Simulation of asymptotic variances

Hendrik Paul Lopuhaä

Delft University of Technology

February 23, 2022

Contents

1 Introduction

This document is intended to prepare for making a comparison of our theoretical results concerning the asymptotic variance of the S-estimator with the results that emerge from the simulation study. This means that we have to implement several constants that appear in the asymptotic variances. This document describes the R-functions that implement these constants and gives some background about specific choices for the parameters in these functions.

2 Moments

The constants appearing in the limiting variances are built from expectations of the form

$$\mathbb{E}_{\mathbf{0},\mathbf{I}_k}\|\mathbf{z}\|^m\{\|\mathbf{z}\|\leq c\},$$

where $m = 0, 1, 2, \ldots$, $\mathbf{z} \sim N_k(\mathbf{0}, \mathbf{I}_k)$ and c is the cut-off constant in the rho-function. Note that $W_k = \|\mathbf{z}\|^2 \sim \chi^2_k$, so that the moments can be computed by means of standard R-functions for the gamma distribution. We have

$$\begin{aligned}
\mathbb{E}_{\mathbf{0},\mathbf{I}_k}\|\mathbf{z}\|^m\{\|\mathbf{z}\|\leq c\} &= \mathbb{E}W_k^{m/2}\{W_k\leq c^2\}\\
&= \int_0^{c^2} w^{m/2}\frac{1}{2^{k/2}\Gamma(k/2)}w^{k/2-1}\mathrm{e}^{-w/2}\,\mathrm{d}w\\
&= \frac{1}{2^{k/2}\Gamma(k/2)}\int_0^{c^2} w^{(m+k)/2-1}\mathrm{e}^{-w/2}\,\mathrm{d}w\\
&= \frac{1}{2^{k/2}\Gamma(k/2)}\int_0^{c^2/2} (2t)^{(m+k)/2-1}\mathrm{e}^{-t}\,\mathrm{d}2t \qquad \text{(change of variables } t = w/2\text{)}\\
&= \frac{2^{(m+k)/2}}{2^{k/2}\Gamma(k/2)}\int_0^{c^2/2} t^{(m+k)/2-1}\mathrm{e}^{-t}\,\mathrm{d}t\\
&= \frac{2^{m/2}\Gamma((m+k)/2)}{\Gamma(k/2)}\int_0^{c^2/2} \frac{t^{(m+k)/2-1}\mathrm{e}^{-t}}{\Gamma((m+k)/2)}\,\mathrm{d}t\\
&= \frac{2^{m/2}\Gamma((m+k)/2)}{\Gamma(k/2)}\texttt{pgamma((c*c)/2,shape=(m+k)/2,scale=1)}
\end{aligned}$$

This is implemented by the following R-function

```
moment=function(m,k,c0){
  # computes the constrained expectation E|Z|^m{|z|<=c}
  # where Z has a multivariate N(0,I_k) distribution
  # m=the power of the moment
  # k=dimension of the normal distribution
  # c0=cut-off value of the rho-function
  const=2^(m/2)*gamma((m+k)/2)/gamma(k/2)
  gamint=pgamma((c0^2)/2,shape=(m+k)/2,scale=1)
  return(const*gamint)
}
```

3 The rho-function

The S-estimator is defined by means of a function $\rho(y)$ with cut-off constant $c > 0$, such that ρ is constant for $|y| \geq c$, symmetric on $[-c, c]$, and non-decreasing on $[0, c]$. A example that is often used is Tukey';s bi-weight function defined in the next subsection.

3.1 Tukey's biweight rho function

The rho-function that appears in most of the literature on S-estimators is Tukey's biweight rho-function (e.g., see Rousseeuw and Yohai [11], Lopuhaä [7], Van Aelst and Willems [15]). It is defined as

$$\rho(y) = \rho(y; c) = \begin{cases} \dfrac{y^2}{2} - \dfrac{y^4}{2c^2} + \dfrac{y^6}{6c^4}, & |y| \leq c \\ \dfrac{c^2}{6} & , |y| \geq c. \end{cases}$$

For later convenience, we write this as

$$\rho(y) = \begin{cases} a_1 y^2 + a_2 y^4 + a_3 y^6 & , |y| \leq c \\ a_4 & , |y| \geq c, \end{cases}$$

where

$$\begin{aligned} a_1 &= 1/2 \\ a_2 &= -1/(2c^2) \\ a_3 &= 1/(6c^4) \\ a_4 &= c^2/6 \end{aligned} \tag{3.1}$$

It is obtained as the continuous primitive of Tukeys's biweight psi-function, defined by

$$\begin{aligned} \psi(y) = \rho'(y) &= \begin{cases} y\left(1 - \dfrac{y^2}{c^2}\right)^2 & , |y| \leq c \\ 0 & , |y| \geq c \end{cases} \\ &= \begin{cases} y - \dfrac{2y^3}{c^2} + \dfrac{y^5}{c^4} & , |y| \leq c \\ 0 & , |y| \geq c. \end{cases} \end{aligned}$$

Finally, the derivative of ψ is also important in computing the constants. It is given by

$$\psi'(y) = \rho''(y) = \begin{cases} 1 - \dfrac{6y^2}{c^2} + \dfrac{5y^4}{c^4} & , |y| \leq c \\ 0 & , |y| \geq c. \end{cases}$$

3.2 Choosing the cut-off constant of Tukey's biweight

The breakdown point of the S-estimator is given by

$$\epsilon_n^*(\boldsymbol{\beta}_n, \mathcal{S}_n) = \epsilon_n^*(\boldsymbol{\theta}_n, \mathcal{S}_n) = \frac{\lceil nr \rceil}{n},$$

where

$$r = \frac{b_0}{\rho(c)},$$

with b_0 being the constant in the S-constrained. From the S-constraint it can be seen that a natural choice is

$$b_0 = \mathbb{E}_{\mathbf{0},\mathbf{I}_k}\rho(\|\mathbf{z}\|).$$

Hence, in order to have a given breakdown point r (for large sample sizes), we choose the cut-off value c, such that

$$\frac{\mathbb{E}_{\mathbf{0},\mathbf{I}_k}\rho(\|\mathbf{z}\|; c)}{\rho(c)} = r.$$

For Tukey's biweight this is the same as solving

$$\frac{\mathbb{E}\|\mathbf{z}\|^2\{\|\mathbf{z}\| \le c\}}{2} - \frac{\mathbb{E}\|\mathbf{z}\|^4\{\|\mathbf{z}\| \le c\}}{2c^2} + \frac{\mathbb{E}\|\mathbf{z}\|^6\{\|\mathbf{z}\| \le c\}}{6c^4} + \frac{c^2}{6}\mathbb{E}\{\|\mathbf{z}\| > c\} - r\frac{c^2}{6} = 0.$$

or equivalently

$$\frac{3\mathbb{E}\|\mathbf{z}\|^2\{\|\mathbf{z}\| \le c\}}{c^2} - \frac{3\mathbb{E}\|\mathbf{z}\|^4\{\|\mathbf{z}\| \le c\}}{c^4} + \frac{\mathbb{E}\|\mathbf{z}\|^6\{\|\mathbf{z}\| \le c\}}{c^6} + \mathbb{E}\{\|\mathbf{z}\| > c\} - r = 0$$

where $\|\mathbf{z}\|^2 \sim \chi^2_k$. In order to solve c from this equation, for given k and r we introduce the following R-function, which computes $\mathbb{E}_{\mathbf{0},\mathbf{I}_k}\rho(\|\mathbf{z}\|)$ as a function of c:

```
expecrho=function(k,c0){
  # computes E[rho(|Z|)]
  # k=dimension multivariate normal
  a1=1/2
  a2=-1/(2*(c0^2))
  a3=1/(6*(c0^4))
  a4=c0^2/6
  return(a1*moment(2,k,c0)+a2*moment(4,k,c0)+a3*moment(6,k,c0)+
            4*(1-moment(0,k,c0)))
}
```

We can then find the cut-off value by means of the following R-function, which applies the R-function `uniroot` on an interval $[a, b]$, for which `expecrho(a)` and `expecrho(b)` have opposite signs:

```
rhoconst=function(k,r,a,b){
  # computes the cut=off constant c0 for which
  # E[rho(|Z|)]=r*rho(c0),
  # where rho is the biweight
  # k=dimension of the multivariate normal
  # r=breakdown point
  # a,b are the boundaries of the interval over which we use uniroot
  objecfun=function(c0){
    expecrho(k,c0)-r*biweightrho(c0,c0)
  }
  return(uniroot(f=objecfun,c(a,b))$root)
}
```

For example, for dimensions $k = 1, 2, 5, 10$, we find

r	$k=1$	$k=2$	$k=5$	$k=10$
0.1	5.182	7.474	11.950	16.961
0.2	3.421	5.069	8.220	11.719
0.3	2.561	3.938	6.505	9.324
0.4	1.988	3.209	5.432	7.840
0.5	1.548	2.661	4.652	6.776

For $k = 1$, the values coincide with the ones in Table 3 in Rousseeuw and Yohai, 1984 [11]. Some of these values are of used to compare the constants in the asymptotic covariances with the results in Lopuhaä (1989) [7].

