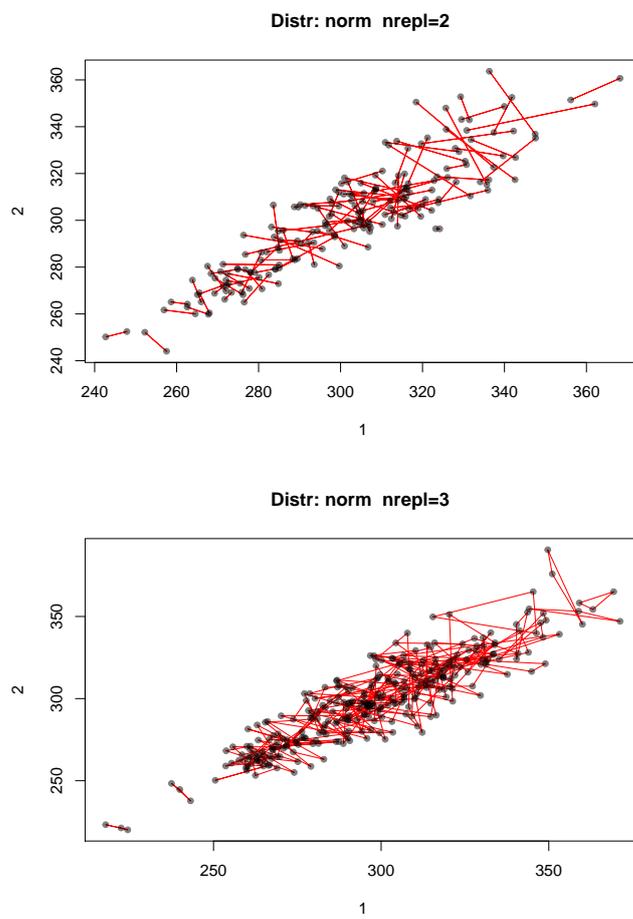
Figure 4.34: **Simulation 2.2.** Example of simulated data.

Coverage of CIs
heteroscedastic errors (const. CV), 2 replications

	a	b	lambda	Sc	Q10	Q20	Q30	Q40	Q50	Q60	Q70	Q80	Q90
ADEM:an.1	0.96	0.93			0.99	0.99	0.99	0.97	0.96	0.92	0.90	0.89	0.89
ADEM:an.2	0.96	0.94			0.99	0.99	0.99	0.98	0.97	0.94	0.91	0.90	0.89
ADEM:jack	0.95	0.96	0.93		0.96	0.96	0.96	0.97	0.96	0.95	0.95	0.95	0.95
ADEM:boot.q	0.94	0.95	0.93		0.94	0.96	0.96	0.96	0.96	0.95	0.94	0.94	0.94
ADEM:boot.bca	0.93	0.93	0.93										
ML:jack	0.95	0.95	0.93		0.96	0.96	0.96	0.97	0.96	0.95	0.95	0.95	0.95
ML:boot.q	0.94	0.95	0.93		0.94	0.96	0.96	0.96	0.96	0.95	0.94	0.94	0.94
ML:boot.bca	0.93	0.93	0.93										
MM:jack	0.96	0.95		0.94	0.96	0.96	0.96	0.96	0.95	0.95	0.95	0.95	0.95
MM:boot.q	0.94	0.93		0.92	0.94	0.95	0.96	0.96	0.96	0.95	0.94	0.94	0.94
MM:boot.bca	0.93	0.93		0.91									
WDEM:jack	0.96	0.96	0.95		0.95	0.96	0.97	0.96	0.96	0.96	0.95	0.95	0.95
WDEM:boot.q	0.94	0.95	0.94		0.95	0.96	0.96	0.96	0.96	0.95	0.95	0.95	0.94
WDEM:boot.bca	0.95	0.95	0.93										

Length of CIs
heteroscedastic errors (const. CV), 2 replications

	a	b	lambda	Sc	Q10	Q20	Q30	Q40	Q50	Q60	Q70	Q80	Q90
ADEM:an.1	14.27	0.14			5.51	4.42	3.82	3.49	3.38	3.49	3.83	4.43	5.49
ADEM:an.2	14.45	0.14			5.66	4.59	4.00	3.68	3.58	3.68	4.01	4.60	5.65
ADEM:jack	13.92	0.15	0.93		4.32	3.35	3.02	3.09	3.44	3.99	4.71	5.67	7.09
ADEM:boot.q	13.50	0.15	0.92		4.26	3.31	2.98	3.03	3.34	3.86	4.55	5.47	6.82
ADEM:boot.bca	13.60	0.15	0.92										
ML:jack	13.94	0.15	0.94		4.33	3.36	3.02	3.09	3.44	3.99	4.71	5.68	7.09
ML:boot.q	13.55	0.15	0.92		4.27	3.32	2.98	3.03	3.34	3.86	4.55	5.47	6.83
ML:boot.bca	13.64	0.15	0.92										
MM:jack	14.79	0.16		58.34	4.58	3.50	3.09	3.11	3.44	4.01	4.76	5.78	7.27
MM:boot.q	14.43	0.16		55.92	4.54	3.48	3.06	3.05	3.36	3.88	4.60	5.57	7.00
MM:boot.bca	14.43	0.16		56.93									
WDEM:jack	9.13	0.11	0.84		3.08	2.79	2.83	3.06	3.40	3.82	4.35	5.03	6.02
WDEM:boot.q	9.14	0.11	0.82		3.06	2.73	2.77	2.98	3.31	3.73	4.25	4.91	5.88
WDEM:boot.bca	9.11	0.11	0.81										

Figure 4.35: **Simulation 2.2.** Coverage (top) and corresponding length (bottom) of CIs. Please note: Sc denotes the variance of the matrix effect.

Coverage of CIs
heteroscedastic errors (const. CV), 3 replications

	a	b	lambda	Sc	Q10	Q20	Q30	Q40	Q50	Q60	Q70	Q80	Q90
ADEM:an.1	0.95	0.91			0.98	0.99	0.98	0.96	0.92	0.88	0.87	0.86	0.85
ADEM:an.2	0.95	0.91			0.99	0.99	0.99	0.97	0.93	0.89	0.87	0.86	0.86
ADEM:jack	0.96	0.96	0.94		0.95	0.96	0.96	0.96	0.96	0.97	0.97	0.96	0.96
ADEM:boot.q	0.95	0.95	0.92		0.94	0.95	0.96	0.95	0.96	0.96	0.95	0.95	0.95
ADEM:boot.bca	0.94	0.94	0.92										
ML:jack	0.96	0.96	0.94		0.95	0.96	0.96	0.96	0.96	0.97	0.97	0.96	0.96
ML:boot.q	0.95	0.95	0.92		0.94	0.95	0.96	0.95	0.96	0.96	0.95	0.95	0.95
ML:boot.bca	0.94	0.94	0.92										
MM:jack	0.96	0.96		0.94	0.95	0.96	0.96	0.96	0.96	0.97	0.96	0.96	0.96
MM:boot.q	0.94	0.95		0.92	0.94	0.95	0.95	0.96	0.96	0.96	0.95	0.95	0.94
MM:boot.bca	0.94	0.94		0.90									
WDEM:jack	0.96	0.96	0.95		0.96	0.96	0.96	0.95	0.96	0.96	0.96	0.96	0.96
WDEM:boot.q	0.96	0.96	0.94		0.95	0.96	0.95	0.95	0.95	0.96	0.95	0.96	0.96
WDEM:boot.bca	0.96	0.96	0.94										

Length of CIs
heteroscedastic errors (const. CV), 3 replications

	a	b	lambda	Sc	Q10	Q20	Q30	Q40	Q50	Q60	Q70	Q80	Q90
ADEM:an.1	11.64	0.11			2.30	1.80	1.60	1.40	1.40	1.40	1.60	1.80	2.30
ADEM:an.2	11.78	0.11			2.30	1.90	1.60	1.50	1.50	1.50	1.60	1.90	2.30
ADEM:jack	11.41	0.13	0.66		3.50	2.70	2.50	2.50	2.80	3.30	3.90	4.70	5.90
ADEM:boot.q	11.04	0.12	0.64		3.40	2.70	2.40	2.50	2.70	3.20	3.70	4.50	5.60
ADEM:boot.bca	11.14	0.12	0.64										
ML:jack	11.42	0.13	0.66		3.50	2.70	2.50	2.50	2.80	3.30	3.90	4.70	5.90
ML:boot.q	11.05	0.12	0.64		3.40	2.70	2.40	2.50	2.70	3.20	3.70	4.50	5.60
ML:boot.bca	11.15	0.12	0.64										
MM:jack	11.82	0.13		35.80	3.60	2.80	2.50	2.50	2.80	3.30	3.90	4.70	5.90
MM:boot.q	11.46	0.12		34.00	3.60	2.80	2.40	2.50	2.70	3.20	3.80	4.50	5.70
MM:boot.bca	11.50	0.12		35.30									
WDEM:jack	7.40	0.09	0.58		2.50	2.30	2.30	2.50	2.80	3.10	3.60	4.10	4.90
WDEM:boot.q	7.47	0.09	0.57		2.50	2.20	2.30	2.40	2.70	3.00	3.50	4.00	4.80
WDEM:boot.bca	7.47	0.09	0.57										

Figure 4.36: **Simulation 2.2.** Coverage (top) and corresponding length (bottom) of CIs. Please note: Sc denotes the variance of the matrix effect.

Distr: norm nrepl=2

Distr: norm nrepl=3

Figure 4.37: **Simulation 2.3.** Example of simulated data.

Coverage of CIs
Laplace distributed errors, 2 replications

	a	b	lambda	Sc	Q10	Q20	Q30	Q40	Q50	Q60	Q70	Q80	Q90
ADEM:an.1	0.89	0.85			0.99	1.00	1.00	0.99	0.94	0.88	0.83	0.81	0.79
ADEM:an.2	0.89	0.86			0.99	1.00	1.00	0.99	0.94	0.89	0.84	0.81	0.80
ADEM:jack	0.94	0.94	0.07		0.95	0.95	0.95	0.94	0.94	0.94	0.94	0.94	0.95
ADEM:boot.q	0.93	0.92	0.00		0.93	0.93	0.94	0.93	0.93	0.93	0.93	0.93	0.93
ADEM:boot.bca	0.90	0.90	0.00		0.89	0.91	0.93	0.92	0.92	0.92	0.91	0.91	0.90
ML:jack	0.94	0.94	0.05		0.95	0.95	0.95	0.94	0.94	0.94	0.94	0.94	0.95
ML:boot.q	0.93	0.92	0.00		0.93	0.93	0.94	0.93	0.93	0.93	0.93	0.93	0.93
ML:boot.bca	0.90	0.90	0.00		0.89	0.91	0.94	0.93	0.92	0.92	0.92	0.91	0.91
MM:jack	0.94	0.94		0.97	0.95	0.95	0.95	0.94	0.94	0.94	0.94	0.94	0.94
MM:boot.q	0.93	0.92		0.91	0.93	0.93	0.95	0.93	0.93	0.93	0.93	0.93	0.93
MM:boot.bca	0.90	0.90		0.83	0.89	0.91	0.93	0.92	0.92	0.92	0.92	0.91	0.91
WDEM:jack	0.95	0.95	0.03		0.96	0.96	0.95	0.93	0.94	0.94	0.94	0.94	0.94
WDEM:boot.q	0.93	0.93	0.00		0.93	0.95	0.94	0.93	0.93	0.93	0.93	0.93	0.93
WDEM:boot.bca	0.91	0.91	0.00		0.92	0.94	0.93	0.92	0.92	0.92	0.92	0.92	0.92

Length of CIs
Laplace distributed errors, 2 replications

	a	b	lambda	Sc	Q10	Q20	Q30	Q40	Q50	Q60	Q70	Q80	Q90
ADEM:an.1	36.34	0.18			7.18	5.81	5.03	4.60	4.46	4.60	5.04	5.82	7.19
ADEM:an.2	36.72	0.18			7.29	5.92	5.13	4.70	4.57	4.71	5.15	5.93	7.30
ADEM:jack	45.26	0.24	22.41		5.16	3.46	3.04	3.54	4.54	5.80	7.26	9.00	11.46
ADEM:boot.q	42.79	0.23	21.84		5.01	3.43	3.02	3.47	4.38	5.55	6.90	8.52	10.83
ADEM:boot.bca	44.08	0.24	23.98		5.20	3.50	3.04	3.48	4.41	5.60	6.99	8.66	11.04
ML:jack	45.19	0.24	20.77		5.14	3.45	3.04	3.54	4.54	5.80	7.26	9.00	11.45
ML:boot.q	42.72	0.23	19.69		5.00	3.42	3.02	3.47	4.38	5.54	6.90	8.52	10.82
ML:boot.bca	44.44	0.24	21.41		5.26	3.54	3.03	3.48	4.41	5.61	7.02	8.70	11.11
MM:jack	45.20	0.24		151.45	5.14	3.45	3.04	3.54	4.54	5.80	7.26	9.00	11.45
MM:boot.q	42.72	0.23		143.40	4.99	3.42	3.02	3.47	4.38	5.54	6.90	8.52	10.82
MM:boot.bca	43.98	0.24		156.68	5.18	3.49	3.03	3.48	4.40	5.60	6.99	8.66	11.02
WDEM:jack	32.44	0.18	17.99		3.23	2.57	2.88	3.59	4.46	5.45	6.55	7.85	9.68
WDEM:boot.q	32.01	0.18	17.62		3.32	2.60	2.83	3.49	4.31	5.27	6.33	7.59	9.36
WDEM:boot.bca	32.69	0.18	19.01		3.38	2.60	2.84	3.50	4.34	5.30	6.38	7.66	9.48

Figure 4.38: **Simulation 2.3.** Coverage (top) and corresponding length (bottom) of CIs. Please note: Sc denotes the variance of the matrix effect.

