Limits of agreement: How to use the regression of differences on averages.

25 April 2008

Bendix Carstensen Steno Diabetes Center, Gentofte, Denmark
bxc@steno.dk
http://www.biostat.ku.dk/~bxc

Contents

Abstract

Method comparison studies are usually analyzed by computing limits of agreement. If only one measurement by each method is taken on each person, and the difference across the range is not constant, it has been recommended [1] to regress the differences on the averages and use the resulting equation to construct limits of agreement. The meaning of the regression differences on means is clarified in the framework of a proper model and improved prediction equations linking one method to another are devised. The performance of this model based method is evaluated against the simple approach earlier proposed and against Deming regression.

KEYWORDS: Method comparison, Bland-Altman plot, differences vs. averages, non-constant difference, regression.

1 Introduction

Comparing two methods of measurement is normally done by computing limits of agreement, i.e. prediction limits for a future difference between measurements with the two methods. When the difference is not constant it is not clear what this means, since the difference between the methods depends on the average, so unlike the case where the difference is constant, limits of agreement cannot directly be translated into a prediction interval for a measurement by one method given that of another, although this has been suggested [1]: "Of course, in clinical practice, when only one method is being used, the observed value by that method provides the value of a [the average of the two methods — my addition]."

In this paper I set up a proper model for data for method comparison studies which in the case of constant difference between methods lead to the classical limits of agreement (LoA), and in the case of linear bias gives a simple formula for the prediction.

2 Model for limits of agreement

The classical computation of limits of agreement is based on a model for measurements y_{mi} by method $m = 1, 2$ on item $i = 1 \ldots n$ (item is here used generically for individual or sample) which can be written:

$$y_{mi} = \alpha_m + \mu_i + e_{mi}, \quad e_{mi} \sim \mathcal{N}(0, \sigma_m^2) \tag{1}$$

Prediction limits for a difference between measurements by the two methods on a future item, 0, say, are based on the distribution of the difference under the model:

$$D_0 = y_{10} - y_{20} = \alpha_1 - \alpha_2 + e_{1i} - e_{2i} \sim \mathcal{N}(\alpha_1 - \alpha_2, \sigma_1^2 + \sigma_2^2)$$

Even though the separate variances cannot be determined the sum of them can be estimated by the empirical variance of the differences. Likewise the separate αs cannot be estimated, only their difference can be estimated as $\bar{D}_{\cdot}$.

The standard error of $\bar{D}_\cdot$ is $\sqrt{(\sigma_1^2+\sigma_2^2)/n}$, so a prediction interval for the difference between two future measurements on the same item may be computed as:

$$\bar{D}_\cdot \pm 2 \times \text{s.d.}(D_i)\sqrt{1+\frac{1}{n}}$$

Limits of agreement are usually computed using

$$\bar{D}_\cdot \pm 2 \times \text{s.d.}(D_i)$$

This can be converted to a prediction interval for a future measurement by method 2, y_{20} given a measurement by method y_{10}:

$$y_{2\cdot 1} = y_{10} + \bar{D}_\cdot \pm 2 \times \text{s.d.}(D_i)$$

3 Non-constant difference

3.1 Model

If it is observed that the assumption of constant difference between methods is violated, i.e. if there is clear slope in the Bland-Altman plot, we can regress the differences on the averages (or sums) and use the regression to construct prediction intervals of the difference between two future measurements.

The obvious extension of model (1) is a model where measurements by each of the methods are related linearly to the "true" mean, μ_i:

$$\begin{aligned} y_{1i} &= \alpha_1 + \beta_1\mu_i + e_{1i}, \qquad e_{1i} \sim \mathcal{N}(0,\sigma_1^2) \\ y_{2i} &= \alpha_2 + \beta_2\mu_i + e_{2i}, \qquad e_{2i} \sim \mathcal{N}(0,\sigma_2^2) \end{aligned} \tag{2}$$

In this model, prediction of y_2 from y_1 is derived by isolating μ_i from the first equation, because the observation of y_{10} on a new item is the only data we have for estimation of the corresponding unknown μ_0: $\hat{\mu}_0 = (y_{10}-\alpha_1)/\beta_1$. Inserting this in the second equation gives:

$$y_{2\cdot 1} = \alpha_2 + \beta_2\mu_0 + e_{20} = \left(\alpha_2 - \alpha_1\frac{\beta_2}{\beta_1}\right) + \frac{\beta_2}{\beta_1}y_{10} + \frac{\beta_2}{\beta_1}e_{10} + e_{20} \tag{3}$$

Hence, the parameters of interest are:

$$\begin{aligned} \text{Intercept:} \quad \alpha_{2\cdot 1} &= \alpha_2 - \alpha_1\frac{\beta_2}{\beta_1} \\ \text{Slope:} \quad \beta_{2\cdot 1} &= \frac{\beta_2}{\beta_1} \\ \text{Prediction variance:} \quad \sigma^2_{2\cdot 1} &= \left(\frac{\beta_2}{\beta_1}\right)^2\sigma_1^2 + \sigma_2^2 \end{aligned} \tag{4}$$

The 95% prediction limits for the value by method 2, y_{20}, given a measurement by method 1, y_{10}, on a new item are then

$$\alpha_{2\cdot 1} + \beta_{2\cdot 1}y_{10} \pm 2 \times \sigma_{2\cdot 1}$$

3.2 Regression of differences on averages

Now, consider the differences $D_i = y_{1i} - y_{2i}$ and the averages $A_i = (y_{1i} + y_{2i})/2$. Assuming the model (2) we have:

$$\begin{aligned} D_i &= (\alpha_1 - \alpha_2) + (\beta_1 - \beta_2)\mu_i + e_{1i} - e_{2i} \\ A_i &= (\alpha_1 + \alpha_2)/2 + (\beta_1 + \beta_2)\mu_i/2 + (e_{1i} + e_{2i})/2 \end{aligned}$$

The relationship between D_i and A_i can be expressed by isolating μ_i from the expression for A_i:

$$\mu_i = [2A_i - (\alpha_1 + \alpha_2) - (e_{1i} + e_{2i})]/(\beta_1 + \beta_2)$$

and inserting this in the expression for D_i:

$$\begin{aligned} D_i &= (\alpha_1 - \alpha_2) + \frac{\beta_1 - \beta_2}{\beta_1 + \beta_2}[2A_i - (\alpha_1 + \alpha_2) - (e_{1i} + e_{2i})] + (e_{1i} - e_{2i}) \\ &= (\alpha_1 - \alpha_2) - (\alpha_1 + \alpha_2)\frac{\beta_1 - \beta_2}{\beta_1 + \beta_2} \\ &\quad + \frac{\beta_1 - \beta_2}{\beta_1 + \beta_2}2A_i \\ &\quad + e_{1i}\left(1 - \frac{\beta_1 - \beta_2}{\beta_1 + \beta_2}\right) - e_{2i}\left(1 + \frac{\beta_1 - \beta_2}{\beta_1 + \beta_2}\right) \end{aligned}$$

