Estimation, testing, and prediction in studies of comparing methods of measurement

Kai Wang
Department of Biostatistics, University of Iowa, Iowa City, IA 52242, USA

Abstract

There are two main issues in comparing different methods of measurement: whether the difference between different methods is constant across subjects and how to predict measurement of one method from others. These issues are complicated in the case of one measurement per method per subject because of the confounding of non-additivity and heteroscedasticity. Because of the error in each measurement, the ordinary least square (OLS) method proposed by [Carstensen(2010)] is biased. We propose a maximum-likelihood-based procedure for estimation and prediction that is applicable to multiple methods. We suggest testing for non-constant difference of measurements by using Mandel's test. The bias of the OLS method is analytically investigated. Although the OLS method is biased, it is shown that the test of non-constant difference for two methods used by [Carstensen(2010)] is valid and is identical to Mandel's test.

Keywords: method comparison; Bland-Altman plot; non-constant difference; prediction; Mandel test

1 Introduction

In medical research there is often a need to compare two or more methods of measurement to determine if they can be used interchangeably after certain adjustment or calibration. The Bland-Altman plot and the 95% limits of agreement (LoA) are popular when the difference between the methods is constant [Bland and Altman(1986)]. This LoA method for assessing method agreement has been extended to the case of non-constant difference [Bland and Altman(1999)].

In a recent research, [Carstensen(2010)] proposed a prediction method for two methods based on the ordinary least square (OLS) regression of the differences on the averages. The goal is to achieve prediction equations that are equivalent regardless which method is chosen to be the predicting one and which is the predicted one. However, as we will show, because of the correlation between the averages and the regression error terms, the OLS estimate of the regression intercept and the slope are biased and so are the prediction equations. [Carstensen(2010)] uses the regression F statistic to test whether there is non-constant difference and claims this procedure is valid even under heteroscedasticity. However, as we will show, this claim is incorrect. This test is identical to the Mandel's F test for non-additivity [Mandel(1961)] and is valid only under homoscedasticity. This result is interesting considering that the OLS estimates of the intercept and the slope are biased.

We propose a method that is based on the maximum likelihood. This method is applicable to the case of more than two methods. We focus on the case of no repeated measurement where there is only one measurement per method per subject. Confounding of non-additivity and heteroscedasticity implies some constraints on model parameters. We note that the non-constant difference of measurements corresponds to method-subject interaction. Hence the existing tests for non-additivity, such as Tukey's test [Tukey(1949)], Mandel's test [Mandel(1961)], and others [Šimeček and Šimečková(2012), Franck et al.(2013)], apply. The Mandel's test is especially pertinent.

This paper is organized as follows. We first define the model and discuss parameter estimation, testing for non-additivity, and prediction in turn. An analytical study is conducted on OLS-based method of [Carstensen(2010)]. The analytical results are demonstrated through simulation studies. After illustrating the proposed procedure through an example, this paper concludes by a discussion.

2 Methods

Suppose that there are I subjects and M methods. The measurement of method m on subject i is modeled

$$y_{im} = \alpha_m + \beta_m \mu_i + e_{im}, \quad e_{im} \sim N(0, \sigma^2),\ i = 1, \ldots, I,\ m = 1, \ldots, M \qquad (1)$$

where α_m and β_m are model parameters and μ_i is the true but unknown value for individual i. That is, there is only one measurement of each method on each subject. Non-constant β_m across m implies interaction between methods and subjects. The error terms e_{im} are assumed to share the same variance σ^2) (i.e., homoscedasticity) as heteroscedasticity confounds with method-subject interaction [Snee(1982)]. These models are the same as those in equation (2) of [Carstensen(2010)] except that M is not limited to 2 and heteroscedasticity is not allowed. Let the subscript $\cdot$ denote an arithmetic average over the values corresponding to all values of that subscript. For instance, $\mu_\cdot = I^{-1} \sum_i \mu_i$. Equations in (1) can be written

$$y_{im} = (\alpha_m + \beta_m \mu_\cdot) + \beta_m(\mu_i - \mu_\cdot) + e_{im}.$$

So without loss of generality, it can be assumed that $\sum_i \mu_i = 0$ and α_m can be interpreted as the expected measurement of method m on an "average" subject whose "true" value is $\mu_\cdot$. Furthermore, multiplying the β_ms by a number while dividing each μ_i by the same number does not change the product $\beta_m \mu_i$. β_m is normalized such that $\sum_m \beta_m = M$. In summary, the number of α_m parameters is M; the number of β_m parameters is $(M-1)$ and the number of μ_i parameters is $(I-1)$. Including σ^2, the total number of parameters is $M + (M-1) + (I-1) + 1 = 2M + I - 1$.