3.3 Asymptotic covariance of the regression S-estimator

When the data are generated from the distribution P of $(\mathbf{y}, \mathbf{X})$, which is such that $\mathbf{y} \mid \mathbf{X} \sim N(\mathbf{X}\boldsymbol{\beta}, \boldsymbol{\Sigma})$, then according to our Corollary 5, the asymptotic covariance of the regression S-estimator is given by

$$\frac{\mathbb{E}_{\mathbf{0},\mathbf{I}_k}\left[\rho'(\|\mathbf{z}\|)^2\right]}{k\alpha^2}\left(\mathbb{E}\left[\mathbf{X}^T\boldsymbol{\Sigma}^{-1}\mathbf{X}\right]\right)^{-1}$$

where

$$\alpha = \mathbb{E}_{\mathbf{0},\mathbf{I}_k}\left[\left(1-\frac{1}{k}\right)\frac{\rho'(\|\mathbf{z}\|)}{\|\mathbf{z}\|} + \frac{1}{k}\rho''(\|\mathbf{z}\|)\right].$$

Let us start with the constant α. First note that from Section 3.1, it follows that

$$\frac{\rho'(y)}{y} = \begin{cases} 1 - \dfrac{2y^2}{c^2} + \dfrac{y^4}{c^4} & , |y| \le c \\ 0 & , |y| \ge c. \end{cases}$$

and recall that

$$\rho''(y) = \begin{cases} 1 - \dfrac{6y^2}{c^2} + \dfrac{5y^4}{c^4} & , |y| \le c \\ 0 & , |y| \ge c. \end{cases}$$

Taking expecations this leads for the following R-function.

```
alpha=function(k,c0){
  # computes the value of the constant alpha in (8.4)
  # k = dimension of the normal distribution
  # c0 = cut-off value of the rho-function

  # Compute E[rho'(|Z|)/|Z|]
  term1=moment(0,k,c0)-2*moment(2,k,c0)/(c0^2)+moment(4,k,c0)/(c0^4)

  # Compute E[rho''(|Z|)]
  term2=moment(0,k,c0)-6*moment(2,k,c0)/(c0^2)+5*moment(4,k,c0)/(c0^4)

  return((1-1/k)*term1+(1/k)*term2)
}
```

For $\mathbb{E}_{\mathbf{0},\mathbf{I}_k}\left[\rho'(\|\mathbf{z}\|)^2\right]$, note that from Section 3.1, we have

$$\rho'(y)^2=\psi^2(y)=\begin{cases} y^2\left(1-\dfrac{y^2}{c^2}\right)^4 & ,|y|\leq c\\ 0 & ,|y|\geq c\end{cases}$$
$$=\begin{cases} y^2\left(1-\dfrac{4y^2}{c^2}+\dfrac{6y^4}{c^4}-\dfrac{4y^6}{c^6}+\dfrac{y^8}{c^8}\right) & ,|y|\leq c\\ 0 & ,|y|\geq c\end{cases},$$
$$=\begin{cases} y^2-\dfrac{4y^4}{c^2}+\dfrac{6y^6}{c^4}-\dfrac{4y^8}{c^6}+\dfrac{y^{10}}{c^8} & ,|y|\leq c\\ 0 & ,|y|\geq c\end{cases}$$

Computation of $\mathbb{E}_{\mathbf{0},\mathbf{I}_k}\left[\rho'(\|\mathbf{z}\|)^2\right]$ is then done with the R-function

```
psisquared=function(k,c0){
  # computes the numerator E[rho'(|Z|)^2] appearing in the
  # limiting covariance of betahat
  # see Corollary 5
  expectation=moment(2,k,c0)-4*moment(4,k,c0)/(c0^2)+
    6*moment(6,k,c0)/(c0^4)-4*moment(8,k,c0)/(c0^6)+
    moment(10,k,c0)/(c0^8)
  return(expectation)
}
```

Both R-functions can be combined to compute the constant

$$\frac{\mathbb{E}_{\mathbf{0},\mathbf{I}_k}\left[\rho'(\|\mathbf{z}\|)^2\right]}{k\alpha^2}$$

by means of the following R-function.

```
constbetahat=function(k,c0){
  # determines the scalar in the limiting covariance of betahat
  # k=dimension of multivariate normal
  # c0=pre-determined cut-off value for given BDP r
  return(psisquared(k,c0)/(k*alpha(k,c0)^2))
}
```

Note that the constant

$$\frac{\mathbb{E}_{\mathbf{0},\mathbf{I}_k}\left[\rho'(\|\mathbf{z}\|)^2\right]}{k\alpha^2}$$

is the same as the scalar λ in the limiting variance $\lambda\boldsymbol{\Sigma}$ of the location S-estimator in the multivariate location-scale model treated in Lopuhaä, 1989 [7]. This means we can compare the value of our R-function with the numbers in Table 1 in [7] to check whether the code is correct. We find the following values

r	$k=1$	$k=2$	$k=10$
0.1	1.035	1.011	1.001
0.2	1.181	1.055	1.006
0.3	1.512	1.157	1.016
0.4	2.165	1.356	1.036
0.5	3.486	1.725	1.072

These values coincide with the ones found in Lopuhaä, 1989 [7]. For $k=1$ the values also coincide with the values in Table 3 in Rousseeuw and Yohai, 1984 [11]. For $k=1,2,3,5,10,20,30$ and $r=0.25,0.50$, the values computed from our R-functions also coincide with the values in Table 3.1 in Van Aelst and Willems, 2005 [15].

3.4 Asymptotic covariance of the covariance S-estimator

When the data are generated from the distribution P of $(\mathbf{y},\mathbf{X})$, which is such that $\mathbf{y}\mid\mathbf{X}\sim N(\mathbf{X}\boldsymbol{\beta},\boldsymbol{\Sigma})$, where $\boldsymbol{\Sigma}=\mathbf{V}(\boldsymbol{\theta})=\sum_{j=1}^{l}\theta_j\mathbf{L}_j$, then according to Corollary 5, the asymptotic covariance of $\sqrt{n}(\boldsymbol{\theta}_n-\boldsymbol{\theta}(P))$ is given by

$$2\sigma_1\Big(\mathbf{L}^T\left(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\right)\mathbf{L}\Big)^{-1}+\sigma_2\boldsymbol{\theta}(P)\boldsymbol{\theta}(P)^T,$$

and $\sqrt{n}(\mathrm{vec}(\mathbf{C}_n)-\mathrm{vec}(\boldsymbol{\Sigma}))$ is asymptotically normal with mean zero and covariance matrix

$$2\sigma_1\mathbf{L}\Big(\mathbf{L}^T\left(\boldsymbol{\Sigma}^{-1}\otimes\boldsymbol{\Sigma}^{-1}\right)\mathbf{L}\Big)^{-1}\mathbf{L}^T+\sigma_2\mathrm{vec}(\boldsymbol{\Sigma})\mathrm{vec}(\boldsymbol{\Sigma})^T,$$

where

$$\begin{aligned}
\sigma_1&=\frac{k(k+2)\mathbb{E}_{\mathbf{0},\mathbf{I}_k}\left[u(\|\mathbf{z}\|)^2\|\mathbf{z}\|^4\right]}{\Big(\mathbb{E}_{0,\mathbf{I}_k}\left[\rho''(\|\mathbf{z}\|)\|\mathbf{z}\|^2+(k+1)\rho'(\|\mathbf{z}\|)\|\mathbf{z}\|\right]\Big)^2}\\
\sigma_2&=-\frac{2}{k}\sigma_1+\frac{4\mathbb{E}_{\mathbf{0},\mathbf{I}_k}\left[(\rho(\|\mathbf{z}\|)-b_0)^2\right]}{\Big(\mathbb{E}_{\mathbf{0},\mathbf{I}_k}\left[\rho'(\|\mathbf{z}\|)\|\mathbf{z}\|\right]\Big)^2}
\end{aligned}$$

Let us start with σ_1. We have from Section 3.1 that

$$\begin{aligned}
u(y)^2y^4=\psi^2(y)y^2&=\begin{cases}y^4\left(1-\dfrac{y^2}{c^2}\right)^4 & ,|y|\le c\\ 0 & ,|y|\ge c\end{cases}\\
&=\begin{cases}y^4\left(1-\dfrac{4y^2}{c^2}+\dfrac{6y^4}{c^4}-\dfrac{4y^6}{c^6}+\dfrac{y^8}{c^8}\right) & ,|y|\le c\\ 0 & ,|y|\ge c\end{cases},\\
&=\begin{cases}y^4-\dfrac{4y^6}{c^2}+\dfrac{6y^8}{c^4}-\dfrac{4y^{10}}{c^6}+\dfrac{y^{12}}{c^8} & ,|y|\le c\\ 0 & ,|y|\ge c\end{cases}
\end{aligned}$$

and

$$\rho''(y)y^2=\begin{cases}y^2-\dfrac{6y^4}{c^2}+\dfrac{5y^6}{c^4} & ,|y|\le c\\ 0 & ,|y|\ge c.\end{cases}$$

whereas

$$\begin{aligned}
\rho'(y)y=\psi(y)y&=\begin{cases}y^2\left(1-\dfrac{y^2}{c^2}\right)^2 & ,|y|\le c\\ 0 & ,|y|\ge c\end{cases}\\
&=\begin{cases}y^2-\dfrac{2y^4}{c^2}+\dfrac{y^6}{c^4} & ,|y|\le c\\ 0 & ,|y|\ge c.\end{cases}
\end{aligned}$$

This leads to the following R-function

```
sigma1=function(k,c0){
  # determines the constant sigma1 in the scalar in the
  # limiting covariance of thetahat
```

```
  # k=dimension of multivariate normal
  # c0=pre-determined cut-off value for given BDP r

  # Compute E[u(d)^2d^4]
  term1=moment(4,k,c0)-
    4*moment(6,k,c0)/(c0^2)+
    6*moment(8,k,c0)/(c0^4)-
    4*moment(10,k,c0)/(c0^6)+
    moment(12,k,c0)/(c0^8)

  # Compute E[rho''(d)d^2]
  term2=moment(2,k,c0)-
    6*moment(4,k,c0)/(c0^2)+
    5*moment(6,k,c0)/(c0^4)

  # Compute E[rho'(d)d]
  term3=moment(2,k,c0)-
    2*moment(4,k,c0)/(c0^2)+
    moment(6,k,c0)/(c0^4)

  num=k*(k+2)*term1
  denum=(term2+(k+1)*term3)^2
  return(num/denum)
}
```

We continue with

$$\sigma_2 = -\frac{2}{k}\sigma_1 + \frac{4\mathbb{E}_{\mathbf{0},\mathbf{I}_k}[(\rho(\|\mathbf{z}\|) - b_0)^2]}{\Big(\mathbb{E}_{\mathbf{0},\mathbf{I}_k}\left[\rho'(\|\mathbf{z}\|)\|\mathbf{z}\|\right]\Big)^2}.$$

Note that we have chosen

$$b_0 = \mathbb{E}_{\mathbf{0},\mathbf{I}_k}\rho(\|\mathbf{z}\|).$$

This means that

$$\mathbb{E}_{\mathbf{0},\mathbf{I}_k}[(\rho(\|\mathbf{z}\|) - b_0)^2] = \mathbb{E}_{\mathbf{0},\mathbf{I}_k}[\rho(\|\mathbf{z}\|)^2] - \left(\mathbb{E}_{\mathbf{0},\mathbf{I}_k}\rho(\|\mathbf{z}\|)\right)^2.$$