The functions of the parameters that correspond to the intercept, the slope and the residual standard deviation in this relation between D_i and A_i can be estimated by a simple linear regression of the differences on the averages:

$$D_i = a + bA_i + e_i, \qquad e_i \sim \mathcal{N}(0, \tau^2) \tag{5}$$

Based on estimates of a, b and τ^2 from this model, the parameters of interest ($\alpha_{2\cdot1}$, $\beta_{2\cdot1}$ and $\sigma_{2\cdot1}$) may then be derived, since these are also functions of the parameters in model 2. The relationships are (using (4)):

$$\left.\begin{aligned} a &= (\alpha_1 - \alpha_2) - (\alpha_1 + \alpha_2)\frac{\beta_1 - \beta_2}{\beta_1 + \beta_2} \\ b &= 2\frac{\beta_1 - \beta_2}{\beta_1 + \beta_2} \\ \tau^2 &= \left(\frac{2\beta_1}{\beta_1 + \beta_2}\right)^2 \left(\frac{\beta_2^2}{\beta_1^2}\sigma_1^2 + \sigma_2^2\right) \end{aligned}\right\} \quad \Leftrightarrow \quad \left\{\begin{aligned} \alpha_{2\cdot1} &= \frac{-a}{1 + b/2} \\ \beta_{2\cdot1} &= \frac{1 - b/2}{1 + b/2} \\ \sigma_{2\cdot1} &= \frac{\tau}{(1 + b/2)} \end{aligned}\right. \tag{6}$$

The formulae for predicting method 1 from method 2 follows from symmetry; if we instead regressed the opposite differences $y_2 - y_1$ on the sums, we would just get opposite values of a and b; τ would remain unchanged.

So from the simple linear regression of the differences on the averages we can compute the parameters needed to provide a prediction equation from one method to another.

Bland and Altman [1] advocate the regression of the differences on the averages and use the regression line and the residual standard deviation from this (τ) directly in producing limits of agreement. This is a correct procedure, however largely irrelevant since it refers to prediction of the difference for a given average, and the average is rarely known unless both measurements are known.

The interesting concept is the prediction of one method from another, which as shown above is *not* obtained by plugging in the value of one method for the average, as also suggested by Bland and Altman [1]. They propose to plug in the observed value to obtain the limits of agreement for a given measurement by one method which is equivalent to constructing a prediction interval for the other method; in terms of the parameters above this would be:

$$y_{2\cdot1} = -a + (1-b)y_1 \pm 2\,\tau$$

This procedure proposed by Bland and Altman gives a smaller slope in the prediction of method 2 from method 1, than the model based approach:

$$\beta_{2\cdot1} = \frac{1-b/2}{1+b/2} = \frac{(1-b/2)^2}{(1+b/2)(1-b/2)} = \frac{1-b+(b/2)^2}{1-(b/2)^2} \geq 1-b$$

The last inequality is because subtraction of a positive constant from the numerator and addition of a positive constant to the denominator (in this case $(b/2)^2$) makes the fraction smaller.

The prediction standard deviation by the method proposed by Bland & Altman is τ, which is too small if the prediction slope is larger than 1 ($b < 0$), and too large if the prediction slope is smaller than 1.

Hence the formulae given in (6) are preferable to just plugging the measurement by method 1 in for the average in the regression of the differences on the averages.

4 A worked example

Table 1 shows data from measurements of plasma volume expressed as a percentage of normal in 99 subjects, using two alternative sets of normal values due to Nadler and Hurley. Data are adapted from the paper by Bland & Altman [1]

Regression the difference between method Hurley and Nadler on the average gives $D = -0.908 - 0.089$ with a residual standard deviation of 2.037 and hence a prediction interval of ± 3.993 (using 1.96 instead of 2). Based on this regression we get the prediction limits

$$\begin{aligned} y_{H\cdot N} &= \frac{a}{1-b} + \frac{1+b}{1-b}\,y_N \pm 2\frac{\tau}{1-b} \\ &= -0.870 + 0.915\,y_N \pm 3.823 \\ \\ y_{N\cdot H} &= \frac{-a}{1+b} + \frac{1-b}{1+b}\,y_H \pm 2\frac{\tau}{1+b} \\ &= 0.951 + 1.093\,y_H \pm 4.179 \end{aligned}$$

Table 1: *Measurements of plasma volume expressed as a percentage of normal in 99 subjects, using two alternative sets of normal values due to Nadler and Hurley. Adapted from table 2 in [1]*

	Nadler	Hurley		Nadler	Hurley		Nadler	Hurley
1	56.9	52.9	34	93.5	86.0	67	104.8	97.1
2	63.2	59.2	35	94.5	84.3	68	105.1	97.3
3	65.5	63.0	36	94.6	87.6	69	105.5	95.1
4	73.6	66.2	37	95.0	84.0	70	105.7	95.8
5	74.1	64.8	38	95.2	85.9	71	106.1	95.5
6	77.1	69.0	39	95.3	84.4	72	106.8	95.9
7	77.3	67.1	40	95.6	85.2	73	107.2	95.4
8	77.5	70.1	41	95.9	85.2	74	107.4	97.3
9	77.8	69.2	42	96.4	89.2	75	107.5	97.7
10	78.9	73.8	43	97.2	87.8	76	107.5	93.0
11	79.5	71.8	44	97.5	88.0	77	108.0	97.6
12	80.8	73.3	45	97.9	88.7	78	108.2	96.1
13	81.2	73.1	46	98.2	91.2	79	108.6	96.2
14	81.9	74.7	47	98.5	91.8	80	109.1	99.5
15	82.2	74.1	48	98.8	92.5	81	110.1	99.8
16	83.1	74.1	49	98.9	88.0	82	111.2	105.3
17	84.4	76.0	50	99.0	93.5	83	111.7	103.6
18	84.9	75.4	51	99.3	89.0	84	111.7	100.2
19	86.0	74.6	52	99.3	89.4	85	112.0	100.0
20	86.3	79.2	53	99.9	89.2	86	113.1	98.8
21	86.3	77.8	54	100.1	91.3	87	116.0	110.0
22	86.6	80.8	55	101.0	90.4	88	116.7	103.5
23	86.6	77.6	56	101.0	91.2	89	118.8	109.4
24	86.6	77.5	57	101.5	91.4	90	119.7	112.1
25	87.1	78.6	58	101.5	93.0	91	120.7	111.3
26	87.5	78.7	59	101.5	91.2	92	122.8	108.6
27	87.8	81.5	60	101.8	92.0	93	124.7	112.4
28	88.6	79.3	61	101.8	91.8	94	126.4	113.8
29	89.3	78.9	62	102.8	96.8	95	127.6	115.6
30	89.6	85.9	63	102.9	92.8	96	128.2	118.1
31	90.3	80.7	64	103.2	94.0	97	129.6	116.8
32	91.1	80.6	65	103.8	93.5	98	130.4	121.6
33	92.1	82.8	66	104.4	95.8	99	133.2	115.8