Coverage of CIs
Laplace distributed errors, 3 replications

	a	b	lambda	Sc	Q10	Q20	Q30	Q40	Q50	Q60	Q70	Q80	Q90
ADEM:an.1	0.86	0.84			0.99	1.00	1.00	0.98	0.93	0.86	0.81	0.78	0.77
ADEM:an.2	0.86	0.84			1.00	1.00	1.00	0.98	0.94	0.87	0.81	0.79	0.77
ADEM:jack	0.93	0.92	0.01		0.96	0.98	0.97	0.95	0.94	0.94	0.93	0.93	0.92
ADEM:boot.q	0.92	0.92	0.00		0.94	0.96	0.96	0.95	0.94	0.93	0.92	0.91	0.91
ADEM:boot.bca	0.88	0.88	0.00		0.90	0.95	0.96	0.95	0.93	0.91	0.90	0.89	0.89
ML:jack	0.93	0.92	0.01		0.96	0.98	0.97	0.95	0.94	0.94	0.93	0.93	0.92
ML:boot.q	0.92	0.91	0.00		0.93	0.96	0.96	0.95	0.94	0.93	0.91	0.91	0.91
ML:boot.bca	0.88	0.88	0.00		0.90	0.95	0.96	0.95	0.93	0.91	0.90	0.89	0.89
MM:jack	0.93	0.92		0.96	0.96	0.98	0.97	0.95	0.94	0.95	0.93	0.93	0.92
MM:boot.q	0.91	0.91		0.92	0.93	0.96	0.96	0.95	0.94	0.93	0.91	0.91	0.91
MM:boot.bca	0.88	0.88		0.87	0.90	0.95	0.96	0.95	0.93	0.91	0.90	0.89	0.89
WDEM:jack	0.94	0.93	0.00		0.97	0.97	0.96	0.95	0.95	0.94	0.94	0.93	0.93
WDEM:boot.q	0.92	0.91	0.00		0.96	0.97	0.95	0.95	0.94	0.93	0.93	0.92	0.91
WDEM:boot.bca	0.90	0.89	0.00		0.94	0.97	0.95	0.94	0.93	0.93	0.92	0.91	0.90

Length of CIs
Laplace distributed errorss, 3 replications

	a	b	lambda	Sc	Q10	Q20	Q30	Q40	Q50	Q60	Q70	Q80	Q90
ADEM:an.1	29.45	0.15			5.86	4.73	4.10	3.75	3.64	3.75	4.11	4.75	5.88
ADEM:an.2	29.75	0.15			5.94	4.80	4.17	3.82	3.71	3.82	4.18	4.82	5.96
ADEM:jack	36.41	0.20	16.98		4.17	2.77	2.46	2.89	3.69	4.69	5.90	7.33	9.34
ADEM:boot.q	34.46	0.18	16.43		4.06	2.75	2.45	2.82	3.56	4.48	5.61	6.95	8.85
ADEM:boot.bca	35.51	0.19	17.81		4.21	2.81	2.46	2.83	3.57	4.52	5.68	7.05	9.00
ML:jack	36.41	0.20	16.25		4.17	2.77	2.46	2.89	3.69	4.69	5.90	7.33	9.34
ML:boot.q	34.45	0.18	15.43		4.06	2.75	2.45	2.82	3.56	4.48	5.61	6.95	8.85
ML:boot.bca	35.49	0.19	16.74		4.22	2.81	2.46	2.83	3.57	4.52	5.68	7.05	9.00
MM:jack	36.45	0.20		90.97	4.18	2.78	2.46	2.89	3.69	4.69	5.90	7.33	9.35
MM:boot.q	34.49	0.19		85.31	4.06	2.75	2.45	2.82	3.56	4.48	5.61	6.95	8.85
MM:boot.bca	35.51	0.19		91.33	4.22	2.81	2.46	2.83	3.57	4.53	5.68	7.05	9.00
WDEM:jack	26.25	0.15	13.69		2.64	2.08	2.34	2.93	3.63	4.41	5.33	6.41	7.91
WDEM:boot.q	25.92	0.14	13.32		2.71	2.11	2.30	2.84	3.51	4.26	5.15	6.20	7.67
WDEM:boot.bca	26.45	0.15	14.18		2.76	2.11	2.31	2.85	3.52	4.29	5.19	6.26	7.75

Figure 4.39: **Simulation 2.3.** Coverage (top) and corresponding length (bottom) of CIs. Please note: Sc denotes the variance of the matrix effect.
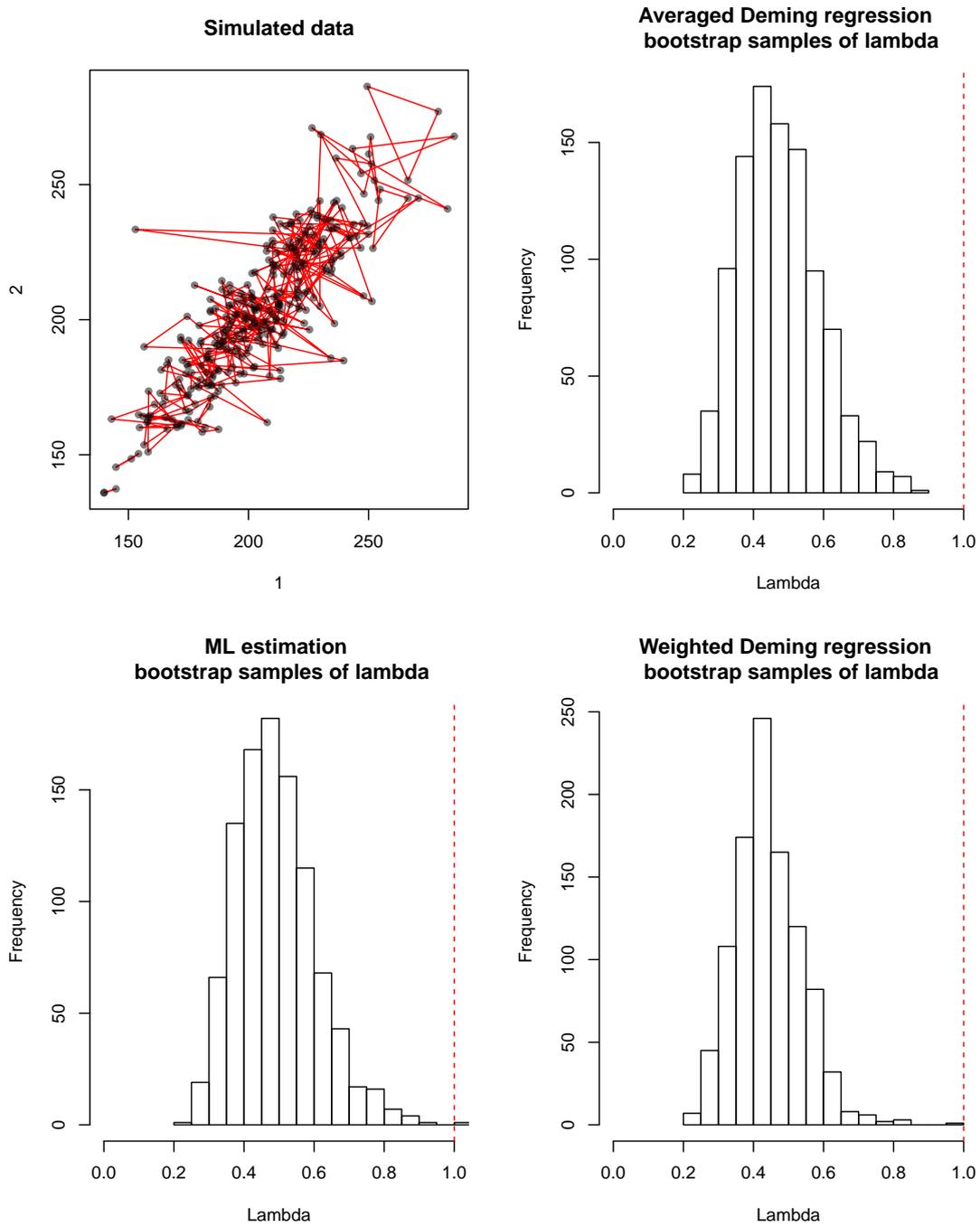
Figure 4.40: **Simulation 2.3.** Bottom left: example of simulated data. The rest of figures shows histigrams of the distribution of the bootstrap sample estimates of λ. The number of replications is fixed at 3. The true λ of the simulated data is 1.

Chapter 5

Examples

In the following we consider two examples of method comparison data with replicated measurements. These data enable us to apply the methods described in the theoretical part to real world problems and use the knowledge about them which we have obtained in simulation studies.

5.1   Example 1: Reagent lot comparison

A novel drug was developed for an inflammatory disease. A clinical study has shown that patients with high concentrations of biomarker A in blood are better responding to this new drug while the other patients (the subgroup with low concentrations) do not respond at all. These groups were separated by a cutoff of 50 ng/ml, which is approximately the median of the distribution of the concentrations of marker A in blood in the whole patient collective. During the initial study the blood concentrations were measured with the reagent lot "Design Reference (DR)" and the cutoff was established. The reagent lot "Master Pilot (MP)" is the first commercial lot that is used for actual patient treatment decisions. The patients with concentrations above 50 ng/ml should receive the new drug. The method comparison study on this data set should show that bias between the measurements of these two lots at the cutoff is within acceptable range and thus does not lead to an incorrect treatment decision.

5.1.1   Data

The data contains measurements of 220 patients. The blood samples were collected during a screening visit of the patient (screening visit) and on the first day of the study (baseline visit). Thus both samples of each patient were collected before the treatment with the new drug was started. The biomarker A was measured twice with the reagent lots Design Reference (DR) and Master Pilot (MP) in almost all samples. All considered models for method comparison analysis assume, that the specimen stays unchanged by measurement replications. That is why we cannot pool

the first two measurement replications (baseline visit) with the next two measurement replications (screening visit). We consider the measurements for the first and second specimens as different data sets. The first data set (baseline visit) contains 211 complete observations. The second data set (screening visit) contains 203 complete observations.

5.1.2   Diagnostic plots

- Figure 5.1 shows scatter plots for measurements of DR lot versus MP lot for the first (left panel) and second (right panel) sample of patients. The scatter plots are very similar. The points below 80 mg/ml are close to the bisecting line with slope 1 and intercept 0. The measurements of the MP lot seem to be slightly higher than the measurements of the DR lot. In the upper area of the measurement range (> 80 mg/dl) the measurements of the MP lot are considerably higher than the measurements of the DR lot.

- We do not see any clusters in the scatter plot, therefore we can assume that there is no strong random matrix effect in the data. Later we check this assumption by estimating of the matrix effect via the MM method.

- In order to select between the averaged or Linnet's weighted Deming regression we first use Bland-Altman plot for the averaged replications (fig. 5.2), which show absolute and relative differences between the measurements of the two lots against the average levels. We can see that the absolute differences increase with increasing averaged levels of measurements. The proportional difference are slightly shifted in the upper direction. Thus we observe proportional differences between the measurements of the DR and MP reagent lots. The slope of the data is expected to be larger than 1.0. The variance of the absolute differences increase with increasing averaged levels of measurements. The Bland-Altman plots suggest slightly heteroscedastic data.

- Scatter plots of differences between duplicate measurements versus their means confirm this result. The variance of differences within methods increase with increasing measurement values (fig. 5.3). In all plots we can observe that the differences between the first replication and the second replication are shifted. Thus, the assumption about the exchangeability of the replications is slightly violated. In all cases, the mean difference is appr. -0.5 ng/ml. This corresponds to appr. 1.0% of the median level. We do not expect that such low bias will not affect the results of the analysis considerably. The plot with duplicates of the measurement of MP lot for baseline visit data contains one clear outlier at the averaged measurement level of 60 ng/ml. The analysis of the data does not shown any reason for such a high difference between the two replications. That is why we do not delete the corresponding data point

from the data set. The averaging of replications will reduce the influence of this observation. The estimate of λ can be more affected by this outlier.