We have

$$\rho(y)^2 = \begin{cases} a_1^2y^4 + a_2^2y^8 + a_3^2y^{12} + 2a_1a_2y^6 + 2a_1a_3y^8 + 2a_2a_3y^{10} & , |y| \leq c \\ a_4^2 & , |y| \geq c. \end{cases}$$

where $a_1, \ldots, a_4$ are from (3.1). To implement the scalar σ_2, we use the following R-function.

```
sigma2=function(k,c0){
  # determines the constant sigma2 in the scalar in the
  # limiting covariance of thetahat
  # k=dimension of multivariate normal
  # c0=pre-determined cut-off value for given BDP r
  a1=1/2
  a2=-1/(2*(c0^2))
  a3=1/(6*(c0^4))
  a4=c0^2/6

  # Compute E[rho(d)^2]
  term1=(a1^2)*moment(4,k,c0)+(a2^2)*moment(8,k,c0)+(a3^2)*moment(12,k,c0)+
    2*a1*a2*moment(6,k,c0)+2*a1*a3*moment(8,k,c0)+2*a2*a3*moment(10,k,c0)+
```

```
    a4^2*(1-moment(0,k,c0))

  # Compute E[rho'(d)d]
  term2=moment(2,k,c0)-
    2*moment(4,k,c0)/(c0^2)+
    moment(6,k,c0)/(c0^4)

  # Compute E[rho(d)]
  term3=expecrho(k,c0)

  num=4*(term1-(term3)^2)
  denum=(term2)^2
  return(-(2/k)*sigma1(k,c0)+num/denum)
}
```

Both functions can be combined in Tyler's efficiency scalar η. As mentioned in [7], $\eta = 2\sigma_1 + \sigma_2$, when $k = 1$ and $\eta = \sigma_1$, when $k \geq 2$.

```
eta=function(k,c0){
  # computes the efficiency index of Tyler
  ifelse(k==1,1,0)*(2*sigma1(k,c0)+sigma2(k,c0))+
    ifelse(k>1,1,0)*sigma1(k,c0)
}
```

In order to check the code, we can compare the values of our R-functions with the efficiency index η in Table 1 in Lopuhaä, 1989 [7]. We find the following values with our code:

r	$k=1$	$k=2$	$k=10$
0.1	2.035	1.018	1.001
0.2	2.176	1.096	1.007
0.3	2.467	1.299	1.020
0.4	2.949	1.735	1.045
0.5	3.711	2.656	1.093

Again, these values coincide with the ones found in Lopuhaä, 1989 [7].

3.5 Computation of the S-estimator

Consider the case in which the covariance $\mathbf{V}$ is linear, i.e.,

$$\mathbf{V} = \sum_{j=1}^{l} \theta_j \mathbf{L}_j.$$

If a solution to the S-minimization exists, then under certain conditions the S-estimator is a solution to the score equations

$$\sum_{i=1}^{n} u(d_i)\mathbf{X}_i^T \mathbf{V}(\boldsymbol{\theta})^{-1}(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta}) = \mathbf{0}$$

and

$$\operatorname{tr}\left(\mathbf{V}^{-1}\mathbf{L}_j\right)\sum_{i=1}^{n} v(d_i) - k\sum_{i=1}^{n} u(d_i)(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta})^T\mathbf{V}^{-1}\mathbf{L}_j\mathbf{V}^{-1}(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta}) = 0, \quad \text{for } j = 1, \ldots, l,$$

where $u(s) = \rho'(s)/s$, $v(s) = u(s)s^2 - \rho(s) + b_0$, and $d_i^2 = (\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta})^T\mathbf{V}^{-1}(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta})$. The first equation leads to a fixed point vector equation for the regression parameter

$$\boldsymbol{\beta} = \left(\sum_{i=1}^n u(d_i)\mathbf{X}_i^T\mathbf{V}(\boldsymbol{\theta})^{-1}\mathbf{X}_i\right)^{-1}\sum_{i=1}^n u(d_i)\mathbf{X}_i^T\mathbf{V}(\boldsymbol{\theta})^{-1}\mathbf{y}_i.$$

The second system of equations can be turned into a second fixed point vector equation for the covariance parameter. First note that we can write

$$\mathrm{tr}\left(\mathbf{V}^{-1}\frac{\partial\mathbf{V}}{\partial\theta_j}\right) = \sum_{s=1}^l \theta_s \mathrm{tr}\left(\mathbf{V}^{-1}\mathbf{L}_j\mathbf{V}^{-1}\mathbf{L}_s\right) = (\mathbf{Q}\boldsymbol{\theta})_j$$

with

$$\mathbf{Q} = \left[\mathrm{tr}\left(\mathbf{V}^{-1}\mathbf{L}_j\mathbf{V}^{-1}\mathbf{L}_s\right)\right]_{j,s=1,\dots,l}.$$

Furthermore, for each $j = 1,\dots,l$, define

$$\mathbf{U}_j = \frac{k}{n}\sum_{i=1}^n u(d_i)(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta})^T\mathbf{V}^{-1}\mathbf{L}_j\mathbf{V}^{-1}(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta})$$

Then the system of equations can be written as

$$\left(\frac{1}{n}\sum_{i=1}^n v(d_i)\right)\mathbf{Q}\boldsymbol{\theta} = \mathbf{U}.$$

When $\sum_{i=1}^n v(d_i) \neq 0$ and $\mathbf{Q}$ is invertible, this is equivalent with the following fixed point equation for the vector of covariance parameters

$$\boldsymbol{\theta} = \left(\frac{1}{n}\sum_{i=1}^n v(d_i)\right)^{-1}\mathbf{Q}^{-1}\mathbf{U}.$$

For the linear mixed effects model in Copt and Victoria-Feser [2], $\mathbf{L}_j = \mathbf{Z}_j\mathbf{Z}_j^T$, so that $\mathbf{Q}$ and $\mathbf{U}$ are the same as in their paper. Note that instead of using $v(s) = u(s)s^2 - \rho(s) + b_0$, Copt and Victoria-Feser [2] use $v(s) = u(s)s^2$ and solve

$$\boldsymbol{\theta} = \left(\frac{1}{n}\sum_{i=1}^n u(d_i)d_i^2\right)^{-1}\mathbf{Q}^{-1}\mathbf{U},$$

see their equation (9). Moreover, they do not re-scale the covariance matrix $\mathbf{V}$, so that it satisfies the S-constraint.

I suggest to mimic the following iteration scheme, proposed by Rocke and Woodruff, 1993 [9], see their Figure 1 on page 30. The iteration scheme for our setup is listed in Figure 1. A couple of remarks concerning this scheme.

Starting values. In order to compute Mahalanobis distances, one needs at least a starting value for $\boldsymbol{\beta}$ and for $\mathbf{V}(\boldsymbol{\theta})$. Since $\mathbf{y}_i$ has expectation $\mathbf{X}_i\boldsymbol{\beta}$ and covariance $\mathbf{V}(\boldsymbol{\theta})$ the following procedure can be used at all times. Determine multivariate location and scatter estimates $\widehat{\boldsymbol{\mu}}$ and $\widehat{\mathbf{V}}$ from $\mathbf{y}_1,\dots,\mathbf{y}_n$. This can either be a robust estimate, such as the weighted OGK or MCD estimates, or just the ordinary sample mean and sample covariance. From this obtain a starting value for $\boldsymbol{\beta}$, by solving the fixed point equation To obtain a starting value for $\boldsymbol{\beta}$, one does one step of the vector fixed point equation

$$\boldsymbol{\beta}^{(0)} = \left(\sum_{i=1}^n w_i\mathbf{X}_i^T\widehat{\mathbf{V}}^{-1}\mathbf{X}_i\right)^{-1}\sum_{i=1}^n w_i\mathbf{X}_i^T\widehat{\mathbf{V}}^{-1}\mathbf{y}_i$$

where w_i are weights (0 or 1) obtained from the multivariate location-scatter estimates (for some robust estimates these are provided, otherwise take $w_i = 1$). Finally, take $\widehat{\mathbf{V}}$ as the starting value for $\mathbf{V}(\boldsymbol{\theta}^{(0)})$.

In some case $\mathbf{X}_i = \mathbf{X}$ is constant and of full rank. This means that $\boldsymbol{\mu} = \mathbf{X}\boldsymbol{\beta}$, so that we can use

$$\boldsymbol{\beta}^{(0)} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\widehat{\boldsymbol{\mu}}$$

This is much simpler, but has a disadvantage when $\mathbf{X}_i$ is not constant! I believe, that option 1 is actually better. Also note that if $\mathbf{X}_i = \mathbf{X}$ is constant and of full rank, after taking all $w_i = 1$, we get the same starting value. When $\mathbf{V}(\boldsymbol{\theta})$ has a simple structure, then it may be possible to solve $\boldsymbol{\theta}^{(0)}$ from

$$\mathbf{V}(\boldsymbol{\theta}^{(0)}) = \widehat{\mathbf{V}}.$$

However, this is not always possible and not necessary to start up the iteration.

Re-scaling. The rescaling in step 3 is precisely what seems to be missing in Copt and Victoria-Feser, 2006 [2]. This rescaling is to make sure that the weights in steps 5 and 6 are computed with Mahalanobis distances that satisfy the S-constraint. Note that if $\mathbf{V}$ is linear, then

$$\begin{aligned}
\left(\frac{d_i^{(j)}}{k^{(j)}}\right)^2 &= \left(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta}^{(j-1)}\right)^T \frac{\mathbf{V}\left(\boldsymbol{\theta}^{(j-1)}\right)^{-1}}{(k^{(j)})^2} \left(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta}^{(j-1)}\right) \\
&= \left(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta}^{(j-1)}\right)^T \left((k^{(j)})^2\mathbf{V}\left(\boldsymbol{\theta}^{(j-1)}\right)\right)^{-1} \left(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta}^{(j-1)}\right) \\
&= \left(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta}^{(j-1)}\right)^T \mathbf{V}\left((k^{(j)})^2\boldsymbol{\theta}^{(j-1)}\right)^{-1} \left(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta}^{(j-1)}\right).
\end{aligned}$$

This means that rescaling of the Mahalanobis distances amounts to rescaling of the vector of covariance parameters, i.e.,

$$\boldsymbol{\theta}^{(j)} = (k^{(j)})^2\boldsymbol{\theta}^{(j-1)}.$$

The re-scaling in step 7 is to make sure that after the iteration steps 5-6, we end up with updated estimates that satisfy the S-constraint.