2.1 MLE of model parameters

The log-likelihood function for model (1) is, up to an additive constant,

$$l(\alpha_1, \ldots, \alpha_M, \beta_1, \ldots, \beta_M, \mu_1, \ldots, \mu_I, \sigma^2) = -\frac{1}{2} \sum_i \sum_m \left[\ln \sigma^2 + \frac{(y_{im} - \alpha_m - \beta_m \mu_i)^2}{\sigma^2} \right].$$

Proposition 1 *The maximum likelihood estimate (MLE) of α_m is*

$$\hat{\alpha}_m = \bar{y}_{\cdot m}, i = 1, \ldots, M.$$

The MLEs of β_m and μ_i are determined jointly by

$$\hat{\beta}_m = 1 + \frac{\sum_i (y_{im} - y_{i\cdot})\hat{\mu}_i}{\sum_i \hat{\mu}_i^2}, \quad m = 1, \ldots, M$$

and

$$\hat{\mu}_i = \frac{\sum_m (y_{im} - y_{\cdot m})\hat{\beta}_m}{\sum_m \hat{\beta}_m^2}, \quad i = 1, \ldots, I.$$

These two expressions can be used to iteratively solve for $\hat{\beta}_m$s and $\hat{\mu}_i$s, for instance, by setting the initial value for μ_i to $y_{i\cdot} - y_{\cdot\cdot}$. Once the MLEs for α_m, β_m and μ_i are obtained, the MLE of σ^2 is given by

$$\hat{\sigma}^2 = \frac{1}{IM} \sum_i \sum_m (y_{im} - \hat{\alpha}_m - \hat{\beta}_m \hat{\mu}_i)^2.$$

A derivation of these equations for MLEs is given in Appendix A.

2.2 Testing for non-constant difference

One important question in comparing different methods of measurement is whether the difference between any two methods is subject-independent. That is, whether there exists subject-method interaction. If there is no interaction, prediction of the measurement of one method from another method is simply obtained by adjusting for the difference. Otherwise it is more complicated. In terms of model (1), this mounts to test the following null hypothesis

$$H_0 : \beta_1 = \beta_2 = \cdots = \beta_M.$$

Because of the restriction $\sum_m \beta_m = M$, this hypothesis is equivalent to requiring $\beta_m = 1, m = 1, \ldots, M$. Under this hypothesis, the MLE of μ_i has an explicit expression, which is $y_{i\cdot} - y_{\cdot\cdot}, i = 1, \ldots, I$. One straightforward way of testing H_0 is to use the likelihood ratio test. However, its performance is unsatisfactory when M is typically smalls (say, less than 10).

The model (1) is methods-linear: for any method m, the expected measurement $E(y_{im})$ is a linear function of μ_i. The test proposed by [Mandel(1961)] for detecting non-additivity in rows-linear models is exactly for testing H_0. The test statistic is

$$T = \frac{\sum_m (b_m - 1)^2 \sum_i (y_{i\cdot} - y_{\cdot\cdot})^2}{\sum_i \sum_m [(y_{im} - y_{\cdot m}) - b_m(y_{i\cdot} - y_{\cdot\cdot})]^2} \cdot (I - 2), \tag{2}$$

where

$$b_m = \frac{\sum_i y_{im}(y_{i\cdot} - y_{\cdot\cdot})}{\sum_i (y_{i\cdot} - y_{\cdot\cdot})^2}.$$

Under H_0, T follows an F distribution with $(M-1)$ and $(M-1)(I-2)$ degrees of freedom.

2.3 Prediction

Once the model parameters in (1) are estimated, model (1) can be used to make prediction of a measurement of a method based on the measurements of other methods. Suppose a measurement is to be predicted for method m_0. Its expected value is

$$E(y_{m_0}) = \alpha_{m_0} + \beta_{m_0}\mu.$$

Since the value of μ is unknown, it is estimated from the measurements of other methods, for instance, by lease square method:

$$\hat{\mu} = \frac{\sum_{m \neq m_0} \beta_m (y_m - \alpha_m)}{\sum_{m \neq m_0} \beta_m^2}.$$

A prediction on y_{m_0} is

$$\alpha_{m_0} + \beta_{m_0}\hat{\mu}.$$

Considering sampling error, the prediction variance is

$$\beta_{m_0}^2 Var(\hat{\mu}) + Var(e_{m_0}) = \sigma^2 \left(\frac{\beta_{m_0}^2}{\sum_{m \neq m_0} \beta_m^2} + 1 \right)$$

The α_ms, β_ms, and σ^2 in the previous two expressions will be replaced by their respective MLEs obtained from training data.