The prediction line and the prediction limits are shown in figure 2 in gray. The prediction equations and limits give identical lines in the, so the plot may be used for prediction both ways. It is clear that the major effect is on the size of the prediction interval; depending on whether the prediction is one way or the other, the interval will be too narrow or too wide. The advantage of using the proposed method is that a prediction interval which is applicable both ways is obtained.

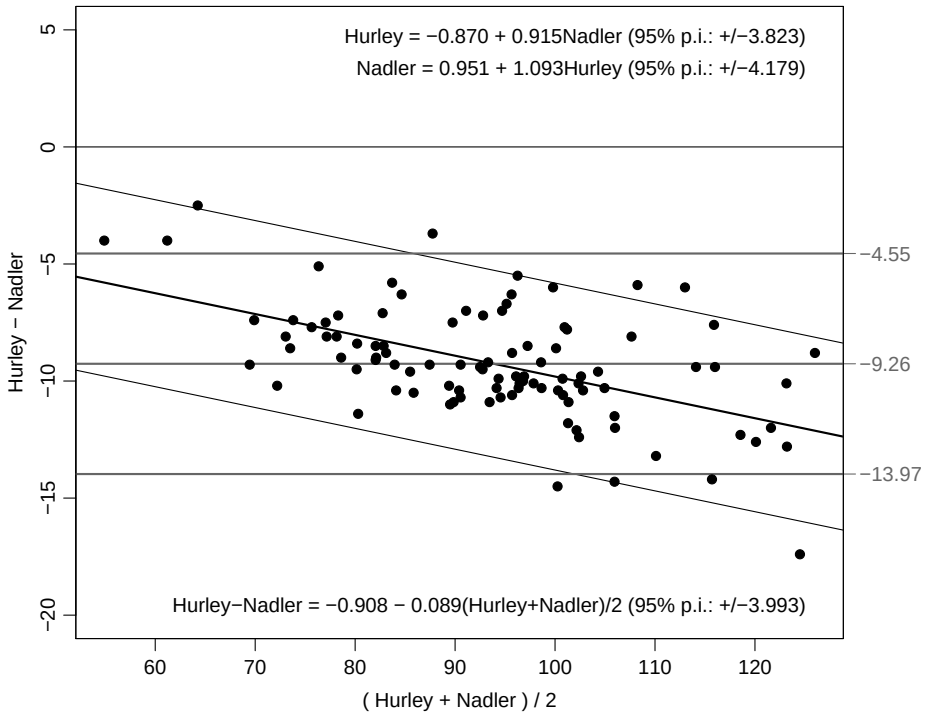

Figure 1: *Bland-Altman plot for the plasma volume data with a regression line of differences on averages. The prediction equations between the methods are the correctly derived ones. Standard output from the function* `BlandAltman` *from the R-package* `MethComp`.

5 Why it's wrong to use the regression of the differences on the averages

5.1 Substantially wrong

Model (2) is qualitatively different from (1) in the sense that it is invariant under arbitrary *scaling* of one of the methods. If for example all measurements by method 2 were multiplied by 2, it would just mean that α_2, β_2 and σ_2 would be twice as large. But the differences and averages would be irrelevant, we could get any result by rescaling measurements by one method. Therefore, the proposed procedure is *only* relevant in situations where the two methods compared can be assumed to be on the same scale.

In the special situation where the relationship between methods is multiplicative, we would log-transform all measurements to obtain a model of the form (2) for the transformed data. In this case, both methods would be on the same scale, namely the relative scale.

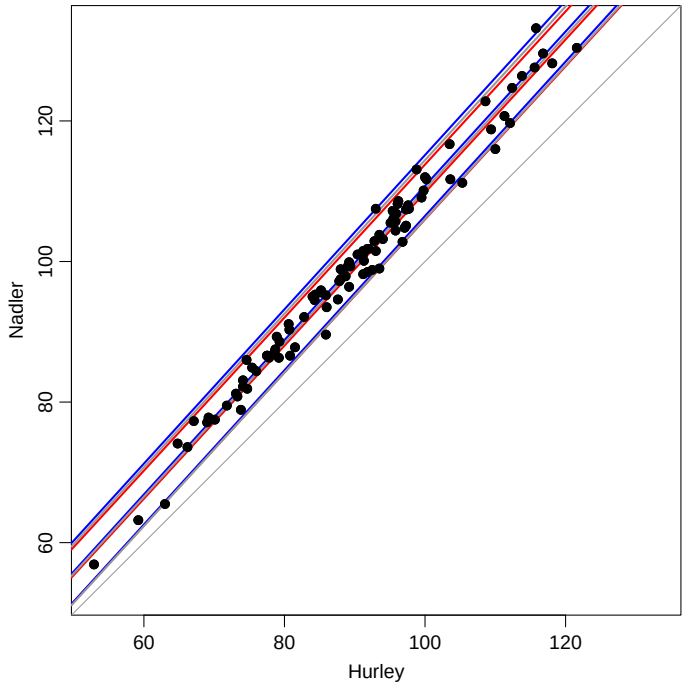

Figure 2: *Prediction limits between the two methods for the plasma volume data.*

5.2 Statistically wrong

In the regression of the differences on the averages we implicitly assume that the averages are independent of the error terms. But this is *not* the case:

$$\begin{aligned}
&\mathrm{cov}(A_i, e_{1i}\left(1-\frac{\beta_1-\beta_2}{\beta_1+\beta_2}\right)-e_{2i}\left(1+\frac{\beta_1-\beta_2}{\beta_1+\beta_2}\right)) \\
&= \mathrm{cov}(\frac{e_{1i}+e_{2_i}}{2}, e_{1i}\left(1-\frac{\beta_1-\beta_2}{\beta_1+\beta_2}\right)-e_{2i}\left(1+\frac{\beta_1-\beta_2}{\beta_1+\beta_2}\right)) \\
&= \frac{1}{2}\left\{\sigma_1^2-\sigma_2^2-\frac{\beta_1-\beta_2}{\beta_1+\beta_2}(\sigma_1^2+\sigma_2^2)\right\}
\end{aligned}$$

This covariance is 0 iff:

$$\frac{\sigma_1^2-\sigma_2^2}{\sigma_1^2+\sigma_2^2}=\frac{\beta_1-\beta_2}{\beta_1+\beta_2} \quad \Leftrightarrow \quad \frac{\beta_2}{\beta_1}=\frac{\sigma_2^2}{\sigma_1^2}$$

i.e. if the slope relating one method to the other equals the ratio of the variances.