NB: after checking it seems that this step is not necessary, since the re-scaling constant $k^{(j)}$ from step 3, converges to 1 during the while-loop. This coincides with the results in Tyler, 1988 [14] on computing M-estimates under constraints. In his setting $k^{(j)} \to 1$. Although our score equations do not satisfy his conditions, I believe we have the same phenomenon when using Tukey's biweight.

Use of ρ function to correct for S-constraint. Note that once we have rescaled the Mahalanobis distances in step 4, the function $v(s) = u(s)s^2 - \rho(s) + b_0$ in step 6 can be replaced by $u(s)s^2$, since

$$\frac{1}{n}\sum_{i=1}^n \rho\left(\widetilde{d}_i^{(j)}\right) = b_0.$$

The function $u(s)s^2$ was also used by Copt and Victoria-Feser, 2006 [2]. It follows that

$$\frac{1}{n}\sum_{i=1}^n v\left(\widetilde{d}_i^{(j)}\right) = \frac{1}{n}\sum_{i=1}^n u\left(\widetilde{d}_i^{(j)}\right)\left(\widetilde{d}_i^{(j)}\right)^2 - \frac{1}{n}\sum_{i=1}^n \rho\left(\widetilde{d}_i^{(j)}\right) + b_0 = \frac{1}{n}\sum_{i=1}^n u\left(\widetilde{d}_i^{(j)}\right)\left(\widetilde{d}_i^{(j)}\right)^2.$$

Copt and Victoria-Feser, 2006 [2] mention in Section 5.1 the rescaling of the Mahalanobis distances, for which they refer to Rocke, 1996 [10], but I believe that has to do with stability of the iteration and has nothing to do with satisfying the S-constraint.

Rocke, 1996 [10] mentions rescaling the Mahalanobis distances by the ratio of the median Mahalanobis distance and a corresponding chi-square quantile $d^2_{(q)}/\chi_k^{-1}(q/(n+1))$, where $q = \lfloor (n+k+1)/2 \rfloor$). This is to repair the instability of the bi-flat rho-function, of which the weight function

$u(s)$ has its mode away from the origin. Copt and Victoria-Feser, 2006 [2] use the translated biweight, of which the weight function is equal to 1 near the origin, so I do not understand why they re-scale.

MM step. After having obtained the CBS (or CTBS) estimators $\widehat{\boldsymbol{\beta}}$ and $\widehat{\boldsymbol{\theta}}$, one can perform an MM-step to gain efficiency. In that case, one keeps

$$\widehat{\mathbf{V}} = \mathbf{V}(\widehat{\boldsymbol{\theta}})$$

and adapts the cut-off constant to c_1, such that it corresponds to a given efficiency, say 95%:

$$\frac{k\alpha^2}{\mathbb{E}_{\mathbf{0},\mathbf{I}_k}\left[\rho'(\|\mathbf{z}\|;c_1)^2\right]} = 0.95.$$

The choice for the rho-function is optional, so one can either use the biweight or translated biweight. Heritier *et al* [5] use the biweight.

With given $\widehat{\mathbf{V}}$ and c_1, one updates the estimate for $\boldsymbol{\beta}$ by solving the fixed point equation

$$\boldsymbol{\beta} = \left(\sum_{i=1}^{n} u(d_i)\mathbf{X}_i^T\widehat{\mathbf{V}}^{-1}\mathbf{X}_i\right)^{-1}\sum_{i=1}^{n} u(d_i)\mathbf{X}_i^T\widehat{\mathbf{V}}^{-1}\mathbf{y}_i$$

starting from $\widehat{\boldsymbol{\theta}}$. Note that here, there is no reason to re-scale since there is no S-constraint involved. Also note that Heritier *et al* [5] use Rocke's adaptation also in this step. I believe that this is also not necessary.

On entry, we have observations $(\mathbf{y}_i, \mathbf{X}_i)$, $i = 1, \ldots, n$, matrices $\mathbf{L}_1, \ldots, \mathbf{L}_l$, and starting values $\boldsymbol{\beta}^{(0)}$ and $\boldsymbol{\theta}^{(0)}$. Moreover, we have chosen

$$b_0 = \mathbb{E}_{\mathbf{0},\mathbf{I}_k} \rho(\|\mathbf{z}\|) = r\rho(c),$$

where r is the desired breakdown point. Set the iteration $j \leftarrow 0$ and `NotConverged`$\leftarrow$`TRUE`.

While (`NotConverged`) do

1. $j \to j + 1$
2. Compute Mahalanobis distances

$$d_i^{(j)} = \sqrt{\left(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta}^{(j-1)}\right)^T \mathbf{V}\left(\boldsymbol{\theta}^{(j-1)}\right)^{-1}\left(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta}^{(j-1)}\right)}, \quad i = 1, \ldots, n$$

3. Compute $k^{(j)}$ as the solution of the constraint equation

$$\frac{1}{n}\sum_{i=1}^{n} \rho\left(\frac{d_i^{(j)}}{k^{(j)}}\right) = b_0$$

4. Replace $d_i^{(j)}$ with adjusted distances $\widetilde{d}_i^{(j)} = d_i^{(j)}/k^{(j)}$
5. Compute the new regression vector

$$\boldsymbol{\beta}^{(j)} = \left(\sum_{i=1}^{n} u\left(\widetilde{d}_i^{(j)}\right)\mathbf{X}_i^T\mathbf{V}\left(\boldsymbol{\theta}^{(j-1)}\right)^{-1}\mathbf{X}_i\right)^{-1}\sum_{i=1}^{n} u\left(\widetilde{d}_i^{(j)}\right)\mathbf{X}_i^T\mathbf{V}\left(\boldsymbol{\theta}^{(j-1)}\right)^{-1}\mathbf{y}_i$$

6. Compute the new covariance vector

$$\boldsymbol{\theta}^{(j)} = \left(\sum_{i=1}^{n} v\left(\widetilde{d}_i^{(j)}\right)\right)^{-1}\left(\mathbf{Q}^{(j-1)}\right)^{-1}\mathbf{U}^{(j-1)},$$

   where $\mathbf{Q}^{(j-1)}$ is the matrix

$$\mathbf{Q}^{(j-1)} = \left[\operatorname{tr}\left(\mathbf{V}\left(\boldsymbol{\theta}^{(j-1)}\right)^{-1}\mathbf{L}_s\mathbf{V}\left(\boldsymbol{\theta}^{(j-1)}\right)^{-1}\mathbf{L}_t\right)\right]_{s,t=1,\ldots,l},$$

   and $\mathbf{U}^{(j-1)}$ is the vector with elements

$$\mathbf{U}_s^{(j-1)} = k\sum_{i=1}^{n} u\left(\widetilde{d}_i^{(j)}\right)\left(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta}^{(j-1)}\right)^T\mathbf{V}\left(\boldsymbol{\theta}^{(j-1)}\right)^{-1}\mathbf{L}_s\mathbf{V}\left(\boldsymbol{\theta}^{(j-1)}\right)^{-1}\left(\mathbf{y}_i - \mathbf{X}_i\boldsymbol{\beta}^{(j-1)}\right),$$

7. Repeat steps 2 and 3 with $\boldsymbol{\beta}^{(j)}$ and $\boldsymbol{\theta}^{(j)}$, giving a new correction factor $\tilde{k}^{(j)}$ and adjust $\boldsymbol{\theta}^{(j)} \to \boldsymbol{\theta}^{(j)} \times \left(\tilde{k}^{(j)}\right)^2$.
8. If $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$ have not changed much, then `NotConverged`$\leftarrow$`FALSE`.

Return $\boldsymbol{\beta}^{(j)}$ and $\boldsymbol{\theta}^{(j)}$.

Figure 1: Iteration Scheme

3.6 Example in Copt and Victoria-Feser, 2016 [2]

Consider the first example from Section 5.1 in Copt and Victoria-Feser [2]. The example is taken from Richardson and Welsh, 1995 [8], and considers the following model, written as one vector-matrix equation

$$\mathbf{y} = \begin{pmatrix} \mathbf{1}_{20} & \mathbf{x} \end{pmatrix} \begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix} + \begin{pmatrix} \mathbf{1}_4 & & & & \\ & \cdot & & & \\ & & \cdot & & \\ & & & \cdot & \\ & & & & \mathbf{1}_4 \end{pmatrix} \begin{pmatrix} \beta_1 \\ \cdot \\ \cdot \\ \cdot \\ \beta_5 \end{pmatrix} + \boldsymbol{\epsilon}$$

see page 1435 in [8]. Translated into our notation this can be written as

$$\mathbf{y}_i = \mathbf{X}_i \boldsymbol{\alpha} + \beta_i \mathbf{1}_4 + \boldsymbol{\epsilon}_i, \quad i = 1, \ldots, 5,$$

or in detail

$$\begin{pmatrix} y_{i1} \\ y_{i2} \\ y_{i3} \\ y_{i4} \end{pmatrix} = \begin{pmatrix} 1 & x_{i1} \\ 1 & x_{i2} \\ 1 & x_{i3} \\ 1 & x_{i4} \end{pmatrix} \begin{pmatrix} \alpha_1 \\ \alpha_2 \end{pmatrix} + \beta_i \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} + \begin{pmatrix} \epsilon_{i1} \\ \epsilon_{i2} \\ \epsilon_{i3} \\ \epsilon_{i4} \end{pmatrix}, \quad i = 1, \ldots, 5.$$

We are dealing with a balanced linear mixed effects model with independent $\mathbf{y}_i$ in dimension $k = 4$, with fixed effects $\boldsymbol{\alpha} = (\alpha_1, \alpha_2)^T$ and corresponding 4×2 design matrices $\mathbf{X}_i$, and random effects $\beta_1, \ldots, \beta_4$ and corresponding design matrices being the same

$$\mathbf{Z}_1 = \mathbf{Z}_2 = \mathbf{Z}_3 = \mathbf{Z}_4 = \mathbf{Z} = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}.$$

For the simulation, the random effects where chosen independently, such that

$$\beta_i \sim N(0, \sigma_\beta^2)$$

and the measurement errors $\boldsymbol{\epsilon}_1, \ldots, \boldsymbol{\epsilon}_4$ independent, such that

$$\boldsymbol{\epsilon}_i = \begin{pmatrix} \epsilon_{i1} \\ \epsilon_{i2} \\ \epsilon_{i3} \\ \epsilon_{i4} \end{pmatrix} \sim N(\mathbf{0}, \sigma_\epsilon^2 \mathbf{I}_4).$$

This means that

$$\mathbf{y}_i \mid \mathbf{X}_i \sim N(\mathbf{X}_i \boldsymbol{\alpha}, \boldsymbol{\Sigma}),$$

such that $\boldsymbol{\Sigma} = \sigma_\beta^2 \mathbf{Z}\mathbf{Z}^T + \sigma_\epsilon^2 \mathbf{I}_4$.