When $M = 2$, the prediction equation for method $m = 1$ from method $m = 2$ is given by

$$E(y_1) = \alpha_1 + \frac{\beta_1}{\beta_2}(y_2 - \alpha_2). \tag{3}$$

with prediction variance

$$\sigma^2(1 + \beta_1^2/\beta_2^2) \tag{4}$$

This prediction equation and the associated prediction variance are the same as those from [Carstensen(2010)]. However, as shown in the next section, the estimates of the model parameters used in [Carstensen(2010)] are biased.

3 Comparison to the method in [Carstensen(2010)]

In an effort to construct a prediction equation between two methods that is invariant to the choice of the predicting method and the predicted method, [Carstensen(2010)] considered the case where $M = 2$. The following regression of differences on averages is used:

$$D_i = a + bA_i + e_i. \tag{5}$$

This model is derived from model (1) with $M = 2$. Regression (5) relates to model (1) in the following way:

$$\begin{aligned} a &= (\alpha_1 - \alpha_2) - (\alpha_1 + \alpha_2)(\beta_1 - \beta_2)/(\beta_1 + \beta_2) \\ b &= 2(\beta_1 - \beta_2)/(\beta_1 + \beta_2) \\ e_i &= 2(\beta_2 e_{i1} - \beta_1 e_{i2})/(\beta_1 + \beta_2). \end{aligned}$$

In addition, e_{i1} and e_{i2} are allowed to have different variances σ_1^2 and σ_2^2, respectively. [Carstensen(2010)] advocates estimating a and b by their ordinary least square (OLS) estimator in regression (5) and construct prediction equations from these estimates (equation (6) of [Carstensen(2010)]). In the R package `MethComp` by the same author, non-constant difference between the two methods is tested by the F-statistic of regression (5), which is

$$F = \frac{\hat{b}^2_{\text{OLS}} \cdot \sum_i (A_i - A.)^2}{\sum_i [(D_i - D.) - \hat{b}_{\text{OLS}}(A_i - A.)]^2/(I-2)} \sim F(1, I-2) \tag{6}$$

where

$$\hat{b}_{\text{OLS}} = \frac{\sum_i (A_i - A.)(D_i - D.)}{\sum_i (A_i - A.)^2}.$$

Unfortunately the OLS estimates of a and b are biased. This is because in regression (5) the regressor A_i is correlated with the residual e_i. Let σ_u^2 denote the population variance of u. The correlation coefficient between A_i and e_i is

$$\begin{aligned} Cor(A_i, e_i) &= Cor(y_{i1} + y_{i2}, \beta_2 e_{i1} - \beta_1 e_{i2}) \\ &= \frac{Cov(y_{i1} + y_{i2}, \beta_2 e_{i1} - \beta_1 e_{i2})}{\sqrt{Var(y_{i1} + y_{i2})}\sqrt{Var(\beta_2 e_{i1} - \beta_1 e_{i2})}} \\ &= \frac{\beta_2 \sigma_1^2 - \beta_1 \sigma_2^2}{\sqrt{(\beta_1^2 + \beta_2^2)\sigma_u^2 + \sigma_1^2 + \sigma_2^2}\sqrt{\beta_2^2 \sigma_1^2 + \beta_1^2 \sigma_2^2}} \end{aligned}$$

which is not 0 unless $\beta_2 \sigma_1^2 - \beta_1 \sigma_2^2 = 0$.

We have the following results:

Proposition 2 *1. The bias in the OLS estimate of b is*

$$plim\ (\hat{b}_{OLS} - b) = \frac{4(\beta_2 \sigma_1^2 - \beta_1 \sigma_2^2)}{[(\beta_1 + \beta_2)^2 \sigma_u^2 + \sigma_1^2 + \sigma_2^2](\beta_1 + \beta_2)} \tag{7}$$

2. *The bias in the OLS estimate of a is*

$$plim\ (\hat{a}_{OLS} - a) = -plim\ (\hat{b}_{OLS} - b) \cdot \frac{\alpha_1 + \alpha_2}{2} \tag{8}$$

3. *The F-test from the regression (5) is identical to the Mandel's test and therefore is valid.*

The OLS of b is biased even if $\beta_1 = \beta_2$ given $\sigma_1^2 \neq \sigma_2^2$. The proof of these results is given in Appendix B. Because the OLS estimates used in [Carstensen(2010)] are biased, the prediction equations are biased as well.