Multiplying all measurements by one method by a factor K changes the slope by a factor K but the ratio of the variances by a factor K^2, so using the correct scaling of the methods relative to each other will make it possible to obtain uncorrelated sums and residuals. However, there is no data to estimate the scaling, which in any case would be an odd thing to do; estimation of the variance ratio should be done based on replicate measurements on each item by each method.

Hence, regressing differences on averages essentially assumes that the estimated slope between methods is also the ratio of residual variances of the methods.

5.3 Why are the limits straight lines?

The prediction limits are straight lines because the estimation variance of $\alpha_{2.1}$ and $\beta_{2.1}$ are ignored. In decently designed studies these are so small that it would not make much difference whether they were included or not. If the study is so small that it matters it is advisable to do a better designed study, or an indication that the chosen range of the measurements is too small.

5.4 What is the relation to standard regression?

The model (2) is *not* a classical regression model, because the individual levels (the μ_is) are explicit parameters in the model. Classical regression models are based on the conditional distribution of one method given that of another, derived from a bivariate normal distribution. In this case this would result in predictions implicitly assuming that the measurement known was taken on an individual drawn from *the same* population as the one used for the method comparison. This is almost always an erroneous assumptions, because samples used for method comparison studies normally will be deliberately chosen to span a relevant range of measurements, and therefore will *not* be representative of future items.

This is the main reason that classical regression analysis is not appropriate for making predictions based on method comparison studies.

5.5 What is the relation to Deming regression?

Deming regression [2, 3] provides the maximum-likelihood estimator in the model (2) (with the usual bias correction for the variance), under the assumption that the ratio of the variances is known.

The prediction variance for predicting method 2 from method 1 is computed as $\sigma_{2\cdot1} = \sqrt{\beta_2\sigma_1^2/\beta_1 + \sigma_2^2}$.

So the Deming regression does not solve the prediction problem unless we are willing to assume a known value for ratio of the variances. In studies without replicates there is no information about the ratio of the residual variances for the two methods.

6 How wrong is it to do it anyway?

To assess the possible discrepancy between the simple method based on regression of differences on sums and the method deriving results from model (2), and to see how well these reproduce the true values, we set up a small simulation study.

Clearly, the actual value of the intercept α is irrelevant for the outcome, only the slopes (i.e. their ratio) and the ratio of residual variances are of interest.

In 5 different scenarios we simulated 100 datasets with 50 items in each, the "true" values (μ_is) uniformly distributed on $[-50, 50]$.

For each dataset we computed estimates of $\alpha_{2\cdot1}$, $\beta_{2\cdot1}$ and $\sigma_{2\cdot1}$, as well as the coverage. The coverage was estimated as the fraction of points from 10 new simulated

datasets that fell within the 95% prediction limits based on the parameter estimates, i.e. for each simulated dataset used for estimation, we simulated 10 new only used for computing the coverage.

We did the calculation by three different methods:

- The simple approach from Bland and Altman [1].
- Deming regression [2, 3], assuming equal variances.
- The model based method proposed.

Within each of the figures 3–8 we used the same values of β_2 for all diagrams and 5 different values of $(\sigma_1, \sigma_2) = (3,1), (3,2), (3,3), (3,5), (3,10)$. The values of β varies between the figures. On the x-axis we have estimates by the method proposed, and on the y-axis we have the corresponding values by the two other methods.

The simulation shows that the proposed method is superior to the the simple approach, although far from unbiased when β is far from 1 or the ratio of residual variances is far from 1.

The simple method underestimates the slope when the true value is far from 1. There is a tendency that the simple method underestimates the prediction variances, whereas the proposed method tends to overestimate them.

The same is true for the coverage, which falls dramatically for the Deming regression and the simple methods as β and the variance ratio moves away from 1. In almost all instances the model based method outperforms the other two in terms of coverage when judging by the average coverage as indicated by the broken lines in the plots.

7 Discussion

I have here proposed a simple twist to the results from regression of the differences on the sums in the case of a linear relationship between two methods of measurement. It is consistent with the obvious underlying model, and exploits the fact that although the parameters of the model cannot be estimated, those functions of the parameters that are needed for creating predictions can be estimated.

The prediction limits provided have the attractive property that if the prediction line with limits is drawn in a coordinate system, the chart will apply both ways, so *both* the line *and* the limits are symmetric. Precisely as the prediction intervals derived from the classical limits of agreement are in the case where the difference between methods is constant.

The drawback is that regression of the differences on the means ignores that the averages are correlated with the residuals (i.e. the error terms), and therefore gives biased estimates if the slope is far from 1 or the residual variances are very different.

Bland and Altman [1] advocate the regression of the differences on the averages and use the regression line and the residual standard deviation from this directly in producing limits of agreement without any explicit model for the observations. These limits do not have the property that the prediction equation for method 1 to method

2 can be obtained by inverting that from method 2 to method 1, and are subject to the same drawbacks as the method proposed here.

8 Summary

When comparing two methods of measurement, and only one measurement by each is available, the following is recommended:

1. Compute limits of agreement. If relevant convert to a prediction interval for one method given the other.

2. Make a Bland-Altman plot to see if the basic assumptions are met.

3. If there is non-constant difference but uniform variance, regress the differences $y_1 - y_2$ on the averages $(y_1 + y_2)/2$, obtain intercept a, slope b and residual standard deviation τ, and construct the prediction between the methods as:

$$\begin{aligned} y_{1\cdot 2} &= \frac{a}{1-b/2} + \frac{1+b/2}{1-b/2}\, y_2 \pm 2\frac{\tau}{1-b/2} \\ y_{2\cdot 1} &= \frac{-a}{1+b/2} + \frac{1-b/2}{1+b/2}\, y_1 \pm 2\frac{\tau}{1+b/2} \end{aligned}$$