In Copt and Victoria-Feser [2] the parameters where set to

$$\alpha_1 = \alpha_2 = \sigma_\beta^2 = \sigma_\epsilon^2 = 1.$$

Moreover, the 20 x-values in $\mathbf{x}$ were generated beforehand independently from a $N(0, 1)$ distribution, and re-scaled to a new 20×2 matrix

$$\begin{pmatrix} \mathbf{1}_{20} & \mathbf{x} \end{pmatrix}$$

such that

$$\begin{pmatrix} \mathbf{1}_{20} & \mathbf{x} \end{pmatrix}^T \begin{pmatrix} \mathbf{1}_{20} & \mathbf{x} \end{pmatrix} = 20\mathbf{I}_2.$$

One way to obtain the re-scaling, is to re-scale for each individual $i = 1, \dots, 5$,

$$\begin{aligned} \sum_{j=1}^{5} x_{ij}^2 &= 4, \\ \sum_{j=1}^{5} x_{ij} &= 0. \end{aligned} \tag{3.2}$$

In this way we will have a linear mixed effects model

$$\mathbf{y}_i = \mathbf{X}_i \boldsymbol{\alpha} + \beta_i \mathbf{Z} + \boldsymbol{\epsilon}_i, \quad i = 1, \dots, 5,$$

where $\boldsymbol{\Sigma} = \sigma_\beta^2 \mathbf{Z}\mathbf{Z}^T + \sigma_\epsilon^2 \mathbf{I}_4$ and the $\mathbf{X}_i$ are independent copies of the random matrix

$$\mathbf{X} = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ 1 & x_3 \\ 1 & x_4 \end{pmatrix},$$

which satisfies

$$\mathbf{X}^T \mathbf{X} = 4\mathbf{I}_2,$$

with probability one.

Note that Copt and Victoria-Feser [2] find different expressions for the asymptotic covariances of $\widehat{\boldsymbol{\beta}}_n$ and $\widehat{\boldsymbol{\theta}}_n$. With the simple settings in Copt and Victoria-Feser [2] these different expressions lead to the same asymptotic covariance matrices as obtained from our results. In order to check which expression is the correct one, I have performed a simulation with a slightly different setting.

Recall that the dimension is $k = 4$ and suppose we choose breakdown point $\epsilon = 0.5$. The R-code

```
k=4
r=0.5
c0=rhoconst(k,r,0.01,100)
b0=expecrho(k,c0)
```

leads to cut-off constant $c = 4.096567$ and $b_0 = 1.398486$ in the S-constraint. I kept the same settings for the unknown parameters with the R-code

```
sigmaeps=1  # variance of error
sigmabeta=1 # variance of random effects
alpha1=1    # fixed effect 1
alpha2=1    # fixed effect 2
fixeff=c(alpha1,alpha2)     # vector of fixed effects parameters
theta=c(sigmabeta,sigmaeps) # vector of covariance parameters
```

and used the same orthogonalized 4×2 design matrix $\mathbf{X}_i = \mathbf{X}$ with the R-command

```
set.seed(123)
X=cbind(1,rnorm(k))
X=cbind(1,data_sphe(X))
```

This produces

$$\mathbf{X} = \begin{pmatrix} 1 & -0.9504967 \\ 1 & -0.5428346 \\ 1 & 1.6650521 \\ 1 & -0.1717207 \end{pmatrix}$$

Moreover, to make sure that the empirical covariances of the estimates are close to the theoretical ones based on asymptotic normality, I took the sample size $n = 100$. In contrast to Copt and Victoria-Feser [2], I used a different design matrix $\mathbf{Z}$ for the random effects

$$\mathbf{Z}_1 = \mathbf{Z}_2 = \mathbf{Z}_3 = \mathbf{Z}_4 = \mathbf{Z} = \begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \end{pmatrix}.$$

This by itself is enough to yield differences in the limiting covariance matrix of $\widehat{\boldsymbol{\beta}}_n$. In order to create differences in the limiting covariance matrix of $\widehat{\boldsymbol{\theta}}_n$, I took a different covariance matrix for the measurement error

$$\boldsymbol{\epsilon}_i = \begin{pmatrix} \epsilon_{i1} \\ \epsilon_{i2} \\ \epsilon_{i3} \\ \epsilon_{i4} \end{pmatrix} \sim N(\mathbf{0}, \sigma_\epsilon^2 \mathbf{R}), \qquad \mathbf{R} = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0 & 4 & 0 & 0 \\ 0 & 0 & 9 & 0 \\ 0 & 0 & 0 & 16 \end{pmatrix}.$$

The choice of different variances on the main diagonal of $\mathbf{R}$ is essential for creating a difference in the asymptotic covariances for $\widehat{\boldsymbol{\theta}}_n$. With the same values on the main diagonal, our expressions and the one in Copt and Victoria-Feser [2] will result in the same covariance matrix. The above settings have been coded as

```
Z=1:4               # THIS SETTING GIVES DIFFERENCES IN COVARIANCE OF BETAHAT
varerror=diag(c(1,4,9,16))

# matrices L in V=sigmabeta*L1+sigmaeps*L2
L1=Z%*%t(Z)        # covariance part due to random effect
L2=varerror        # covariance part due to measurement error
Vtheta=theta[1]*L1+theta[2]*L2
L=as.matrix(cbind(vec(L1),vec(L2)))
```

The asymptotic variances for $\widehat{\boldsymbol{\beta}}_n$ are coded

```
# Expression from our theory
varbeta=constbetahat(k,c0)*
    solve(t(X)%*%solve(Vtheta)%*%X)

# Expression from Copt & Victoria-Feser (2006)
varbetaCopt=constbetahat(k,c0)*
    solve(t(X)%*%X) %*% t(X) %*% Vtheta %*% X %*% solve(t(X)%*%X)
```

This leads to

$$\texttt{varbeta} = \begin{pmatrix} 5.309162 & 2.757777 \\ 2.757777 & 2.486482 \end{pmatrix}$$

and

$$\texttt{varbetaCopt} = \begin{pmatrix} 10.158464 & 2.487822 \\ 2.487822 & 2.552743 \end{pmatrix}.$$

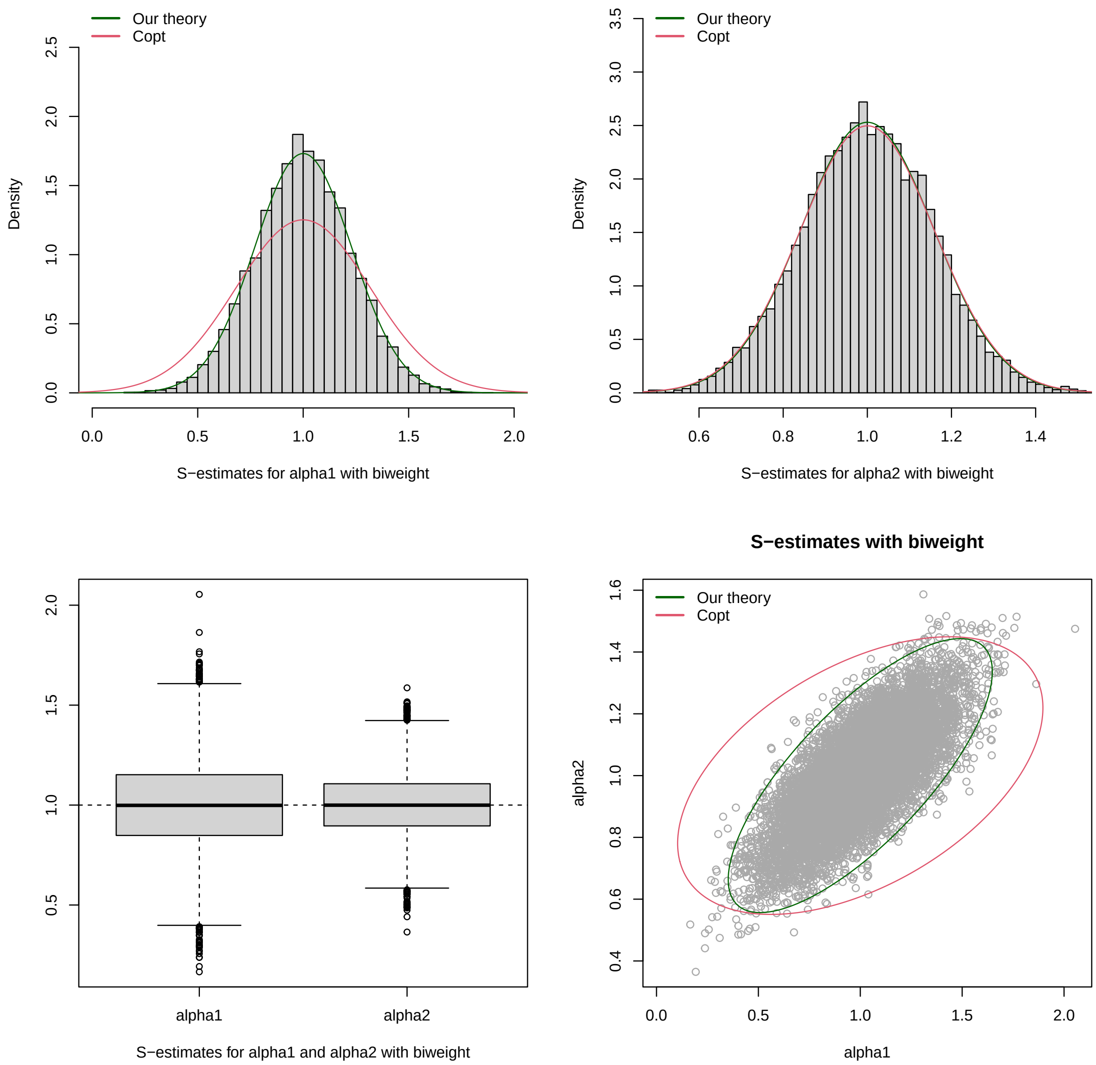

Figure 2: Empirical and asymptotic distributions of $\boldsymbol{\beta}_n$ computed with Tukey's biweight.