Thus we have heteroscedastic data with some moderate outliers. Linnet's weighted Deming regression showed to be the preferred in this case. We can expect proper estimates of the regression coefficients. The estimates of λ may be considerably affected by outliers.
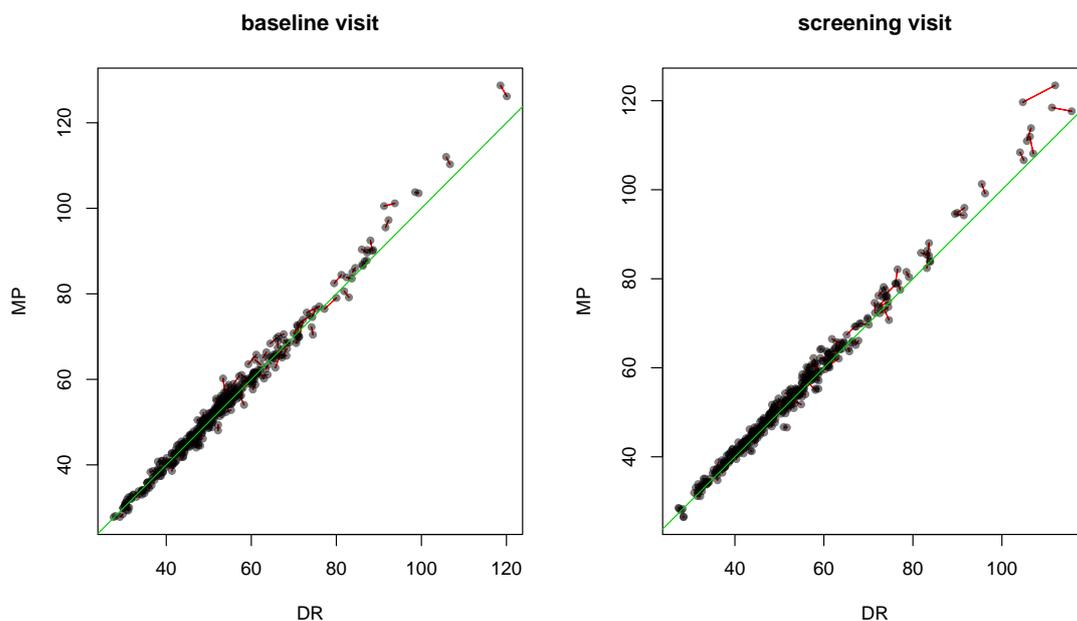
Figure 5.1: Example 1. Left: Scatter plot for measurements with lot DR and MP for the first day of study (baseline visit). Right: Scatter plot for measurements with lot DR and MP for the second screening visit of the study (screening visit). The replications are connected with red lines.
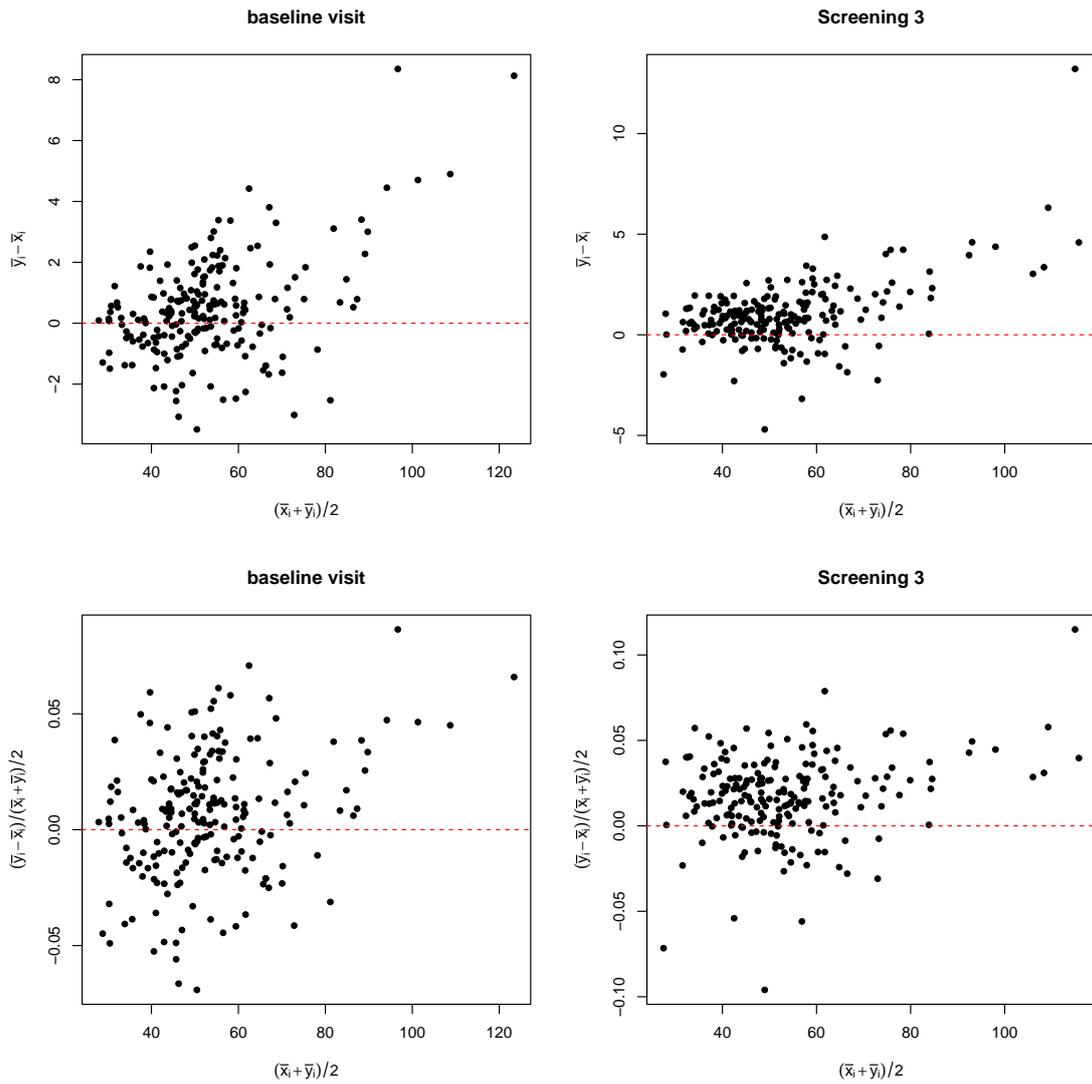
Figure 5.2: Example 1. Bland-Altman plots for the averaged replications of the first (left panel) and second (right) samples of the patients. Top scatter plots show absolute differences between DR and MP, the bottom graphics show the proportional differences.
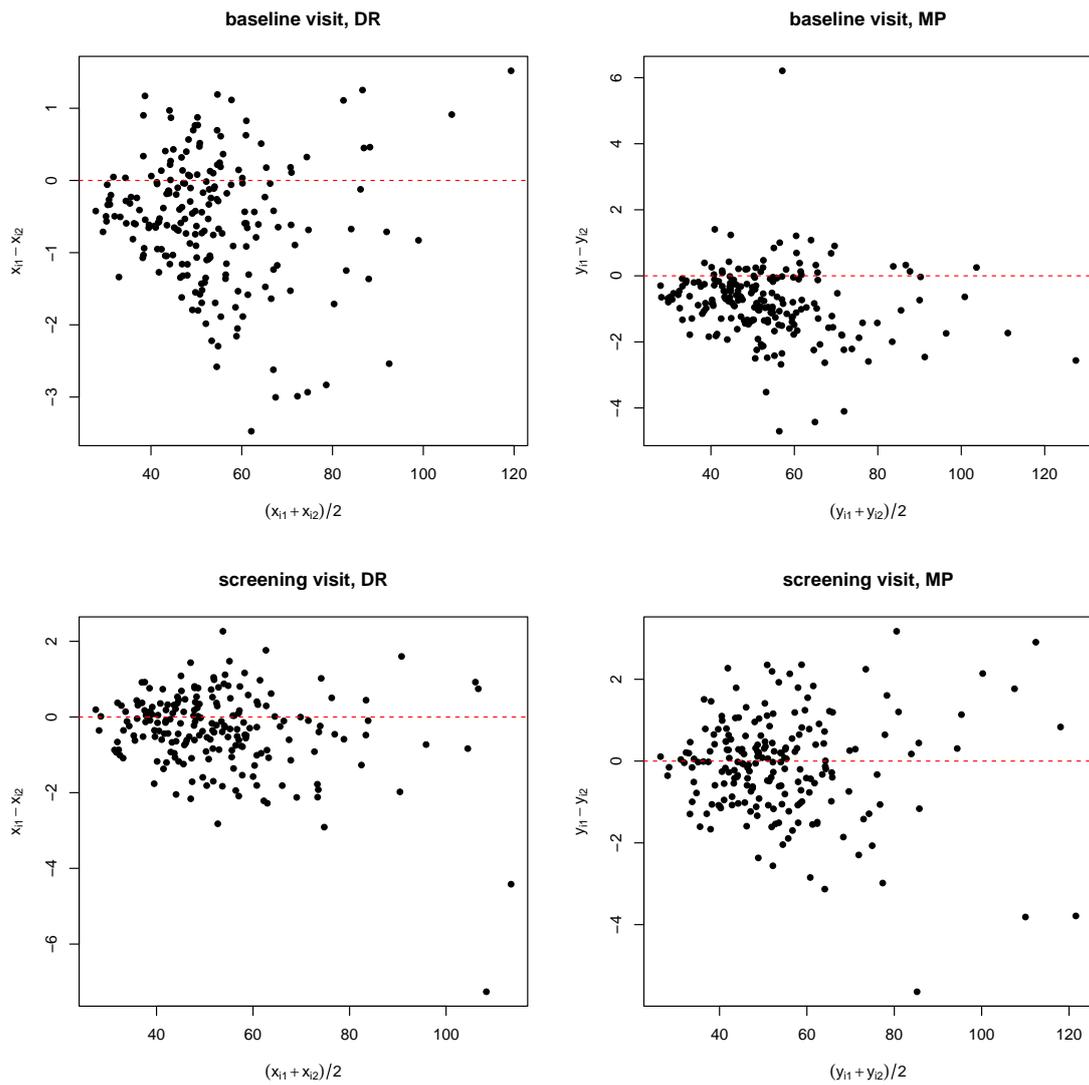
Figure 5.3: Example 1. Absolute differences between duplicate measurements by the methods plotted against the means of the measurements.

102

5.1.3 Regression analysis

Since we have asserted heteroscedastic type of data, Linnet's weighted Deming regression is suggested to be preferred approach for this data. However, we calculate estimates of all methods, which were considered in this master thesis, except for Barnett's approach. So we can compare the methods using real data. For the calculation of CIs for regression coefficients and matrix effect we use the jackknife method, as recommended by CLSI EP9 and Linnet [7]. Confidence intervals for λ are calculated using the bootstrap quantile method, because jackknife CIs may include negative values.

Table 5.1 contains the results of calculations. The results of comparison of the estimates of different methods are consistent with those results of the simulation studies. The analysis shows that:

- The estimates of regression coefficients of the averaged Deming regression are rather close to the ML-and MM-estimates.

- The estimates of slope via Linnet's Deming regression are closer to 1 than the estimates of the averaged Deming regression.

- The estimates of the intercept via Linnet's weighted Deming regression are closer to 0 than the estimates of the averaged Deming regression.

- The CI of the estimation of matrix effect does not contain zero. Thus there is a slight matrix effect in the data. The ratio between estimates of the standard deviations of the matrix effect and of the measurement errors is not larger than 2. That is why we can assume that matrix effect does not affect the estimates of the regression coefficients considerably.

- The estimates of λ via averaged and weighted Deming regression are similar. The ML method has much higher variance. This is consistent with results of the simulation 1.5, which has shown that the ML estimates yield unstable estimates of λ if there is matrix effect in the data.

	n	Intercept	Slope	Lambda	Matrix effect
baseline visit: ave.Deming	210	-2.167 (-3.319; -1.014)	1.049 (1.026; 1.073)	1.644 (1.157; 2.318)	
baseline visit: WDeming	210	-1.339 (-2.082; -0.596)	1.033 (1.018; 1.049)	1.703 (1.242; 2.290)	
baseline visit: ML	210	-2.108 (-3.259; -0.956)	1.048 (1.025; 1.072)	2.681 (1.218; 4.394)	
baseline visit: MM	210	-2.023 (-3.176; -0.870)	1.047 (1.023; 1.070)		1.511 (0.936; 2.087)
screening visit: ave.Deming	203	-1.912 (-3.255; -0.568)	1.055 (1.028; 1.082)	1.225 (0.809; 1.874)	
screening visit: WDeming	203	-0.769 (-1.600; 0.062)	1.033 (1.015; 1.050)	1.345 (1.014; 1.809)	
screening visit: ML	203	-1.895 (-3.259; -0.532)	1.055 (1.027; 1.082)	1.427 (0.454; 3.133)	
screening visit: MM	203	-1.776 (-3.048; -0.504)	1.052 (1.026; 1.078)		1.355 (0.473; 2.237)

Table 5.1: Example 1. Estimated regression parameters.