4 Simulation Studies

Simulation studies were conducted to study the validity of the test of non-constant difference used by [Carstensen(2010)]. So the value of M is fixed at 2 and $\beta_1 = \beta_2 = 1$. Without loss of generality, data from $I = 101$ subjects are simulated from model (1) with

$$\alpha_1 = \alpha_2 = 0, \quad \mu_i = c(-5.0, -4.9, \ldots, -0.1, 0, 0.1, \ldots, 4.9, 5.0).$$

Like [Carstensen(2010)], the variance of the residual e_{i1} is allowed to be different from that of e_{i2}. They are denoted by σ_1^2 and σ_2^2, respectively. The number of simulation replicates is fixed at 1000. It has been verified that the F-statistic (6) from the regression of differences on averages is identical to the Mandes test statistic in every single replicate. Only results from the F-statistic is reported.

Table 1 presents the rejection rate of the F-statistic at significance level $0.1, 0.05$, and 0.01 for different combination of values for (σ_1^2, σ_2^2). It clearly shows that the F-statistic is valid only when there is homoscedasticity. It is highly inflated otherwise. These findings are consistent with the analytical studies presented in the previous section.

[Table 1 about here.]

Further simulation studies was conducted to demonstrate the bias of Carstensen's method and superiority of the MLE estimates. Data were simulated from model (1) with the same set of parameters as in the previous simulation except that (σ_1^2, σ_2^2) is fixed at $(1, 1)$ and β_1 and β_2 are allowed to change subject to $\beta_1 + \beta_2 = 2$. We focus on the estimate of the slope b in the regression of differences on averages (5) because in this setting the bias in the estimate of a by Carstensen's method is expected to be 0 (ref. (8)). Its true value is computed from the values used for β_1 and β_2. Results are presented in Table 2. The maximum likelihood method computes the MLE of b by substituting the MLEs of β_1 and β_2 into $2(\beta_1 - \beta_2)/(\beta_1 + \beta_2)$. These results indicate that the Carstensen's method is biased while the maximum likelihood method is consistent. It is easy to verify the bias is similar in magnitude to the analytical results computed from (7).

[Table 2 about here.]

5 An Example

We consider the blood glucose example used in [Carstensen(2010)]. This data is a subset of the `glucose` data included in the R package `MethComp`. It contains measurements of glucose based on venous plasma and capillary whole blood on 46 subjects. We were able to replicate the 95%prediction limits reported in [Carstensen(2010)], which are

$$\begin{aligned}\text{Plasma} &= -2.695 + 1.402 \text{ Capillary} \pm 2 \times 1.302\\ \text{Capillary} &= 1.922 + 0.713 \text{ Plasma} \pm 2 \times 0.928.\end{aligned}$$

In comparison, the MLE of (α, β) for Plasma is $(7.993, 1.183)$ and that for Capillary is $(7.624, 0.817)$. The MLE of σ^2 is 0.570. By (4), the prediction error is 1.328 for Plasma and is 0.918 for Capillary. By (3), the 95% prediction limits are

$$\begin{aligned}\text{Plasma} &= -3.041 + 1.447 \text{ Capillary} \pm 2 \times 1.328\\ \text{Capillary} &= 2.101 + 0.691 \text{ Capillary} \pm 2 \times 0.918.\end{aligned}$$

6 Discussion

We proposed a maximum likelihood method for studies of agreement of multiple methods. We covered the main issues related to such studies such as parameter estimation, testing, and prediction. As we have seen, it provides a general framework more theoretically sound than OLS methods. Importantly, it is applicable to comparison of more than two methods.

One natural question is that whether the likelihood ratio test can be used for testing non-constant difference. It turns out that since the number of methods M is typically small (for instance, < 10), simulation studies (results not shown) indicate that the likelihood ratio test is seriously inflated. However, the Mandel's test performs satisfactorily.

The programming of the method proposed in this work is straightforward. The author's R code used for the analysis in this paper is available upon request.