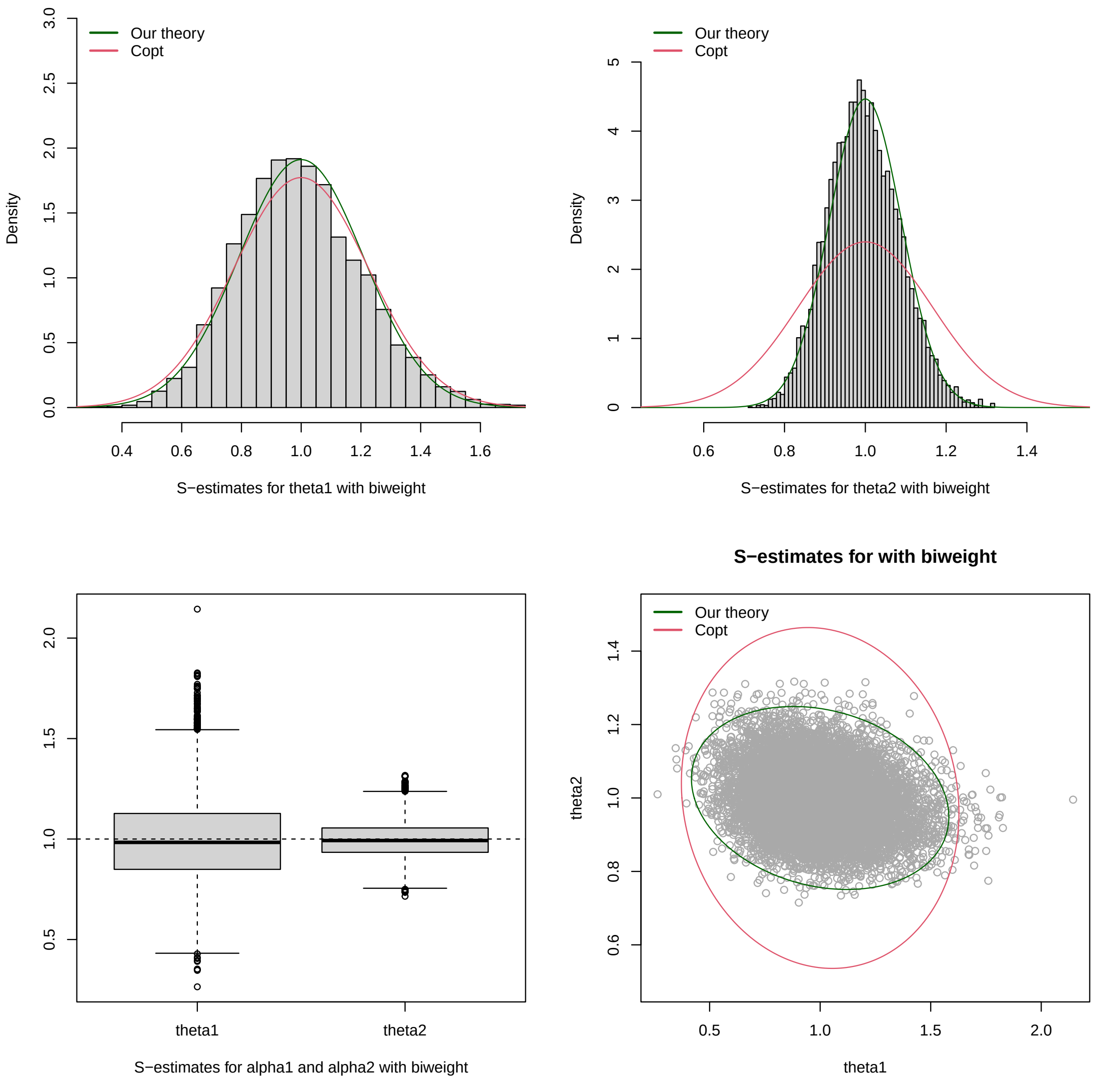

Figure 3: Empirical and asymptotic distributions of $\boldsymbol{\theta}_n$ computed with Tukey's biweight.

The results of a simulation with 10 000 replications of the iteration described in Figure 1 can be seen in Figure 2. We can see a clear match between the empirical (co)variances with our theoretical results (in green), which deviates completely from the predictions by Copt and Victoria-Feser [2] (in red). I should emphasize that even greater differences can be obtained by choosing different settings.

The asymptotic variances for $\widehat{\boldsymbol{\theta}}_n$ are coded

```
# Expression from our theory
vartheta=(2*sigma1(k,c0)*
    solve(t(L)%*%(solve(Vtheta)%x%solve(Vtheta))%*%L)+
    sigma2(k,c0)*theta%*%t(theta))

# Expression from Copt & Victoria-Feser (2006)
varthetaCopt=2*sigma1(k,c0)*
    solve((t(L)%*%L))%*%t(L)%*%(Vtheta%x%Vtheta)%*%L%*%solve((t(L)%*%L))+
    sigma2(k,c0)*theta%*%t(theta)
```

This leads to

$$\texttt{vartheta} = \begin{pmatrix} 4.3569335 & -0.3881412 \\ -0.3881412 & 0.7981275 \end{pmatrix}$$

and

$$\texttt{varthetaCopt} = \begin{pmatrix} 5.0652567 & -0.3297755 \\ -0.3297755 & 2.7692047 \end{pmatrix}.$$

The results of a simulation with 10 000 replications of the iteration described in Figure 1 can be seen in Figure 3. Again we see a clear match between the empirical (co)variances with our theoretical results (in green), which deviates completely from the predictions by Copt and Victoria-Feser [2] (in red). Once more, I should emphasize that even greater differences can be obtained by choosing different settings.

There is one thing that bothers me a little. If you take a close look at the histograms of $\widehat{\theta}_{1n}$ (top left) and $\widehat{\theta}_{2n}$ (top right) in Figure 3, I think you can see a slight bias to the left. I have no idea what causes this phenomena. Perhaps sample size $n = 100$ is too small in order to erase possible asymptotic bias. Note that in the simulation in Copt and Victoria-Feser [2] we see the same phenomenon in Figure 1.

4 Rocke's translated biweight

Instead of Tukey's biweight, Copt and Victoria-Feser [2] use the translated biweight proposed by Rocke, 1996 [10]. For this reason, I have studied the paper by Rocke, 1996 [10] and implemented the translated biweight and important quantities that correspond to this rho function.

4.1 Understanding the paper by Rocke (1996)

Rocke, 1996 [10] investigates the asymptotic rejection rate

$$\text{ARP} = \mathbb{P}_{\mathbf{0},\mathbf{I}_k}\left(\|\mathbf{z}\| \geq c\right) = \mathbb{P}(W_k \geq c^2), \qquad W_k \sim \chi^2_k,$$

of Tukey's biweight function $\rho(\cdot; c)$ in high dimensions, when c is chosen such that the S-estimator has a high breakdown point, say $r = 0.5$. Given a breakdown point r, the cut-off constant c is usually chosen, such that

$$\frac{\mathbb{E}_{\mathbf{0},\mathbf{I}_k}\rho(\|\mathbf{z}\|; c)}{\rho(c; c)} = r.$$

This means that for different dimensions, the value for c changes. It turns out that for larger dimensions k, the corresponding ARP decreases to zero extremely fast as k increases, which means in practice that outliers in high dimensions that are relatively far from the center of the data, are not given weight zero.

In Lopuhaä [7] and other papers, determination of the cut-off constant c is adapted when the dimension k is changing. For Tukey's biweight this is done by solving c from the equation

$$\frac{\mathbb{E}_k\|\mathbf{z}\|^2\{\|\mathbf{z}\| \leq c\}}{2} - \frac{\mathbb{E}_k\|\mathbf{z}\|^4\{\|\mathbf{z}\| \leq c\}}{2c^2} + \frac{\mathbb{E}_k\|\mathbf{z}\|^6\{\|\mathbf{z}\| \leq c\}}{6c^4} + \frac{c^2}{6}\mathbb{E}_k\{\|\mathbf{z}\| > c\} = r\frac{c^2}{6},$$

where $\mathbb{E}_k$ refers to taking expectation with respect to the k-variate normal, or equivalently, solving c from the equation

$$\frac{3\mathbb{E}_k\|\mathbf{z}\|^2\{\|\mathbf{z}\| \leq c\}}{c^2} - \frac{3\mathbb{E}_k\|\mathbf{z}\|^4\{\|\mathbf{z}\| \leq c\}}{c^4} + \frac{\mathbb{E}_k\|\mathbf{z}\|^6\{\|\mathbf{z}\| \leq c\}}{c^6} + \mathbb{E}_k\{\|\mathbf{z}\| > c\} = r,$$

where $\|\mathbf{z}\|^2 \sim \chi^2_k$. For Tukey's biweight this means that also the form of the rho-function is adapted. Rocke [10] has a different, but equivalent approach. The rho-function including the cut-off constant, say c_0, is fixed, and then the rho-function is scaled internally. That is, the S-constraint becomes

$$\frac{1}{n}\sum_{i=1}^{n}\rho\left(\frac{d_i}{c_k}; c_0\right) = b_0 = \mathbb{E}_{\mathbf{0},\mathbf{I}_k}\rho\left(\frac{\|\mathbf{z}\|}{c_k}; c_0\right)$$

and c_k is solved from

$$\mathbb{E}_{\mathbf{0},\mathbf{I}_k}\rho\left(\frac{\|\mathbf{z}\|}{c_k}; c_0\right) = r\rho(c_0; c_0),$$

(I believe that Theorem 1 in Rocke [10] contains a typo: $b_p = r\rho(c_p)$, should be $b_p = \rho(c_0)$). For Tukey's biweight we have

$$\rho\left(\frac{y}{c_k}; c_0\right) = \begin{cases} \dfrac{y^2}{2c_k^2} - \dfrac{y^4}{2c_k^4c_0^2} + \dfrac{y^6}{6c_k^6c_0^4} & , |y| \leq c_kc_0 \\ \dfrac{c_0^2}{6} & , |y| \geq c_kc_0. \end{cases}$$

and

$$\frac{\rho\left(\dfrac{y}{c_k}; c_0\right)}{\rho(c_0; c_0)} = \begin{cases} \dfrac{3y^2}{c_k^2c_0^2} - \dfrac{3y^4}{c_k^4c_0^4} + \dfrac{y^6}{c_k^6c_0^6} & , |y| \leq c_kc_0 \\ 1 & , |y| \geq c_kc_0. \end{cases}$$

For Tukey's biweight, determining c_k is therefore equivalent with solving c_k from the equation

$$\frac{3\mathbb{E}_k\|\mathbf{z}\|^2\{\|\mathbf{z}\| \leq c_k c_0\}}{c_k^2 c_0^2} - \frac{3\mathbb{E}_k\|\mathbf{z}\|^4\{\|\mathbf{z}\| \leq c_k c_0\}}{c_k^4 c_0^4} + \frac{\mathbb{E}_k\|\mathbf{z}\|^6\{\|\mathbf{z}\| \leq c_k c_0\}}{c_k^6 c_0^6} + \mathbb{E}_k\{\|\mathbf{z}\| > c_k c_0\} = r,$$

where $\|\mathbf{z}\|^2 \sim \chi_k^2$. Clearly, the use of the original biweight (with cut-off c) or the re-scaled biweight (with cut-off $c = c_k c_0$) leads to the same rho-function that corresponds to a breakdown point r.