Figure 5.4 compares different kinds of CIs for the estimates of slope and intercept. Looking at figure 5.4 we can conclude that for this data

- All kinds of CIs for the estimates of Linnet's weighted Deming regression are very similar.

- CIs for the estimates of the other regression methods are less similar. If we consider for example the CIs for slope which was estimated via ML method, we can see that bootstrap BC_a intervals are more skewed compared with quantile bootstrap confidence intervals.

- The analytical CIs of the estimates of the averaged Deming regression are considerably narrower than the nonparametric CIs. We cannot use them for this data because of heteroscedasticity of measurement errors.

To check the model assumptions of averaged and weighted Deming regression, and MM-method, we use residual plots (fig. 5.5). We can see that Linnet's weighted Deming regression is the most appropriate for this data, although the difference to the residual plot of the averaged Deming regression is rather small. In the "screening visit" data set there are 2 observations which are relatively far from the 3 standard deviation limits. We consider them to be outliers.
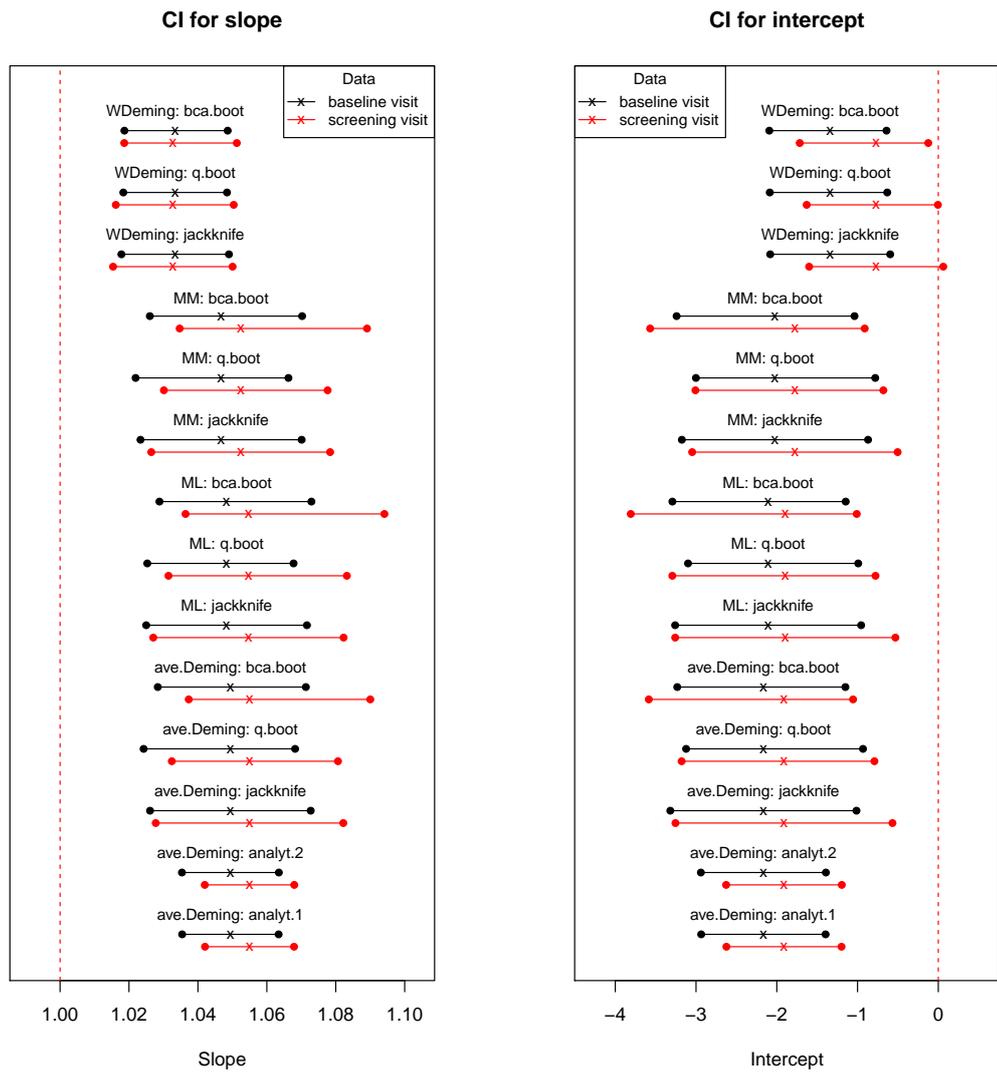
Figure 5.4: Example 1. Comparison of confidence intervals for regression coefficients.
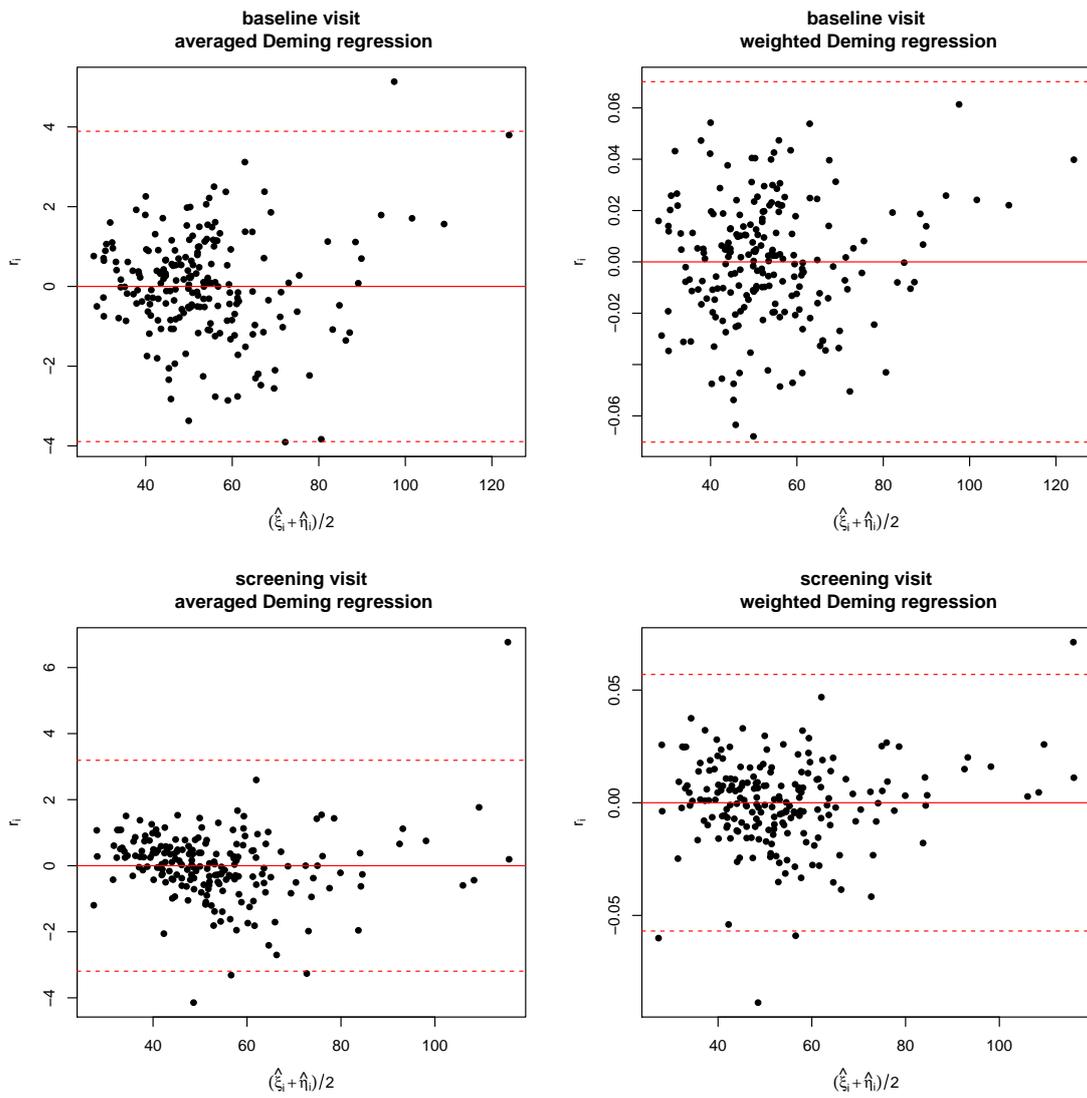
Figure 5.5: Example 1. Residual plots for the averaged Deming regression fit and Linnet's weighted Deming regression fit. The red dotted line shows the bounds of 3 standard deviations of the distribution of residuals. The data points above (under) this bounds can be considered to be outliers.
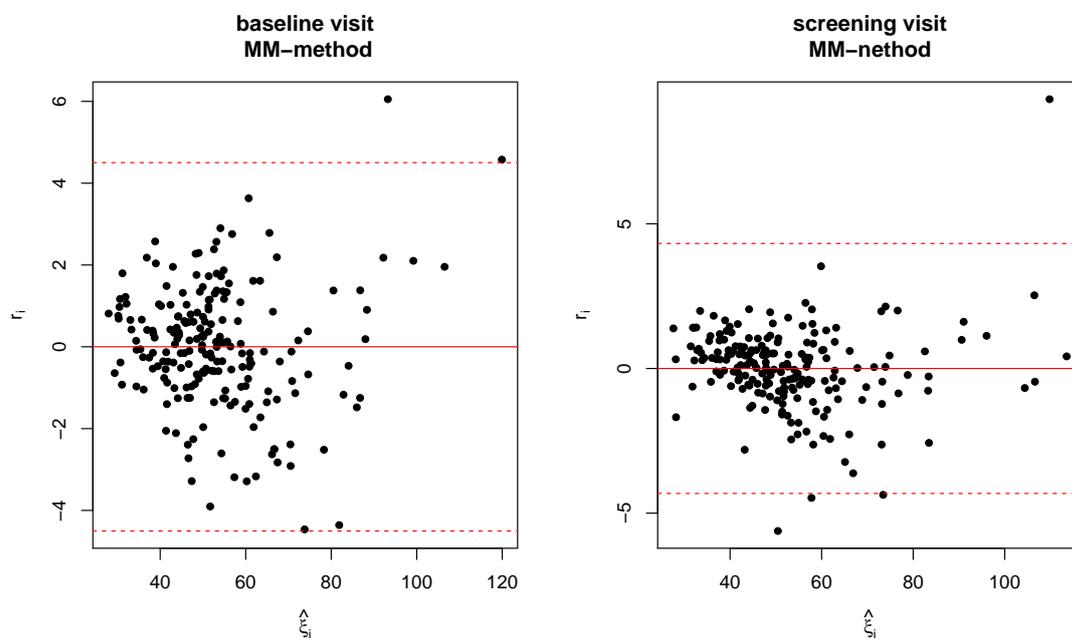
Figure 5.6: Example 1. Residual plot for MM method. The red dotted line shows the bounds of 3 standard deviations of the distribution of residuals. The data points above (under) this bounds can be considered to be outliers.

Now we consider only the estimates of the averaged- and weighted Deming regressions. Figure 5.7 shows data with the corresponding regression fits. Both lines lie above the identity line, very close to each other. The weighted Deming regression fit is slightly closer to the identity line than the averaged Deming regression fit.
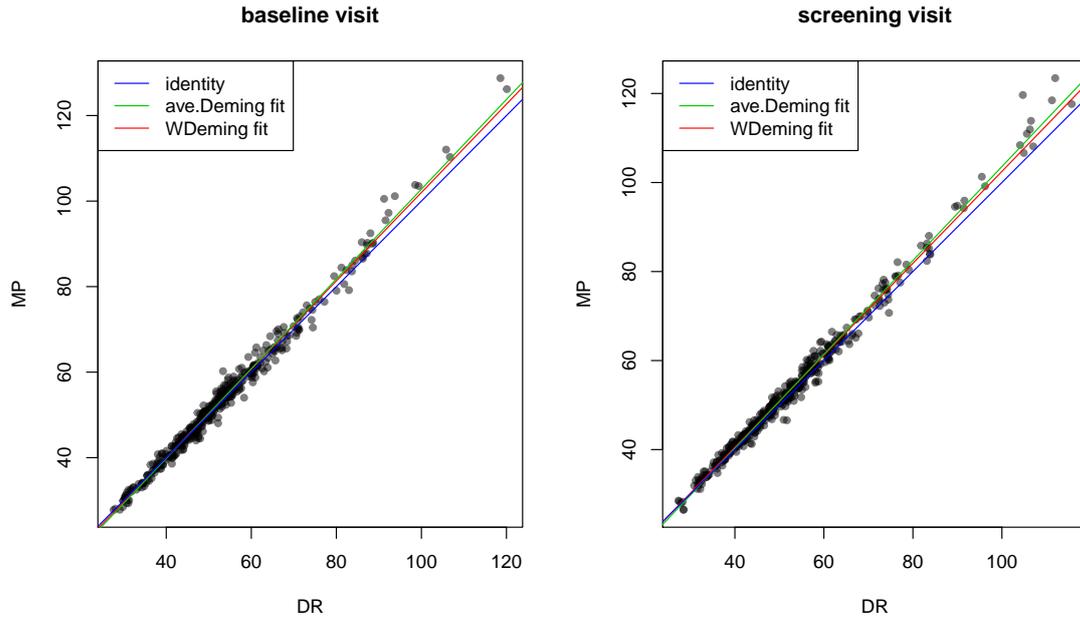
Figure 5.7: Example 1. Scatter plots for measurements with lot DR and MP for the first day of the study (baseline visit) and the second (screening visit) of the study with averaged Deming and Linnet's weighted Deming regression fit.