Appendix A Proof of Proposition 1

The partial derivatives for the log-likelihood function $l(\alpha_1, \ldots, \alpha_M, \beta_1, \ldots, \beta_M, \mu_1, \ldots, \mu_I, \sigma^2)$ are

$$\begin{aligned}\frac{\partial l}{\partial \alpha_m} &= \frac{1}{\sigma^2}\sum_i (y_{im} - \alpha_m - \beta_m \mu_i), \quad m = 1, \ldots, M\\ \frac{\partial l}{\partial \beta_m} &= \frac{1}{\sigma^2}\sum_i (y_{im} - \alpha_m - \beta_m \mu_i)\mu_i - \frac{1}{\sigma^2}\sum_i (y_{iM} - \alpha_M - \beta_M \mu_i)\mu_i\\ &= \frac{1}{\sigma^2}\sum_i (y_{im} - \beta_m \mu_i)\mu_i - \frac{1}{\sigma^2}\sum_i (y_{iM} - \beta_M \mu_i)\mu_i, \quad m = 1, \ldots, M-1\end{aligned}$$

$$
\begin{aligned}
\frac{\partial l}{\partial \mu_i} &= \frac{1}{\sigma^2}\sum_m [(y_{im} - y_{Im}) - \beta_m(\mu_i - \mu_I)]\beta_m, \quad i = 1, \ldots, I-1 \\
\frac{\partial l}{\partial \sigma^2} &= -\frac{IM}{2\sigma^2} + \frac{1}{2(\sigma^2)^2}\sum_i\sum_m (y_{im} - \alpha_m - \beta_m\mu_i)^2.
\end{aligned}
$$

Setting these equations to 0 and solve them for the unknown parameters, only $\hat{\alpha}_m$ has an explicit solution:

$$
\hat{\alpha}_m = \frac{1}{I}\sum_i (y_{im} - \hat{\beta}_m\hat{\mu}_i) = y_{\cdot m}.
$$

Other parameters mutually dependent. After simplification, we have

$$
\begin{aligned}
\hat{\sigma}^2 &= \frac{1}{IM}\sum_i\sum_m (y_{im} - \hat{\alpha}_m - \hat{\beta}_m\hat{\mu}_i)^2 \\
\beta_m &= 1 + \frac{\sum_i (y_{im} - y_{i\cdot})\mu_i}{\sum_i \mu_i^2}
\end{aligned}
$$

and

$$
\hat{\mu}_i - \hat{\mu}_I = \frac{\sum_m (y_{im} - y_{Im})\hat{\beta}_m}{\sum_m \hat{\beta}_m^2},
$$

or

$$
\hat{\mu}_i = \frac{\sum_m (y_{im} - y_{\cdot m})\hat{\beta}_m}{\sum_m \hat{\beta}_m^2}, \quad i = 1, \ldots, I.
$$

Appendix B Proof of Proposition 2

1.

$$
\begin{aligned}
\text{plim } \hat{b}_{\text{OLS}} - b &= \text{plim } \frac{\sum_i (A_i - A.)(D_i - D.)}{\sum_i (A_i - A.)^2} - \frac{2(\beta_1 - \beta_2)}{\beta_1 + \beta_2} \\
&= \frac{2[(\beta_1 + \beta_2)(\beta_1 - \beta_2)\sigma_u^2 + \sigma_1^2 - \sigma_2^2]}{(\beta_1 + \beta_2)^2\sigma_u^2 + \sigma_1^2 + \sigma_2^2} - \frac{2(\beta_1 - \beta_2)}{\beta_1 + \beta_2} \\
&= \frac{4(\beta_2\sigma_1^2 - \beta_1\sigma_2^2)}{[(\beta_1 + \beta_2)^2\sigma_u^2 + \sigma_1^2 + \sigma_2^2](\beta_1 + \beta_2)}
\end{aligned}
$$

2.

$$
\begin{aligned}
\text{plim } \hat{a}_{\text{OLS}} - a &= \text{plim } (D. - \hat{b}_{\text{OLS}}A.) - a \\
&= \text{plim } [(b - \hat{b}_{\text{OLS}})A. + (D. - bA. - a)] \\
&= -\frac{(\alpha_1 + \alpha_2) + (\beta_1 + \beta_2)\mu.}{2}(\text{plim } \hat{b}_{\text{OLS}} - b) \\
&= -\frac{\alpha_1 + \alpha_2}{2}(\text{plim } \hat{b}_{\text{OLS}} - b)
\end{aligned}
$$