4.2 Rocke's translated biweight rho-function

Rocke, 1996 [10], proposed a translated biweight function to improve the asymptotic rejection rates

$$\text{ARP} = \mathbb{P}_{\mathbf{0},\mathbf{I}_k}\left(\|\mathbf{z}\| \geq c\right)$$

of the regular biweight function in high dimensions. It has two cut-off parameters M and c and is defined on $[0, \infty)$ as

$$\rho_t(d) = \rho_t(d; M, c) = \begin{cases} d^2/2 & , 0 \leq d < M \\ M^2/2 - M^2(M^4 - 5M^2c^2 + 15c^4)/(30c^4) & \\ \qquad + d^2(1/2 + M^4/(2c^4) - M^2/c^2) & \\ \qquad + d^3(4M/(3c^2) - 4M^3/(3c^4)) & \\ \qquad + d^4(3M^2/(2c^4) - 1/(2c^2)) & , M \leq d \leq M + c \\ \qquad - 4Md^5/(5c^4) & \\ \qquad + d^6/(6c^4) & \\ M^2/2 + c(5c + 16M)/30 & , d > M + c \end{cases}$$

For later convenience, we write this as

$$\rho_t(d) = \begin{cases} d^2/2 & , 0 \leq d < M \\ a_{1t} + a_{2t} + a_{3t}d^2 + a_{4t}d^3 + a_{5t}d^4 + a_{6t}d^5 + a_{7t}d^6 & , M \leq d \leq M + c \\ a_{8t} & , d > M + c \end{cases}$$

where

$$\begin{aligned} a_{1t} &= M^2/2 \\ a_{2t} &= -M^2(M^4 - 5M^2c^2 + 15c^4)/(30c^4) \\ a_{3t} &= 1/2 + M^4/(2c^4) - M^2/c^2 \\ a_{4t} &= 4M/(3c^2) - 4M^3/(3c^4) \\ a_{5t} &= 3M^2/(2c^4) - 1/(2c^2) \\ a_{6t} &= -4M/(5c^4) \\ a_{7t} &= 1/(6c^4) \\ a_{8t} &= M^2/2 + c(5c + 16M)/30. \end{aligned} \tag{4.1}$$

There are two limiting cases: (i) for $M = 0$ one obtains Tukey's biweight rho-function and (ii) for $c \downarrow 0$ one obtains the rho-function that corresponds to the winsorized mean:

$$\rho_w(d) = \begin{cases} d^2/2 & , 0 \leq d < M \\ M^2/2 & , d > M. \end{cases}$$

The rho-function ρ_t is the continuous primitive of Rocke's translated biweight psi-function given by

$$\psi_t(d) = \rho_t'(d) = \begin{cases} d & ,0 \le d < M \\ d\left(1 - \dfrac{(d-M)^2}{c^2}\right)^2 & ,M \le d \le M+c \\ 0 & ,d > M+c \end{cases}$$

$$= \begin{cases} d & ,0 \le d < M \\ d(1 + M^4/c^4 - 2M^2/c^2) & \\ \quad + d^2(4M/c^2 - 4M^3/c^4) & \\ \quad + d^3(6M^2/c^4 - 2/c^2) & ,M \le d \le M+c \\ \quad - 4Md^4/c^4 & \\ \quad + d^5/c^4 & \\ 0 & ,d > M+c \end{cases}$$

Finally, the derivative of ψ_t is also important in computing the constants. It is given by

$$\psi_t'(d) = \rho_t''(d) = \begin{cases} 1 & ,0 \le d < M \\ 1 + M^4/c^4 - 2M^2/c^2 & \\ \quad + d(8M/c^2 - 8M^3/c^4) & \\ \quad + d^2(18M^2/c^4 - 6/c^2) & ,M \le d \le M+c \\ \quad - 16Md^3/c^4 & \\ \quad + 5d^4/c^4 & \\ 0 & ,d > M+c \end{cases}$$

For comparison, the graphs of the ρ-functions and some of their derivatives are displayed in Figure 4. Tukey's biweight (black) with cut-off parameter c corresponding to 50% breakdown in dimension $k = 4$, Rocke's translated biweight (red) with cut-off parameters M and c corresponding to ARP $\pi = 0.01$ and 50% breakdown in dimension $k = 4$, and the winsorized mean (green) with the same value M.

The function $u(d) = \psi(d)/d$ (bottom left) represents the weight function in the fixed point equations for the regression and covariance parameters. The function $u(d)d^2 = \psi(d)d$ (bottom right) represents the weight function in the fixed point equation for the vector of covariance parameters.

4.3 Choosing the cut-off constants of the translated biweight

In Section 3.3 of Copt and Victoria-Feser [2], it is explained how the CTBS (constraint translated biweight S) estimator is computed. Given a breakdown point r and an asymptotic rejection rate $\pi = 1 - F_{\chi^2_k}((M+c)^2)$, the two parameters should satisfy

$$\mathbb{E}_{\mathbf{0},\mathbf{I}_k}\rho_t(\|\mathbf{z}\|) = r\rho_t(M+c),$$

and

$$M + c = \sqrt{F^{-1}_{\chi^2_k}(1-\pi)},$$

where $F^{-1}_{\chi^2_k}$ denotes the quantile function of the chi-square distribution with k degrees of freedom. The natural thing to do seems to express M in terms of c:

$$M = \sqrt{F^{-1}_{\chi^2_k}(1-\pi)} - c$$

and solve c from

$$\mathbb{E}_{\mathbf{0},\mathbf{I}_k}\rho_t(\|\mathbf{z}\|) = r\rho_t(M+c) = r\rho_t\left(\sqrt{F^{-1}_{\chi^2_k}(1-\pi)}\right).$$

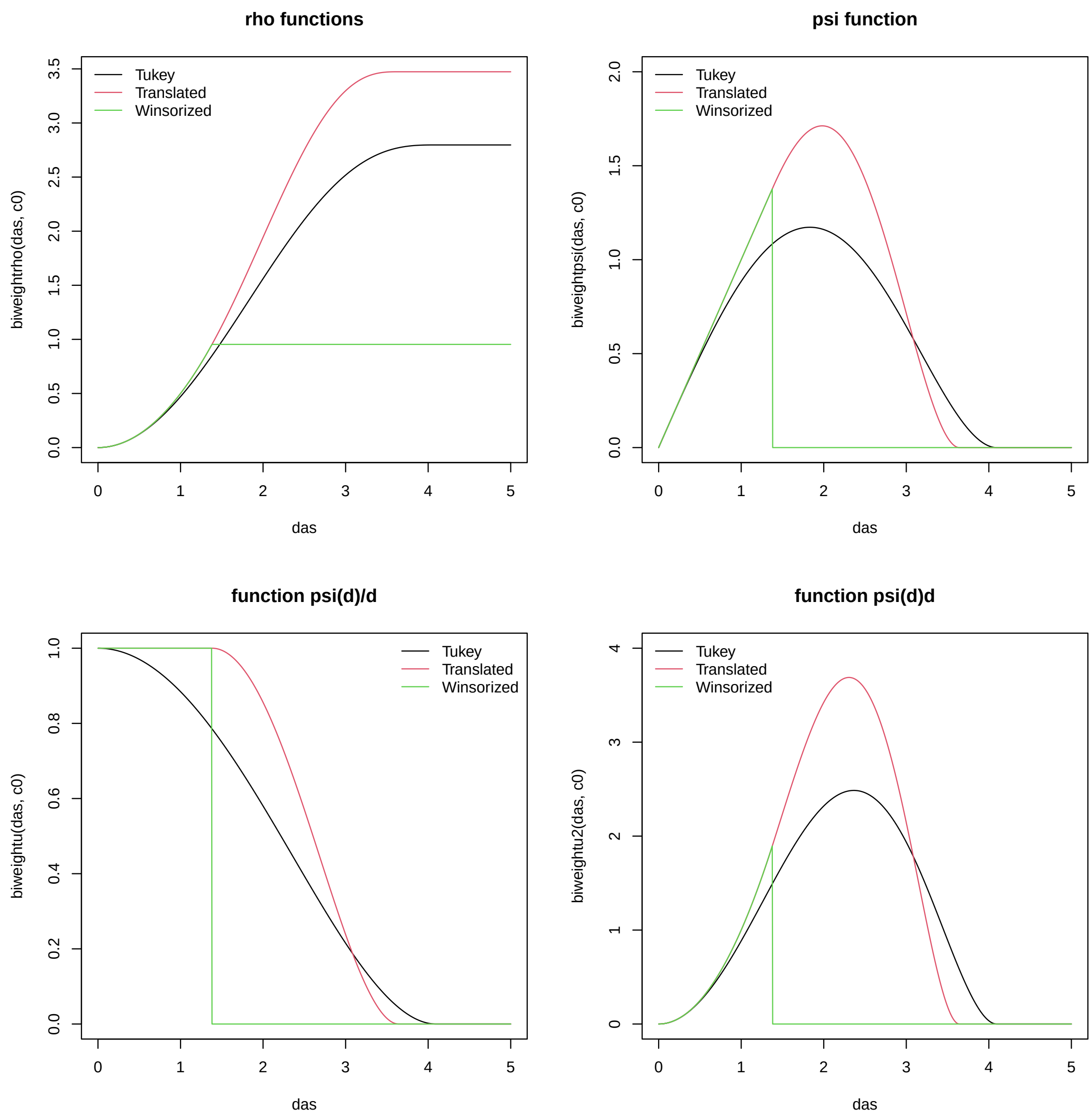

Figure 4: Biweight and translated biweight functions

This is indeed implemented in the R-function `constantC` in the R-script `simulation.r`. One may check that in dimensions $k =, 1, 2$, but also $k = 30, 40, 50$ this R-function does not give a value!