The absolute bias between the two measurements at some medical decision point X_c can be estimated by $\hat{a} + (\hat{b} - 1)Xc$. An estimate of the proportional bias is $\hat{a}/X_c - (\hat{b} - 1)$. For this data set the medical decision point of interest (MDP) is at 50 ng/ml. A proportional bias of 3% is desired. The proportional bias of 5 % is considered to be acceptable. We compare estimates of the proportional bias calculated with averaged Deming regression and Linnet's weighted Deming regression. For calculation of confidence intervals we use the jackknife method.

Table 5.2 shows the estimation of bias of averaged Deming regression and Linnet's weighted Deming regression. The upper limits of proportional bias at the MDP 50 ng/ml lie below 0.03 in both data sets. The estimates of bias of the both methods are very similar. The reason is that the value Xc = 50 ng/ml is very close to the center of data.

Figures 5.8-5.9 show the estimation of the absolute (upper) and proportional (lower) bias using averaged and Linnet's weighted Deming regression. The 95% confidence limits (dotted line) were calculated using the jackknife method. The CIs of proportional bias of both data sets lie within

the limits $\pm$ 3% or contain either +3% or -3% for almost all possible MDPs.

	baseline visit	screening visit
abs.bias: ave.Deming	0.305 (0.121; 0.489)	0.838 (0.663; 1.012)
abs.bias: WDeming	0.332 (0.150; 0.513)	0.867 (0.695; 1.039)
prop.bias: ave.Deming	0.006 (0.002; 0.010)	0.017 (0.013; 0.020)
prop.bias: WDeming	0.007 (0.003; 0.010)	0.017 (0.014; 0.021)

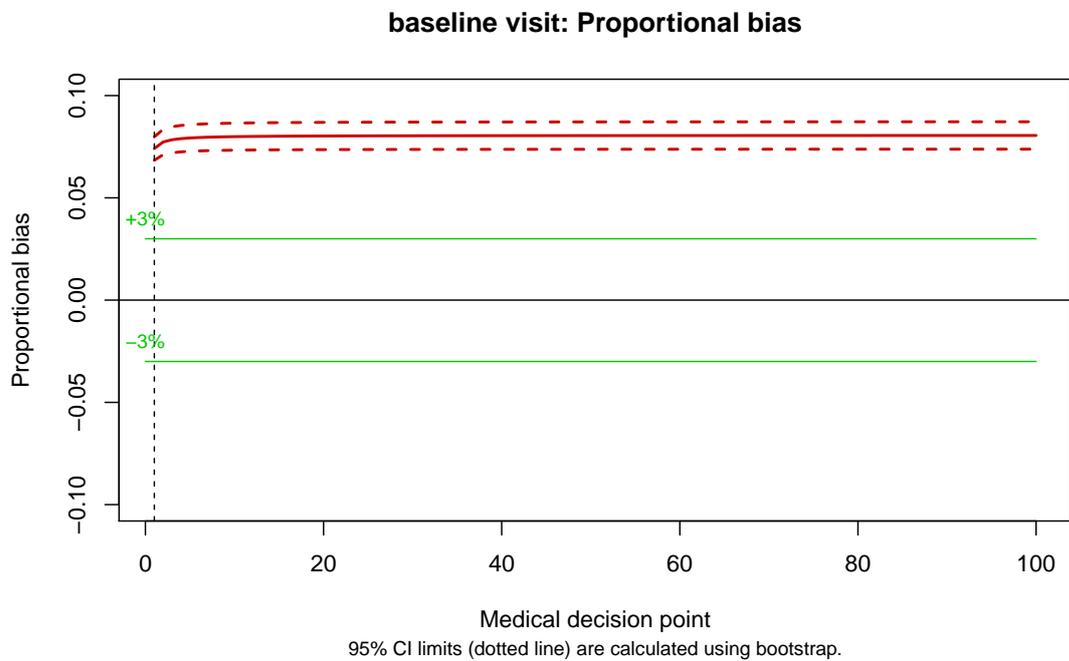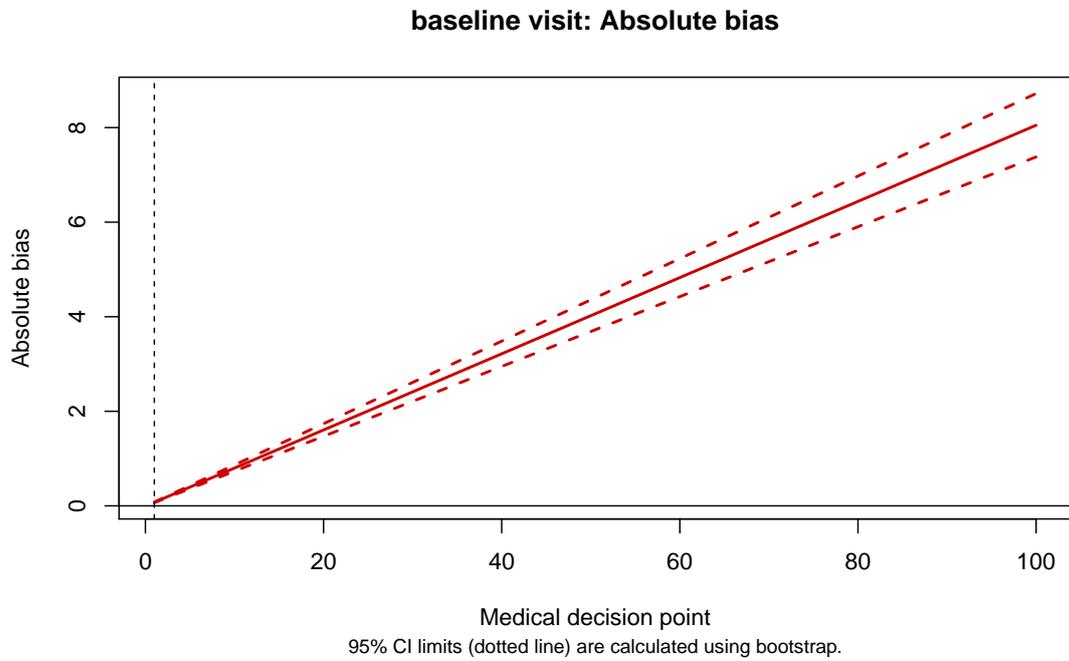Table 5.2: Estimation of bias and proportional bias at the MDP of 50 ng/ml with jackknife confidence intervals.

Conclusions

- The measurements of both instruments: the DR lot and MP lot have acceptable proportional bias at the medical decision point of 50 ng/ml 0.007 (0.003; 0.010) for baseline visit and 0.017 (0.014; 0.021) for screening visit.

- The method comparison analysis using data set "baseline visit" yields a slope of 1.033 (1.018; 1.049) and an intercept of -1.339 (-2.082; -0.596).

- The method comparison analysis using the data set "screening visit" yields a slope of 1.033 (1.015; 1.050) and an intercept of -0.769 (-1.600; 0.062).

- Estimates of λ are larger than 1.0 in both data sets. We can assume that the measurement error of the MP lot is larger than the measurement error of the DR lot. Simulation 2.3 suggests that estimation of λ can be considerably biased by outliers.

baseline visit: Absolute bias

baseline visit: Proportional bias
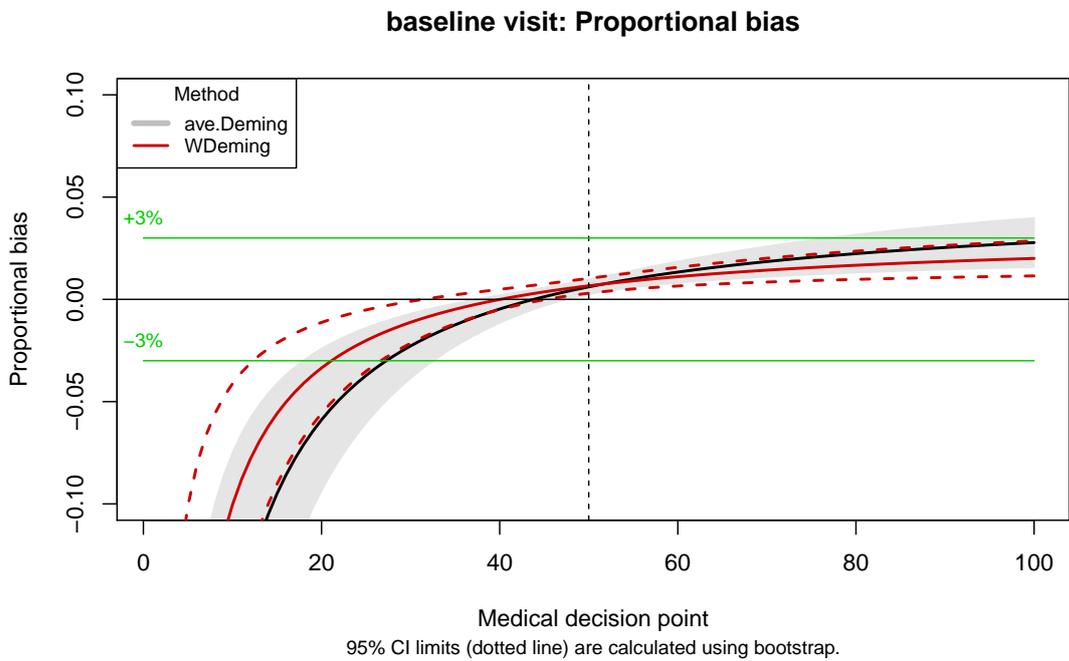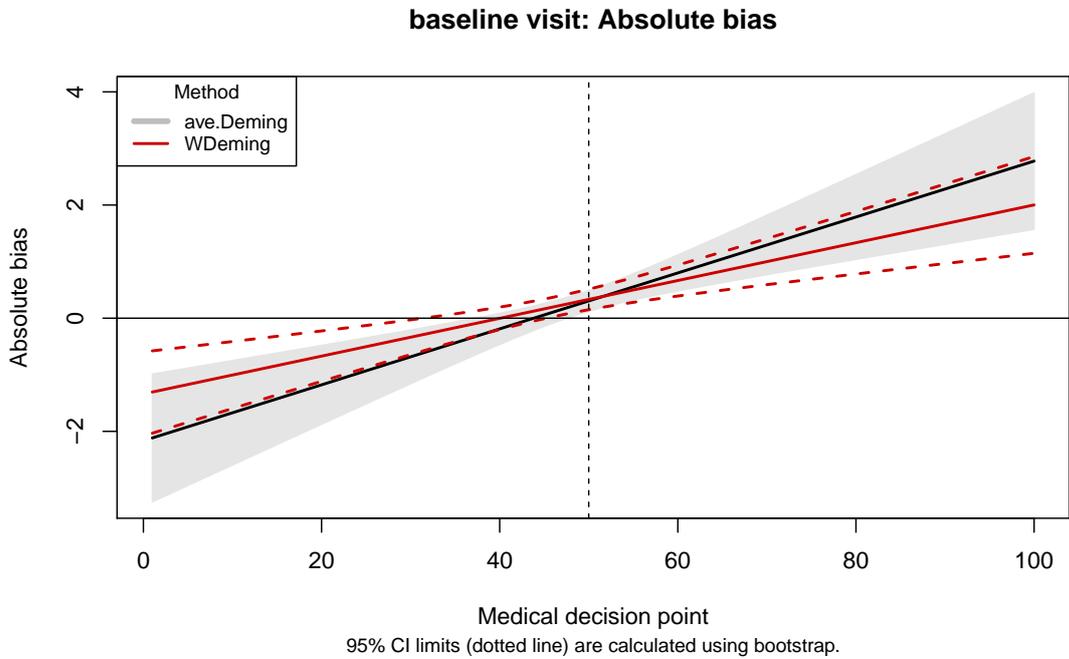
Figure 5.8: Example 1. baseline visit: Estimation of the absolute (uper) and proportional (lower) bias using averaged and Linnet's weighted Deming regression The 95% confidence limits (dotted lines) were calculated using the jackknife method.