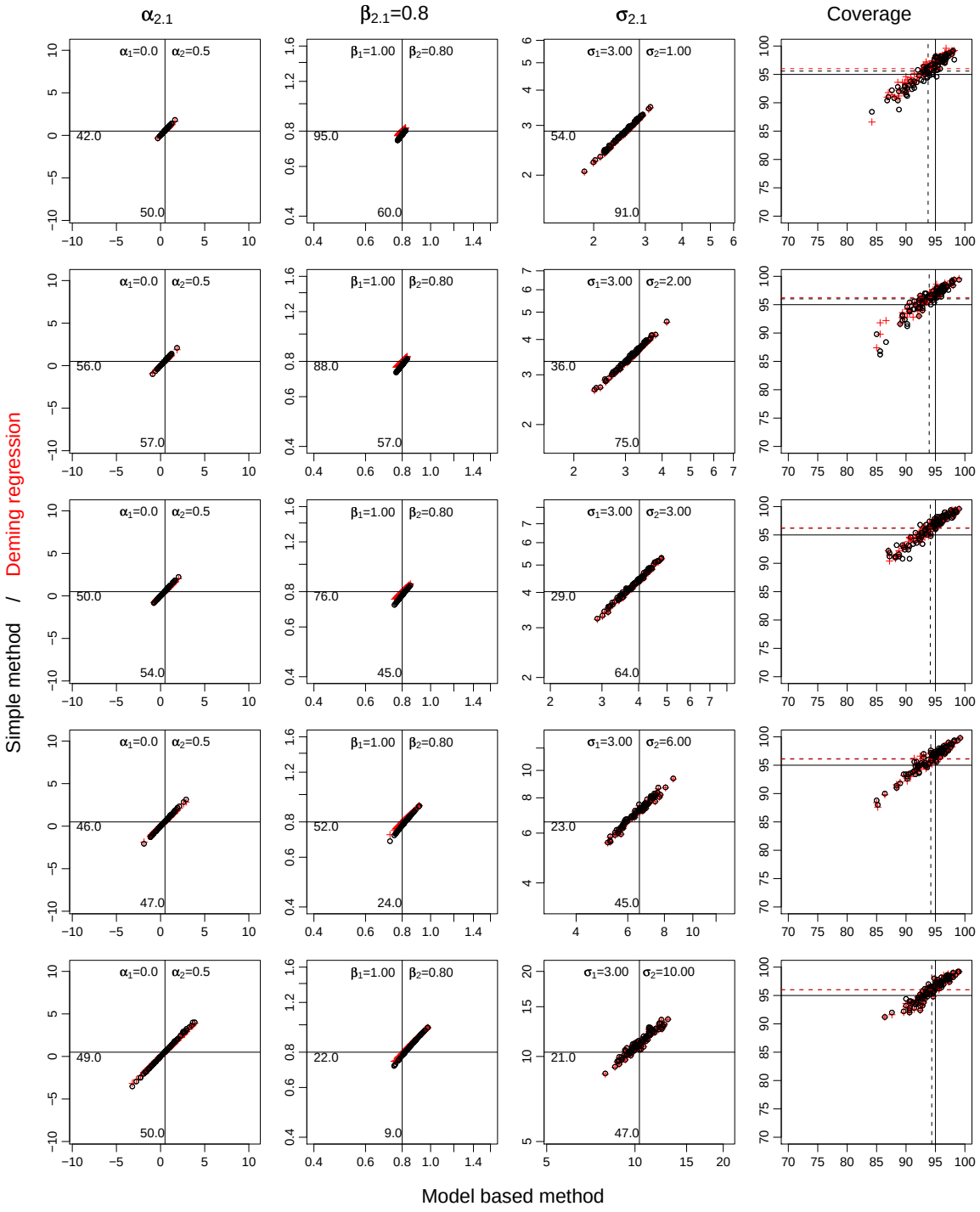

Figure 3: *Comparison of the results from the model based approach (x-axis) and the simple procedure and Deming regression (y-axis). The parameters compared are the intercept ($\alpha_{2\cdot1}$), the prediction slope ($\beta_{2\cdot1}$) and the prediction standard deviation ($\sigma_{2\cdot1}$). The lines indicate the true values used in the simulation. The small number indicates the percentage of simulated values that fall below the true value. The rightmost panels gives the coverage computed as the average fraction of points from ten independently simulated datasets that fall within the estimated prediction limits.*