3. The term b_1 in expression (2) satisfies

$$\begin{aligned} b_1 - 1 &= \frac{1}{\sum_i (y_{i\cdot} - y_{\cdot\cdot})^2} \sum_i (y_{i\cdot} - y_{\cdot\cdot})(y_{i1} - y_{i\cdot} + y_{\cdot\cdot}) \\ &= \frac{1}{\sum_i (A_i - A_\cdot)^2} \sum_i (A_i - A_\cdot)(D_i/2 + A_\cdot) \\ &= \frac{1}{2\sum_i (A_i - A_\cdot)^2} \sum_i (A_i - A_\cdot) D_i \\ &= \frac{\hat{b}_{\mathrm{OLS}}}{2}. \end{aligned}$$

Similarly,

$$b_2 - 1 = -\frac{\hat{b}_{\mathrm{OLS}}}{2}.$$

Furthermore,

$$\begin{aligned} y_{i1} - y_{\cdot 1} - b_1(y_{i\cdot} - y_{\cdot\cdot}) &= y_{i1} - y_{\cdot 1} - (\hat{b}_{\mathrm{OLS}}/2 + 1)(A_i - A_\cdot) \\ &= (y_{i1} - y_{\cdot 1} - A_i + A_\cdot) - (\hat{b}_{\mathrm{OLS}}/2 + 1)(A_i - A_\cdot) \\ &= \frac{1}{2}[(D_i - D_\cdot) - \hat{b}_{\mathrm{OLS}}(A_i - A_\cdot)] \end{aligned}$$

and

$$y_{i2} - y_{\cdot 2} - b_2(y_{i\cdot} - y_{\cdot\cdot}) = -\frac{1}{2}[(D_i - D_\cdot) - \hat{b}_{\mathrm{OLS}}(A_i - A_\cdot)]$$

Hence the denominator of the Mandel's test in (2) is one fourth times the residual sum of squares of the regression (5). Combining all these results, the T statistic in (2) is indeed equal to the F statistic in (6).

References

[Carstensen(2010)] Carstensen B. Comparing methods of measurement: Extending the LoA by regression. *Stat Med* 2010; **29**(3):401–410.

[Bland and Altman(1986)] Bland JM, Altman DG. Statistical methods for assessing agreement between two methods of clinical measurement. *Lancet* 1986; **i**:307–310.

[Bland and Altman(1999)] Bland JM, Altman DG. Measuring agreement in method comparison studies. *Stat Methods in Med Res* 1999; **8**(2):135–160.

[Mandel(1961)] Mandel J. Non-additivity in two-way analysis of variance. *Journal of the American Statistical Association* 1961; **56**(296):878–888.

[Tukey(1949)] Tukey J. One degree of freedom for non-additivity. *Biometrics* 1949; **5**(3):232–242.

[Šimeček and Šimečková(2012)] Šimeček P, Šimečková M. Modification of Tukey's additivity test. *arXiv* 2012; :1207.2883v1.

[Franck et al.(2013)] Franck CT, Nielsen DM, Osborne JA. A method for detecting hidden additivity in two-factor unreplicated experiments. *Computational Statistics and Data Analysis* 2013; **67**:95–104.

[Snee(1982)] Snee RD. Nonadditivity in a two-way classification: is it interaction of nonhomogeneous variance? *Journal of the American Statistical Association* 1982; **77**(379).

Table 1: Simulated rejection rate when there are two methods and their difference is constant (i.e., $\beta_1 = \beta_2 = 1$)

	Nominal Significance Level		
(σ_1^2, σ_2^2)	0.1	0.05	0.01
$(1,1)$	0.101	0.048	0.011
$(1,2)$	0.236	0.155	0.049
$(1,3)$	0.492	0.380	0.165
$(1,4)$	0.724	0.599	0.333

Table 2: Estimated slope b in the regression of differences on averages (5) by Carstensen's method and the maximum likelihood method. The value in the parenthesis is the standard error over 1000 simulation replicates.

	True Value	Method	
(β_1, β_2)	$(= 2(\beta_1 - \beta_2)/(\beta_1 + \beta_2))$	Carstensen	Maximum Likelihood
$(1, 1)$	0	0.001411 (0.048420)	0.001498 (0.051221)
$(1.2, 0.8)$	0.4	0.378420 (0.045241)	0.400397 (0.048032)
$(1.4, 0.6)$	0.8	0.756821 (0.048581)	0.800738 (0.145569)
$(1.6, 0.4)$	1.2	1.133011 (0.054472)	1.198421 (0.058434)