screening visit: Absolute bias
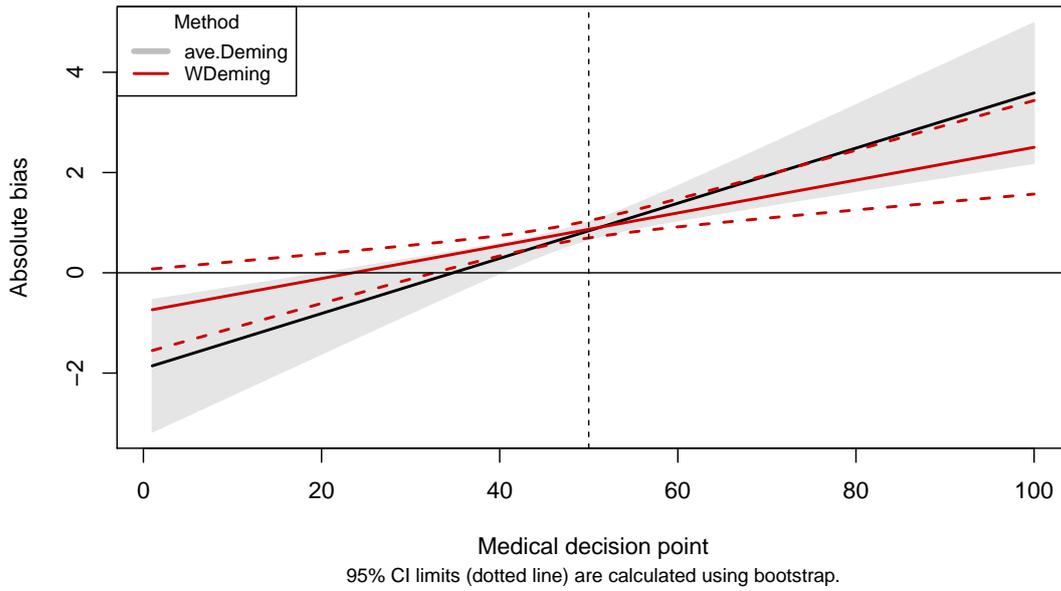
screening visit: Proportional bias
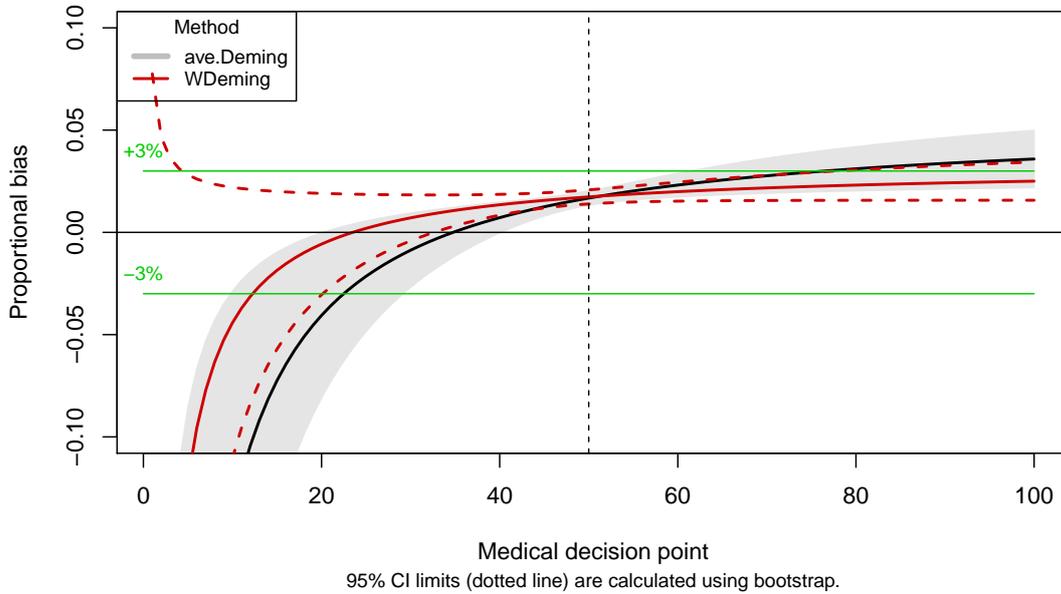
Figure 5.9: Example 1. Screening visit: Estimation of the absolute (uper) and proportional (lower) bias using averaged and Linnet's weighted Deming regression The 95% confidence limits (dotted lines) were calculated using the jackknife method.

5.2 Example 2: Instrument comparison

5.2.1 Data

The data contains measurements of biomarker B in blood. The blood samples were produced artificially for the instrument comparison. For each specimen there are 6 measurements. Three measurements were conducted using the Elecsys e601 system in combination with 3 instruments: E, EC, and EE. Another three measurements were conducted using the Elecsys e602 system in combination with 3 instruments: E, EC, and EE. We would like to compare system 602 with system 601 and estimate proportional bias at a fixed medical decision point of 1.0 (the measurement values are scaled so that the cutoff of interest equals 1.0).

We assume that the replications which are made with instruments E, EC and EE are exchangeable. This assumption may be violated. To check it, we generate a new data set with randomly permuted replications. In the following we compare the results of regression analysis on the basis of original and permuted data set. If the results are similar, we can conclude that the analysis was correct, because the assumption about exchangeability of the replications is not violated.
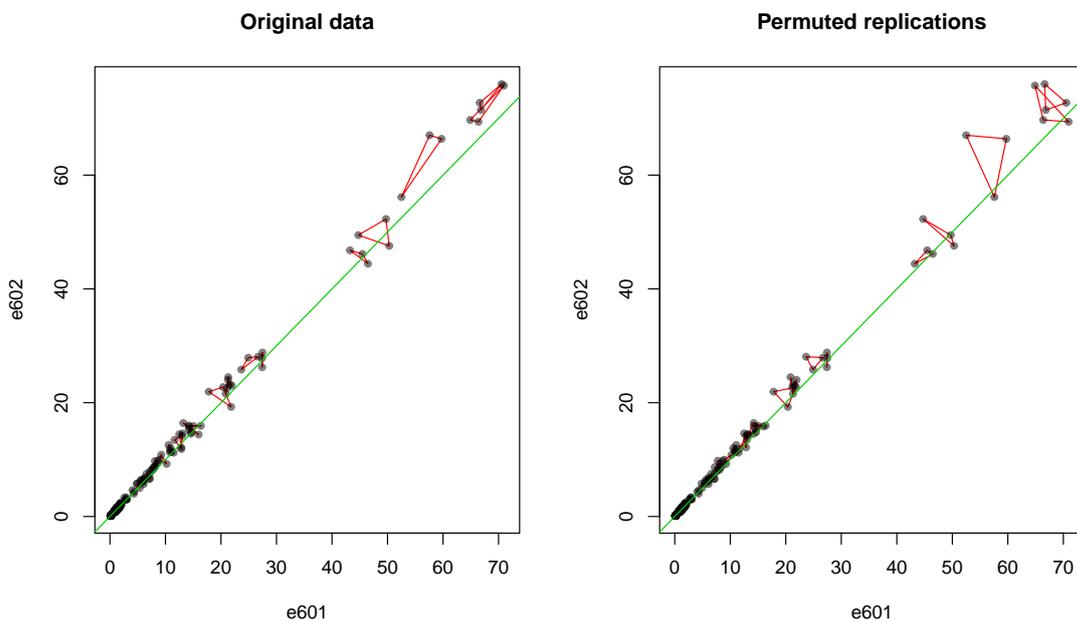
Figure 5.10: Example 2. Scatter plot for measurements of system e601 vs. system e602. Left: original data, right: data with randomly permuted replications. The replications are connected with red lines.

5.2.2 Diagnostic plots

- Figure 5.10 shows scatter plots of the original data and data with permuted replications. The scatter plots are quite similar. That is why the assumption about the exchangeablilty of replications may hold. The distribution of the measurements is strongly skewed: most observations are in range from 0 till 30.

- In order to choose between the averaged and Linnet's weighted regression we use Bland-Altman plots for the averaged replications (fig. 5.11). Comparison of Bland-Altman plots for absolute and relative differences between two systems yields clear choice of weighted Deming regression. We can also see that the relative differences lay above zero: there is positive proportional bias between systems e602 and e601.

- We can confirm the choice of Linnet's weighted Deming regression plotting absolute deviations between individual and averaged measurement replications vs. averaged measurement replications. We can see that both of the measurement errors are heteroscedastic (fig 5.12).
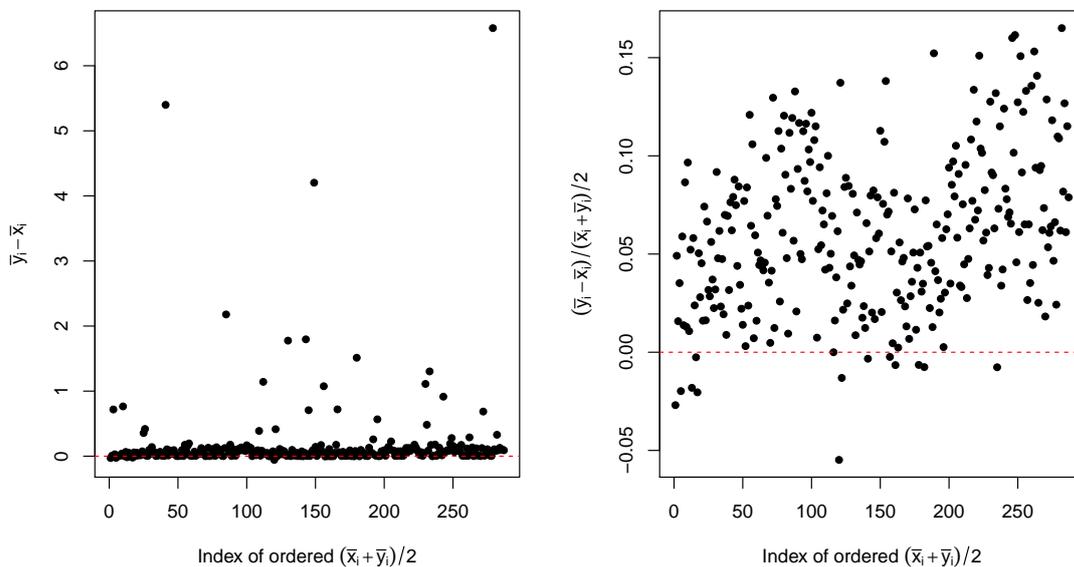
Figure 5.11: Example 2. Bland-Altman plots for the averaged replications of original data.Left scatter plot shows absolute differences between systems e601 and e602, the right figure shows the proportional differences. Here "X" means "e601", "Y" means "e602". **Note:** The distribution of concentrations is extremely skewed. That is why the x-axis contains instead of the averaged means their indexes.

114
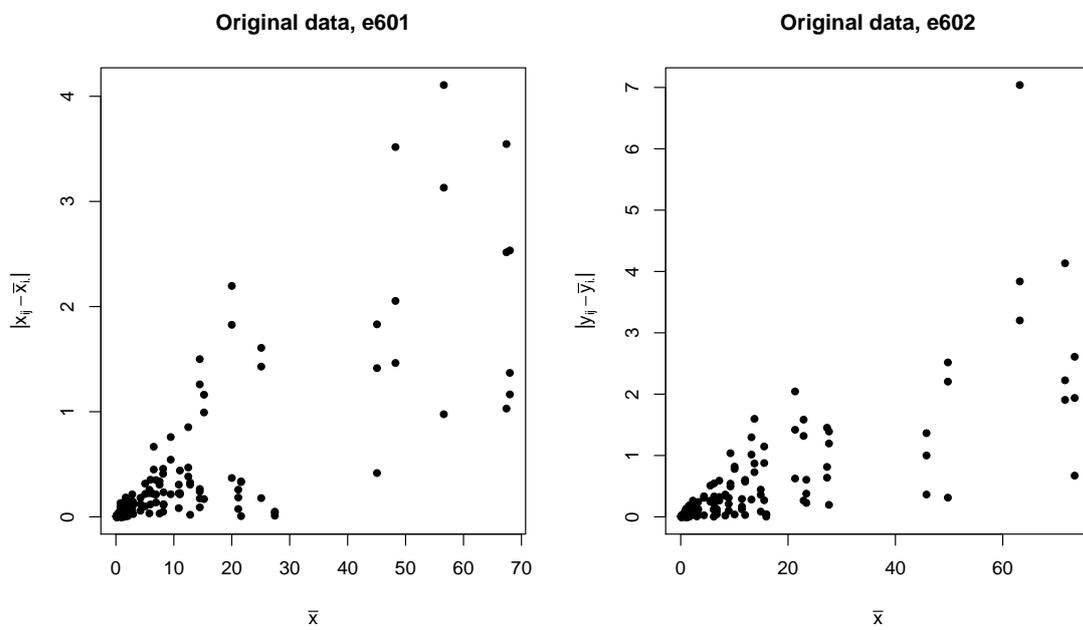
Figure 5.12: Example 2. Absolute differences between replications and averaged means of replications plotted against the averaged replications of each item. Here "X" means "e601", "Y" means "e602".