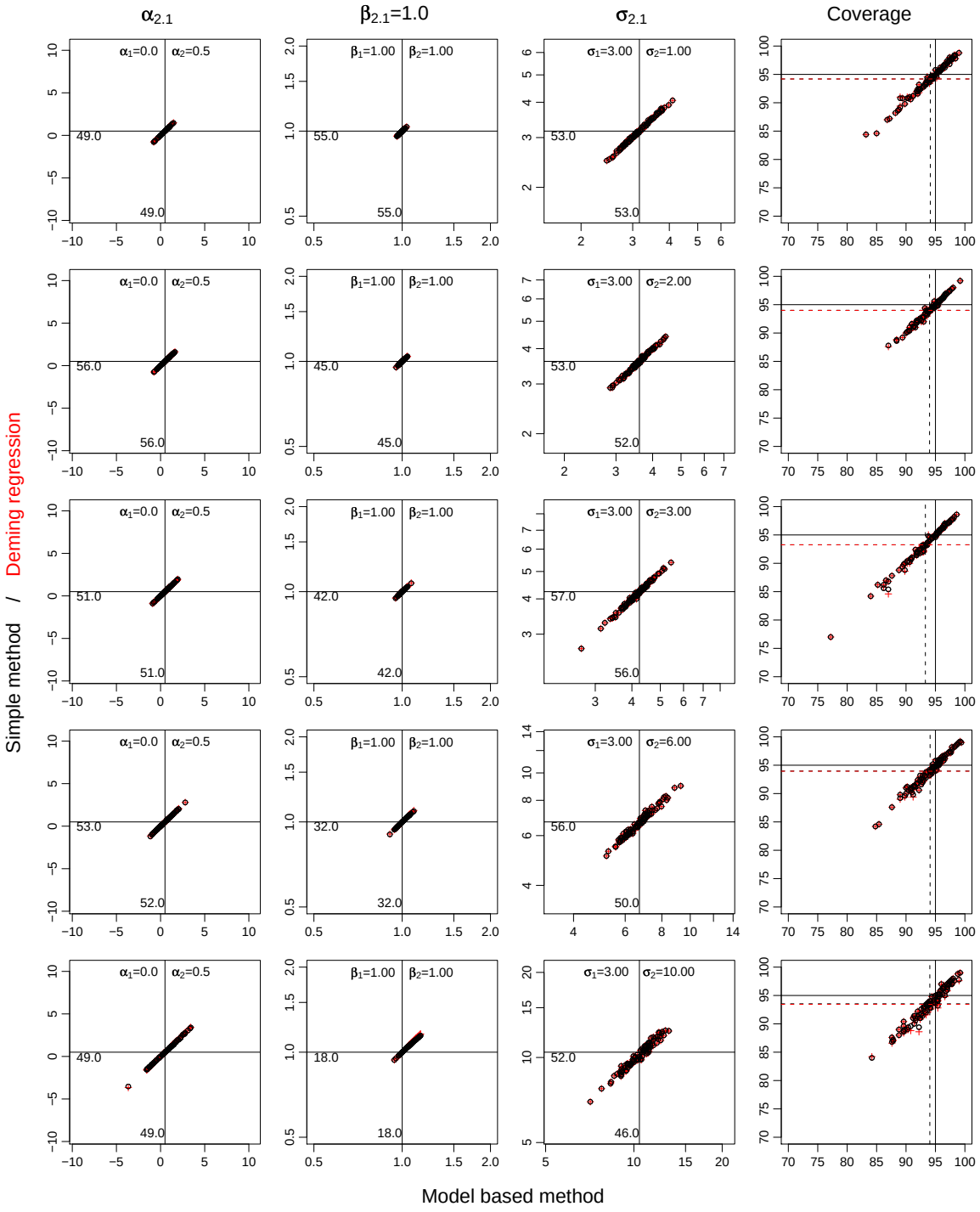

Figure 4: *Comparison of the results from the model based approach (x-axis) and the simple procedure and Deming regression (y-axis). The parameters compared are the intercept ($\alpha_{2\cdot1}$), the prediction slope ($\beta_{2\cdot1}$) and the prediction standard deviation ($\sigma_{2\cdot1}$). The lines indicate the true values used in the simulation. The small number indicates the percentage of simulated values that fall below the true value. The rightmost panels gives the coverage computed as the average fraction of points from ten independently simulated datasets that fall within the estimated prediction limits.*

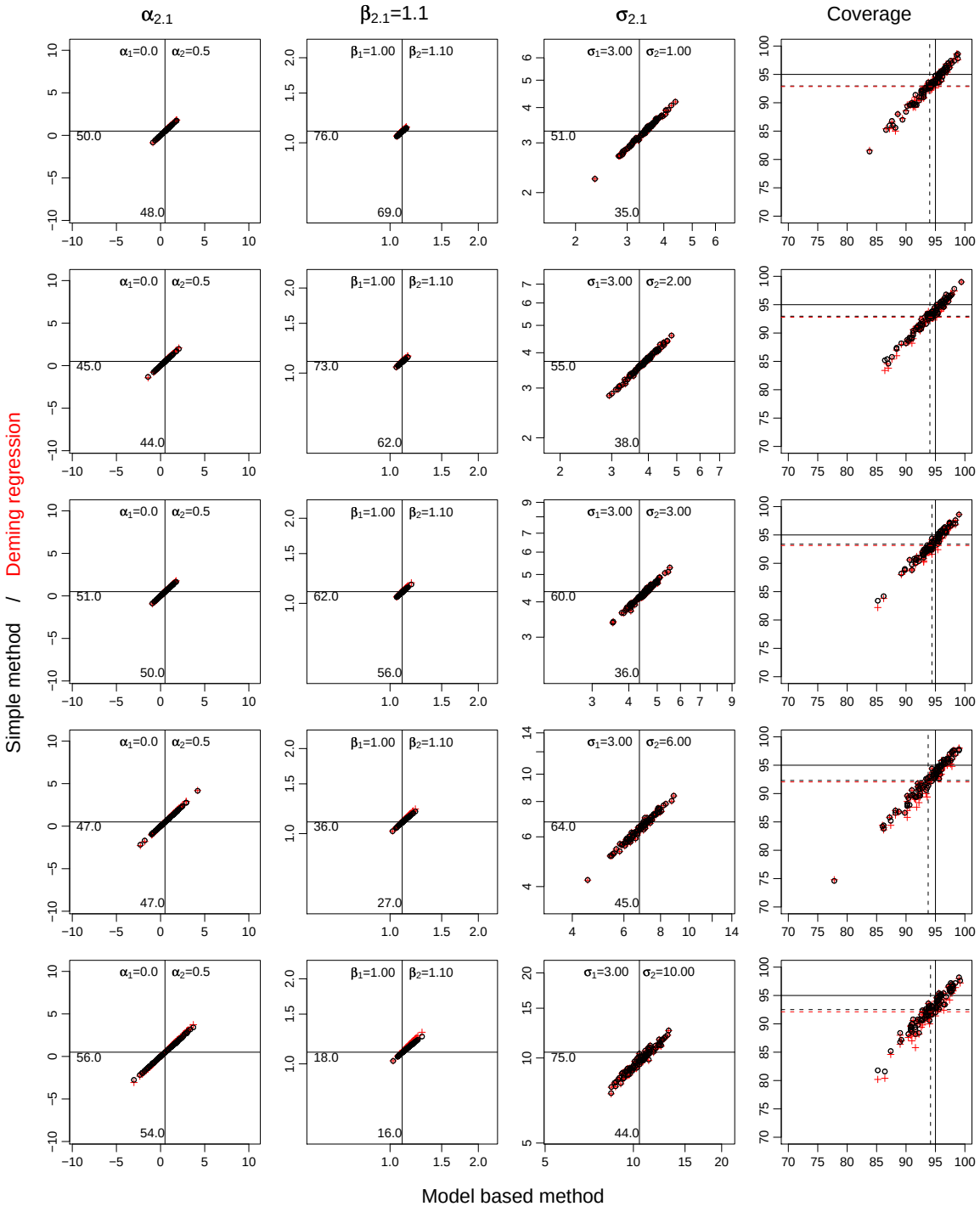

Figure 5: *Comparison of the results from the model based approach (x-axis) and the simple procedure and Deming regression (y-axis). The parameters compared are the intercept ($\alpha_{2\cdot1}$), the prediction slope ($\beta_{2\cdot1}$) and the prediction standard deviation ($\sigma_{2\cdot1}$). The lines indicate the true values used in the simulation. The small number indicates the percentage of simulated values that fall below the true value. The rightmost panels gives the coverage computed as the average fraction of points from ten independently simulated datasets that fall within the estimated prediction limits.*