5.2.3 Regression analysis

Thus the analysis suggests, Linnet's weighted Deming aproach is the best method to choose for this data. However we consider estimates of all those methods which were looked at this master thesis. We would like to compare them using real world data. Table 5.3 shows estimates of regression coefficients, λ and matrix effect. The CIs for the regression coefficients and matrix effect are estimated using jackknife method. The CIs for λ are estimated using bootstrap quantile method. The calculations have shown that:

- We can see that the intercept is very close to zero (-0.007 (-0.008; -0.006)).

- The estimation of slope via Linnet's weighted Deming regression (1.081 (1.074; 1.087)) is bigger than the estimation of the averaged Deming regression (1.067 (1.042; 1.091)), ML and MM estimates.

- The CI for matrix effect (-0.180; 0.004) contains zero. We can assume that there is no matrix effect in this data. The simulation studies jhave shown that in such case the estimates of ML and MM estimates and the estimates of Linnet's weighted Deming regression are very similar.

- The estimation of λ using weighted Deming regression is 1.864 (1.636; 2.113). The measurement error of e602 is considerably higher than the measurement error of e601.

	n	Intercept	Slope	Lambda	Matrix effect
ave.Deming	287	0.001 (-0.033; 0.035)	1.067 (1.042; 1.091)	1.500 (0.880; 2.212)	
WDeming	287	-0.007 (-0.008; -0.006)	1.081 (1.074; 1.087)	1.864 (1.636; 2.113)	
ML	287	0.001 (-0.033; 0.035)	1.067 (1.042; 1.091)	1.461 (0.884; 2.019)	
MM	287	-0.001 (-0.035; 0.034)	1.067 (1.042; 1.092)		-0.088 (-0.180; 0.004)

Table 5.3: Example 2. Estimated regression parameters.
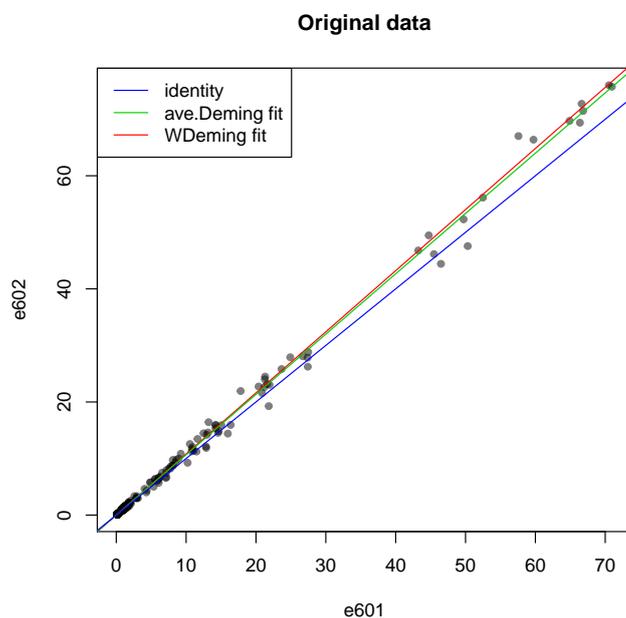
Figure 5.13: Example 2. Scatter plots for measurements of system e601 vs. measurements e602 with averaged Deming and Linnet's weighted Deming regression fit.

Figure 5.13 shows scatter plot of the measurements of system e601 vs. measurements of system e602 together with the averaged- and weighted Deming regression fit. We can see that both fits are very close to each other, but considerably far from the identity line.

Figure 5.15 compares CIs for slope and intercept which were estimated using different methods. For this data:

- We can see that analytical confidence intervals are too narrow.

- The nonparametrical methods yield very similar CIs.

- The CIs for ML- and MM-estimates and for estimates of the averaged Deming regression are considerably longer than the CIs of Linnet's weighted Deming regression.

- Red CIs were calculated using data with randomly permuted replications. We can see that permutation of replications doesn't affected considerably the estimates. Thus we can assume that the assumption about the exangeability of the replications is not violated.

	Bias
abs.bias: WDeming	0.074 (0.068; 0.080)
prop.bias: WDeming	0.074 (0.068; 0.080)

Table 5.4: Estimation of bias and proportional bias at the MDP of 1.0 with jackknife confidence intervals.

Averaged Deming regression　　　　**Linnet's weighted Deming regression**
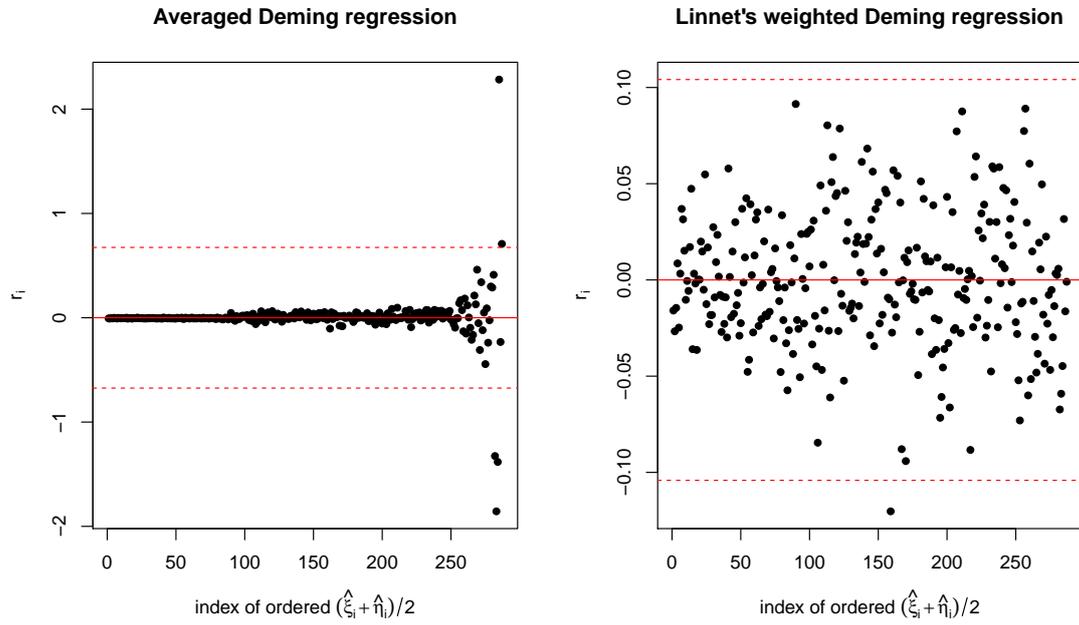
Figure 5.14: Example 2. Residual plots for the averaged- and Linnet's weighted Deming regression fit for original data. The red dotted line shows the bounds of 3 standard deviations of the distribution of residuals. The data points over (under) this bounds we can consider as outliers.

Fig. 5.14 shows residual plots of the averaged- and Linnet's weighted Deming regression fit. It is clear that Linnet's weighted Deming regression fit is the correct model for this data.
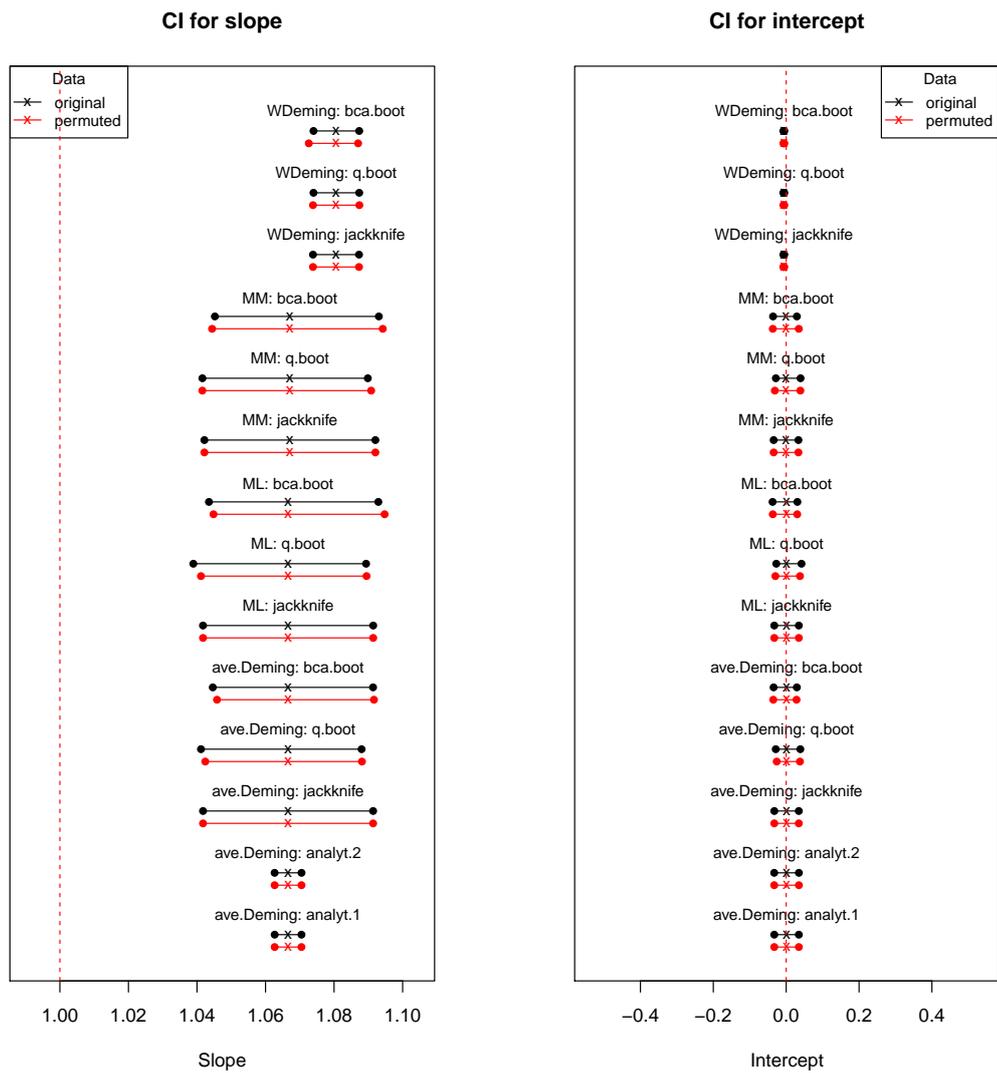
Figure 5.15: Comparison of confidence intervals for regression coefficients for original data (black) and for data with permuted replications (red).

Table 5.4 contains estimation of bias at medical decision point of 1.0. Estimation of proportional bias is considerably high (0.074 (0.068; 0.080)). It lays above 5% (look at pict. 5.16).

Conclusions The analysis of data suggests that:

- The measurements of both systems e601 ans e602 have high proportional bias at the medical decision point $Xc = 1.0$ (0.074 (0.068; 0.080)).

- The Linnet's weighted Deming regression yields slope of 1.081 (1.074; 1.087), intercept of -0.007 (-0.008; -0.006) and λ of 1.864 (1.636; 2.113).

- The measurements of both the systems are not exchangeable because of high bias and considerably higher variance of the measurement of the system e602.

baseline visit: Absolute bias

Absolute bias

Medical decision point

95% CI limits (dotted line) are calculated using bootstrap.

baseline visit: Proportional bias

Proportional bias

+3%

−3%

Medical decision point

95% CI limits (dotted line) are calculated using bootstrap.

Figure 5.16: Example 2. Estimation of the absolute (up) and proportional (low) bias using Linnet's weighted Deming regression The 95% confidence limits (dotted lines) were calculated using jackknife method.

Chapter 6

Conclusions and discussion

The master thesis deals with the analysis of the outcome of method comparison experiments with replications. The conduction and analysis of such experiments should follow the recommendations of the Clinical and Laboratory Standards Institute (CLSI), which are described in a special guide-line "Method comparison and bias estimation using patient samples; approved guideline" (CLSI EP9). This document suggests the following approaches for the analysis of method comparison data:

- Averaged Deming regression (in case of homoscedastic data).

- Linnet's weighted Deming regression (in case of heteroscedastic data with constant CV).

- Passing-Bablok method (robust nonparametric approach which makes very few assumptions about the distribution of measurement errors).

The investigation of the theoretical foundation and practical performance of the first two methods is the main scope of this work.

The averaged Deming regression method is based on the linear structural error-in-variables model with replications and is closely related to the Deming regression approach for the same model without replications, which assumes that the ratio of variances of measurement errors λ is known. A literature search of this topic has revealed that the averaged Deming regression approach is closely connected to the maximum likelihood approach for linear structural ME model with replications, described in Chan and Mak [6]. The simulation studies, which are conducted in this work show that the estimates of both methods are very similar to each other in most situations. The simulations suggest that in case of homoscedastic data and in case of heteroscedastic data with constant CV, both the ML-method and the averaged Deming regression yield unbiased estimates independent of the combination of the true intercept, slope and λ of simulated data. The estimates of regression coefficients of both methods are relatively robust against violation of normality assumptions. The results of both methods were acceptable even in case of simulated data with Laplace distributed measurement errors with 2 replications.

Linnet's weighted Deming regression was developed by Linnet [7] for the case of heteroscedastic data with constant CV. This is a weighted version of the averaged Deming regression. The weights of the data points are updated iteratively. The algorithm is very fast and converges usually after 3 or 4 iterations, if all measurements are positive. The simulationsshow that this method yields unbiased estimates in case of homoscedastic and heteroscedastic data, independent of the combination of the true slope, intercept and λ of the simulated data. In case of heteroscedastic data the estimates have considerably smaller variance than the estimates of the averaged Deming regression. The choice between the averaged and weighted Deming regression can be supported by diagnostic plots, described in section 3.5.

In subsection 3.3.5, Barnett's approach [15] for the linear functional ME model with replications is considered. This method assumes individual variances of measurement errors for each item. Simulation studies have shown that this method yields biased estimates if the data have an intercept which differs from zero. The bigger the intercept the bigger the bias of the estimated regression coefficients. Centering of the data improves the estimates considerably, but the variance of the results is rather high. Hence we would not recommend to use the current implementation of this approach for the analysis of method comparison data.

In section 3.4 we deal with the problem of a matrix effect in the test method. A matrix effect is the difference between the response of equally concentrated analytes in solvent and in matrix extracts (urine or blood, for example). According to the recommendations of Dunn et.al. [10] and Carstensen [16] we consider the matrix effect to be a random effect. If the variance of the matrix effect is considerably larger than the variance of the measurement errors it is considered to be strong matrix effect. In this case the data are clustered. It may be identified by looking at a scatter plot of the data. Our simulation studies have shown that the averaged Deming regression, ML-method and weighted Deming regression yield considerably biased results if there is strong matrix effect in the test method. In this case the ML estimates of regression coefficients are less biased than the estimates of the averaged Deming regression and Linnet's weighted Deming regression. Thus, ignoring of matrix effect in the test method leads to biased estimates of regression coefficients.

In section 3.4 we consider the linear structural ME model with error in equation, which allows for a matrix effect in the test method. The performance of MM-estimates of the parameters of this model, derived at Fuller [2] and Oman et.al. [9] are investigated in simulation studies. They show that this method yields unbiased estimates in case of data with matrix effect. The simulations suggest that in case of data without matrix effect in the test method, the estimates of the linear structural ME model with error in equation are unbiased and close to the ML estimates of the same model without error in equation. The estimation of the variance of the matrix effect is symmetrical around zero. That is why we can use it as a kind of indicator: if the CI of matrix effect does not contain zero, there is matrix effect in the data [9].

CLSI EP9 requires no matrix effect in the reference method [18]. However we have investigated the performance of all considered methods in case both measurement procedures contain matrix effect with the same variance. The simulation studies show that in this case the MM estimates of the structural model with error in equations are biased while the esimations of the structural model without error in equation (ML-estimation, the averaged and Linnet's weighted Deming regression approaches) yield reasonable results. It is not easy to visually detect, whether only the test method or both methods contain a matrix effect. This problem may be solved by comparison of the within variance of each item with between-items variances in X and Y direction. The master thesis does not investigate this problem.

In Chapter 3 different methods for the calculation of CIs for estimates are considered:

- two analytical methods for the estimates of regression parameters of averaged Deming regression. One of them (formulas 3.30, 3.33) is recommended by CLSI EP9 [18].

- the jackknife method, recommended by Linnet [7] and CLSI EP9 [18].

- the nonparametric bootstrap (quantile abd BC_a CIs),

- the parametric bootstrap method.

The analytical approach can be used if the model assumptions of the averaged Deming regression (normality and homoscedasticity of measurement errors) hold. The nonparametric or semiparametric approaches should be used if the model assumptions are violated or if we dont't have any analytical method for calculation of variances of estimates.

The performance of all methods mentioned above (except parametric bootstrap) are investigated via simulated studies. Therefore three scenarios of simulated data (with 2 and 3 measurement replications) are considered:

- homoscedastic measurement errors,

- heteroscedastic measurement errors with constant CV,

- heteroscedastic data with Laplace distributed errors.

Our simulations show that the analytical CIs for the averaged Deming regression have proper coverage in case of homoscdastic errors. In case of heteroscedastic data the analytical methods yield poor results.

The jackknife method of calculation of CIs shows the best performance in case of data with homoscedastic errors and in case of data with heteroscdastic errors with constant CV. In case of data with Laplace distributed errors with two replications the jackknife CIs for regression coefficients still achieve a coverage of 0.94-0.95.

The bootstrap quantile CIs are slighty shorter than jackknife CIs, but have mostly similar performance. Only in a few simulations the coverage is worse to jackknife CIs. For example in case of data with Laplace distributed measurement errors the results of the bootstrap CIs are poor. This had to be expected because the bootstrap CIs may be strongly affected by outliers.

Bootstrap BC_a CIs are the improved quantile CIs which theoretically should have better performance. Our simulations suggest that, against expectations, the coverage of the bootstrap BC_a CIs is slightly worse than the coverage of the corresponding bootstrap quantile CIs, for example in case of heteroscedastic data with constant CV with 2 replications.

Our simulations suggest that a higher number of replications can improve performance of all nonparametric CIs in scenarios with homoscedastic errors and in case of heteroscedastic errors with constant CV. The coverage gets closer to the desired level of 0.95 and the CIs get smaller.

The choice of the correct model leads to a better performance of the CIs: for example, in case of heteroscedastic data with constant CV, all CIs of the estimates of the Linnet's weighted Deming regression are shorter. The coverage of the bootstrap BC_a CIs for the estimates of this method achieve 0.95, while the same CIs for the estimates of the averaged Deming regression have coverage of 0.93 in case of data with 2 replications.

The poorest results show the CIs for the estimates of λ. In case of homoscedastic data or heteroscedastic data with constant CV, the coverage of the CIs is 0.93 in most cases. In case of data with Laplace distributed errors our simulation suggests a coverage of zero. The estimates of measurement errors are probably strongly affected by many outliers, that is why the estimates of λ are biased and the CIs have poor coverage. It is interesting that although λ is incorrectly estimated, the estimates of the regression coefficients stay reasonable in case of Laplace distributed errors.

The performance of the parametric bootstrap is not considered in the master thesis. This method has some advantages compared with the nonparametric bootstrap. In case of replicated measurements one item may contain a lot of information. By bootstrapping items (nonparametric bootstrap), we exclude a large amount of information from the bootstrap data set and thus overestimate the variance of estimates. This is especially important if we have unbalanced data (each item having different number of replications). Using parametric bootstrap we can generate balanced samples from the estimated distribution of the data.

Most of the methods of estimations considered in this work assume balanced data. In this case the nonparametric bootstrap may have less disadvantages compared with the parametric bootstrap. Otherwise, use of parametric bootstrap needs clear assumptions about the underlying model. In case of Linnet's weighted Deming regression it is not clear, how to generate the sample data. That is why we have considered only the confidence intervals based on the nonparametric bootstrap in this master thesis.

In general, we can assert, that the methods, which are recommended by CLSI EP9 (averaged Deming regression and Linnet's weighted Deming regression) in combination with jackknife or bootstrap quantile confidence intervals yield reasonable estimates of regression coefficients even if the model assumptions are considerably violated, if there is no strong matrix effect in the test method.

Chapter 7

Contents of the attached CD-ROM

File/object name	Folder	Description
mcCalcCI.r.r	algorithms	Contains functions for calculation of the different kinds of CIs.
calcEP9.r	algorithms	Calculation of the averaged Deming regression fit.
calcML.r	algorithms	Calculation of the ML estimation for the linear structural ME model with replications.
calcML_Barnett_old.r	algorithms	Calculation of the ML estimations for the linear functional ME model with replications
calcML_Barnett.r	algorithms	Calculation of the ML estimations for the linear functional ME model with replications. Improved version.
calcMM.r	algorithms	Calculation of MM estimation for the linear structural ME model with error in equation (for the case of matrix effect in test method)
compare.r	algorithms	Comparison of fits of the difference methods
diagnostic_plots.r	algorithms	Contains functions for drawing of diagnostic plots.
		Continued on next page

File/object name	Folder	Description
funPolynoms.r	algorithms	Contains functions for calculation of roots of the polynomials of 2nd, 3rd, and 4th degree. This functions are used to calculation of the ML estimation for the linear structural ME model with replications.
jackknife.r	algorithms	Returns results of the jackknife resampling.
makeWData.r	algorithms	Creates WData object
rounddigits.r	algorithms	Returns rounded numbers in string format.
runall.r	algorithms	Runs all programs from folder algorithms
show.r	algorithms	Draws scatter plot of the data, which contains WData-object.
simData.r	algorithms	Programs for simulation of data.
unparBoot.r	algorithms	Returns the results of bootstrap resampling.
wdeming.r	algorithms	Calculation of Linnets weighted Deming regression fit.
Demo.r	algorithms	Demonstration of the most important algorithms.
functions_constSD.r	/simulations	Different programs for conduction of simulations 1.1-1.5 and the analysis of their results.
functions_heterosked.r	/simulations	Different programs for conduction of the simulations 1.6-1.8, 2.1-2.3 and the analysis of their results.
functions_simulations.r	/simulations	Functions for the analysis of the simulation results.
simulation2_1-2_3.r	/simulations	Siulations 2.1-2.3
simulations1_1-1_8.r	/simulations	Simulations 1.1-1.8

Continued on next page

File/object name	Folder	Description
analysis_of_sim_results.r	/simulations	Analysis of the simulation results (Chapter 4).
	/simulations/ Simulation results	Folder with results of the simulations
functions.r	/data_analysis	Functions for the analysis of the example data sets.
data_analysis.r	/data_analysis	Analysis of the example data sets (Chapter 5).
	/data_analysis /data	Data sets and additional objects
example1.rdata	/data_analysis /data	Data for the example 1 (Chapter 5)
data01sas.csv	/data_analysis /data	Data for the example 2 (Chapter 5)

Bibliography

[1]   C.Cheng, J.Ness: *Statistical Regression with Measurement Error.* London, 1999.

[2]   W.A.Fuller: *Measurement error models.* New York, 1987.

[3]   R.G.Miller: *Grundlagen der angewandten statistik.* Oldenbourg Wissenschaftsverlag, 1995.

[4]   L.Held: *Methoden der statistischen Inferenz.* Heidelberg, 2008.

[5]   B.Efron, R.J.Tibshirani: *An introduction to the bootstrap.* Chapmann & Hall/CRC, 1993.

[6]   L.K.Chan, T.K.Mak: *Maximum likelihood estimation of a linear relationship with replication. J.R.Statist.Soc.* B , **41**, 2, 263-268(1979).

[7]   K.Linnet: *Estimation of the linear relationship between the measurements of two methods with proportional errors. Statistics in Medicine*, **9**, 1463-1473 (1990).

[8]   K.Linnet: *Evaluation of Regression Procedures for Methods Comparison Studies. CLIN.CHEM*, **39**, 3, 424-432 (1993).

[9]   S.D.Oman, N.Meir: *Comparing two measures of creatinine clearence: an application of errors-in-variables and bootstrap techniques. Appl.Statist.*, **48**, 1, 39-52 (1999).

[10]   G.Dunn, C.Roberts: *Modeling method comparison data. Stat. Methods. Med. Res.*, **8**, 1, 161-179 (1999).

[11]   G.Kelly: *The influence function in the errors in variables problem. The annals of statistics*, **12**, 1, 87-100 (1984).

[12]   M.Lagutin: *The evident mathematical statistics.* Moscow, 2007.

[13]   S. W. Looney: *Statistical Methods for Assessing Biomarkers. Methods in Molecular Biology*, **184**: Biostatistical Methods, 86-108 (2001).

[14]   D.G.Altman, J.M.Bland: *Measurement in medicine: the analysis of method comparison studies. Statistician*, **32**, 307317 (1983).

[15]   V.D.Barnett: *Fitting straight lines - the linear functional relationship with replicated observations. Journal of the royal statistical society. Series C (Applied statistics)*, **19:** 2, 135-144 (1970).

[16]   B.Carstensen: *Comparing clinical measurement methods.* John Wiley & Sons Ltd, 2010.

[17]   L.Fahrmeir, T.Kneib, S.Lang: *Regression: Modelle, Methoden und Anwendungen (Statistik und ihre Anwendungen).* Berlin, 2007.

[18]   Clinical and Laboratory Standards Institute: *Method comparison and bias estimation using patient samples; approved guideline.* 2. Wayne, PA, USA: CLSI; 2005. CLSI document EP9-A2.

List of Figures