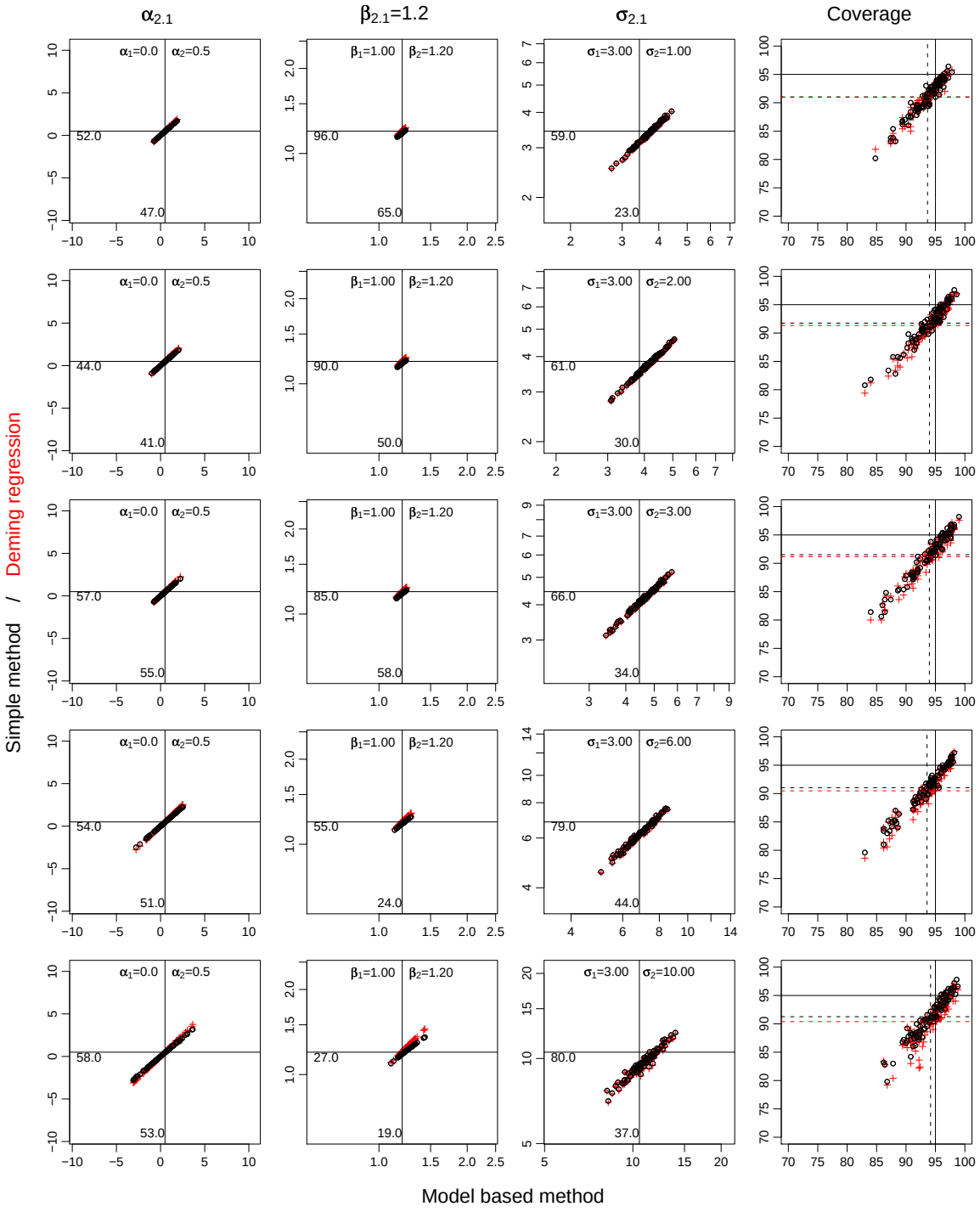

Figure 6: *Comparison of the results from the model based approach (x-axis) and the simple procedure and Deming regression (y-axis). The parameters compared are the intercept ($\alpha_{2\cdot1}$), the prediction slope ($\beta_{2\cdot1}$) and the prediction standard deviation ($\sigma_{2\cdot1}$). The lines indicate the true values used in the simulation. The small number indicates the percentage of simulated values that fall below the true value. The rightmost panels gives the coverage computed as the average fraction of points from ten independently simulated datasets that fall within the estimated prediction limits.*

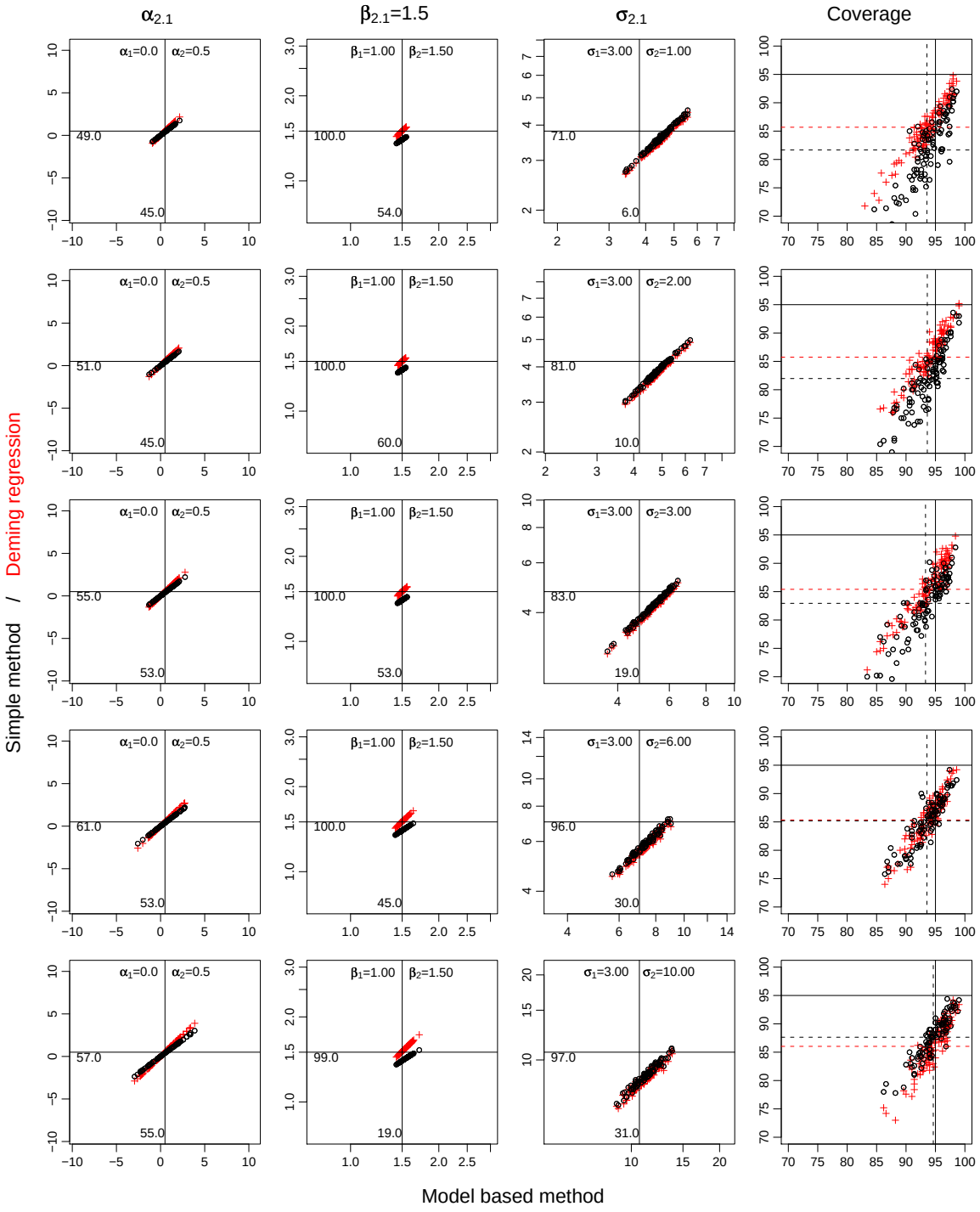

Figure 7: *Comparison of the results from the model based approach (x-axis) and the simple procedure and Deming regression (y-axis). The parameters compared are the intercept ($\alpha_{2\cdot1}$), the prediction slope ($\beta_{2\cdot1}$) and the prediction standard deviation ($\sigma_{2\cdot1}$). The lines indicate the true values used in the simulation. The small number indicates the percentage of simulated values that fall below the true value. The rightmost panels gives the coverage computed as the average fraction of points from ten independently simulated datasets that fall within the estimated prediction limits.*

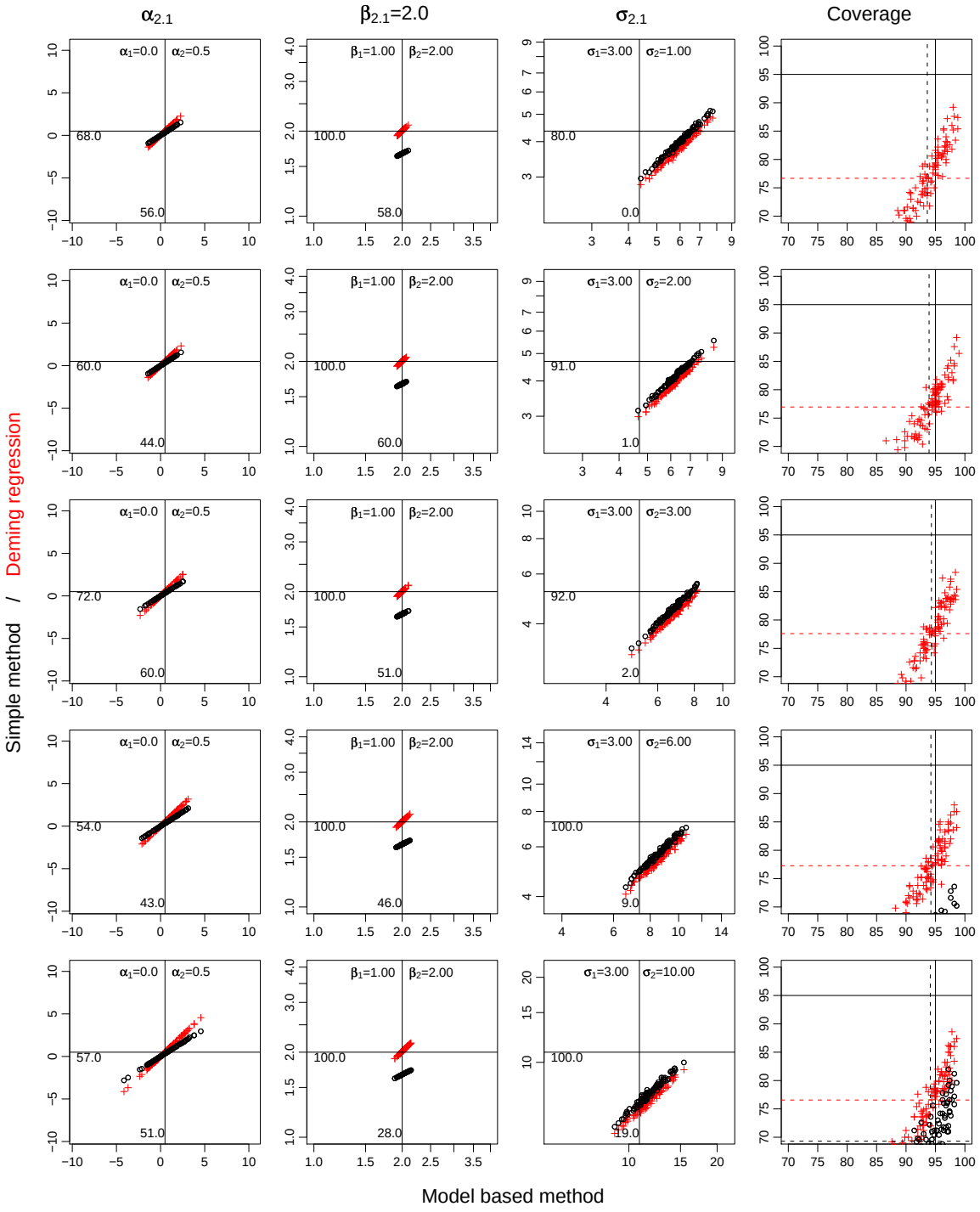

Figure 8: *Comparison of the results from the model based approach (x-axis) and the simple procedure and Deming regression (y-axis). The parameters compared are the intercept ($\alpha_{2\cdot1}$), the prediction slope ($\beta_{2\cdot1}$) and the prediction standard deviation ($\sigma_{2\cdot1}$). The lines indicate the true values used in the simulation. The small number indicates the percentage of simulated values that fall below the true value. The rightmost panels gives the coverage computed as the average fraction of points from ten independently simulated datasets that fall within the estimated prediction limits.*

9 Acknowledgments

I am grateful to Lene Theil Skovgaard from Department of Biostatistics, University of Copenhagen for helpful discussions and valuable suggestions.

References

[1] J.M. Bland and D.G. Altman. Measuring agreement in method comparison studies. *Statistical Methods in Medical Research*, 8:136–160, 1999.

[2] PJ Cornbleet and N Gochman. Incorrect least-squares regression coefficients in method-comparison analysis. *Clin. Chem.*, 25(3):432–438, 1973.

[3] AC Jensen. Deming regression. Technical report, http://staff.pubhealth.ku.dk/~bxc/MethComp/Deming.pdf, 2007. Vignette for the `MethComp` package for R.