Nevertheless, at this point I want to compare the results of the simulation with our theoretical expressions. Therefore, I have implemented a similar R-function that makes use of explicitly computed moments. To this end we first implement $\mathbb{E}_{\mathbf{0},\mathbf{I}_k}\rho_t(\|\mathbf{z}\|)$ by the following R-functions. First we implement the moments

$$\mathbb{E}_{\mathbf{0},\mathbf{I}_k}\{a \le \|\mathbf{z}\| \le b\} = \mathbb{E}_{\mathbf{0},\mathbf{I}_k}\{\|\mathbf{z}\| \le b\} - \mathbb{E}_{\mathbf{0},\mathbf{I}_k}\{\|\mathbf{z}\| \le a\}$$

that are involved, by means of the following R-function.

```
momenttranslated=function(d,k,a,b){
  # computes E[|Z|^d{a<|Z|<b}]
  # needed for expectations of translated biweight functions
  return(moment(d,k,b)-moment(d,k,a))
  }
```

Next, we implement $\mathbb{E}_{\mathbf{0},\mathbf{I}_k}\rho_t(\|\mathbf{z}\|)$ by means of the following R-function.

```
expecrhotranslated=function(k,M,c0){
  # computes E[rho_M(|Z|)]
  # M=first parameter of translated biweight
  # c0=second parameter of translated biweight
  # k=dimension of multivariate normal

  # compute E[|Z|^2/2{|Z|>=M}]
  term1=moment(2,k,M)/2

  # compute the expectation over [M,M+c0]
  term2=momenttranslated(0,k,M,M+c0)*M^2/2-
    momenttranslated(0,k,M,M+c0)*
      M^2*(M^4-5*M^2*c0^2+15*c0^4)/(30*c0^4)+
    momenttranslated(2,k,M,M+c0)*(1/2+M^4/(2*c0^4)-M^2/c0^2)+
    momenttranslated(3,k,M,M+c0)*(4*M/(3*c0^2)-4*M^3/(3*c0^4))+
    momenttranslated(4,k,M,M+c0)*(3*M^2/(2*c0^4)-1/(2*c0^2))-
    momenttranslated(5,k,M,M+c0)*4*M/(5*c0^4)+
    momenttranslated(6,k,M,M+c0)/(6*c0^4)

  # compute expectation over [M+c0,Inf]
  term3=(1-moment(0,k,M+c0))*(M^2/2+c0*(5*c0+16*M)/30)

  return(term1+term2+term3)
}
```

Finally, the following R-function determines the constant c of ρ_t, given the dimension k, the breakdown point r, the ARP π, and M such that $M + c = \sqrt{F^{-1}_{\chi^2_k}(1-\pi)}$.

```
rhotranslatedconst=function(k,r,p,a,b){
  # computes the cut=off constant c0 for which
  # E[rho_t(|Z|)]=r*rho_t(M+c0),
  # where rho_t is the translated biweight
  # M chosen such that p=pchisq((M+c0)^2,df=k)
  # k=dimension of the multivariate normal
  # r=breakdown point
  # p=asymptotic rejection rate
  # a,b are the boundaries of the interval over which we use uniroot
```

```
  # compute E[rho_t(|Z|)]-r*rho_t(M+c0)
  objectfun=function(c0){
    # M chosen such that p=pchisq((M+c0)^2,df=k)
    M=sqrt(qchisq(1-p,df=k))-c0
    expecrhotranslated(k,M,c0)-
      r*biweightrhotranslated(M+c0,M,c0)
  }
  return(uniroot(f=objectfun,c(a,b))$root)
}
```

For example, for $k = 5$, $r = 0.5$, and $\pi = 0.01$, we find $c = 1.866563$ and $M = 2.017542$, which corresponds to a translated rho-function that has cut-off point $M + c = 3.884105$. Note that the computation $\mathbb{E}_{\mathbf{0},\mathbf{I}_k}\rho_t(\|\mathbf{z}\|)$ is explicit. In contrast, five repetitions of Copt's function `constantC` gives values ranging between 1.846311 and 1.895152!

4.4 Asymptotic covariance of the regression CTBS estimator

When the data are generated from the distribution P of $(\mathbf{y}, \mathbf{X})$, which is such that $\mathbf{y} \mid \mathbf{X} \sim N(\mathbf{X}\boldsymbol{\beta}, \boldsymbol{\Sigma})$, then according to Corollary 5, the asymptotic covariance of the constrained translated beiweight S-estimator (CTBS) for regression defined with ρ_t is given by

$$\frac{\mathbb{E}_{\mathbf{0},\mathbf{I}_k}\left[\rho_t'(\|\mathbf{z}\|)^2\right]}{k\alpha_t^2}\left(\mathbb{E}\left[\mathbf{X}^T\boldsymbol{\Sigma}^{-1}\mathbf{X}\right]\right)^{-1}$$

where

$$\alpha_t = \mathbb{E}_{\mathbf{0},\mathbf{I}_k}\left[\left(1-\frac{1}{k}\right)\frac{\rho_t'(\|\mathbf{z}\|)}{\|\mathbf{z}\|} + \frac{1}{k}\rho_t''(\|\mathbf{z}\|)\right].$$

Let us start with the constant α_t. From Section 4.2 we have

$$\begin{aligned}
\frac{\rho_t'(d)}{d} &= \begin{cases} 1 & ,0 \le d < M \\ \left(1 - \dfrac{(d-M)^2}{c^2}\right)^2 & ,M \le d \le M+c \\ 0 & ,d > M+c \end{cases} \\
&= \begin{cases} 1 & ,0 \le d < M \\ \begin{aligned}&1 + M^4/c^4 - 2M^2/c^2 \\ &\quad + d(4M/c^2 - 4M^3/c^4) \\ &\quad + d^2(6M^2/c^4 - 2/c^2) \\ &\quad - 4Md^3/c^4 \\ &\quad + d^4/c^4\end{aligned} & ,M \le d \le M+c \\ 0 & ,d > M+c \end{cases}
\end{aligned}$$

and

$$\rho_t''(d) = \begin{cases} 1 & ,0 \le d < M \\ \begin{aligned}&1 + M^4/c^4 - 2M^2/c^2 \\ &\quad + d(8M/c^2 - 8M^3/c^4) \\ &\quad + d^2(18M^2/c^4 - 6/c^2) \\ &\quad - 16Md^3/c^4 \\ &\quad + 5d^4/c^4\end{aligned} & ,M \le d \le M+c \\ 0 & ,d > M+c \end{cases}$$

This leads to the following R-function.

```
alphatranslated=function(k,M,c0){
  # computes the value of the constant alpha in (8.4)
  # for the translated biweight rho function
  # k = dimension of the normal distribution
  # M = first parameter of the translated rho-function
  # c0 = second parameter of the translated rho-function

  # compute E[rho_t'(d)/d]
  term1=moment(0,k,M)+
    momenttranslated(0,k,M,M+c0)*(1+M^4/(c0^4)-2*M^2/(c0^2))+
    momenttranslated(1,k,M,M+c0)*(4*M/(c0^2)-4*M^3/(c0^4))+
    momenttranslated(2,k,M,M+c0)*(6*M^2/(c0^4)-2/(c0^2))-
    momenttranslated(3,k,M,M+c0)*4*M/(c0^4)+
    momenttranslated(4,k,M,M+c0)/(c0^4)

  # compute E[rho_t''(d)]
  term2=moment(0,k,M)+
    momenttranslated(0,k,M,M+c0)*(1+M^4/(c0^4)-2*M^2/(c0^2))+
    momenttranslated(1,k,M,M+c0)*(8*M/(c0^2)-8*M^3/(c0^4))+
    momenttranslated(2,k,M,M+c0)*(18*M^2/(c0^4)-6/c0^2)-
    momenttranslated(3,k,M,M+c0)*16*M/(c0^4)+
    momenttranslated(4,k,M,M+c0)*5/c0^4

  return((1-1/k)*term1+(1/k)*term2)
}
```

For $\mathbb{E}_{\mathbf{0},\mathbf{I}_k}\left[\rho_t'(\|\mathbf{z}\|)^2\right]$, note that from Section 4.2, we have

$$\rho_t'(d)^2 = \begin{cases} d^2 & ,0 \le d < M \\ d^2\left(1-\dfrac{(d-M)^2}{c^2}\right)^4 & ,M \le d \le M+c \\ 0 & ,d > M+c \end{cases}$$
$$= \begin{cases} d^2 & ,0 \le d < M \\ \begin{aligned} &d^2 - (4/c^2)d^2(d-M)^2 \\ &\quad + (6/c^4)d^2(d-M)^4 \\ &\quad - (4/c^6)d^2(d-M)^6 \\ &\quad + (1/c^8)d^2(d-M)^8 \end{aligned} & ,M \le d \le M+c \\ 0 & ,d > M+c \end{cases}$$

This suggest to first create an R-function that computes

$$\mathbb{E}_{\mathbf{0},\mathbf{I}_k}\left[\|\mathbf{z}\|^l\Big(\|\mathbf{z}\|-M\Big)^m\{M \le \|\mathbf{z}\| \le M+c\}\right]$$

where

$$\|\mathbf{z}\|^l\Big(\|\mathbf{z}\|-M\Big)^m = \|\mathbf{z}\|^l\sum_{i=0}^m\binom{m}{i}(-M)^i\|\mathbf{z}\|^{m-i} = \sum_{i=0}^m\binom{m}{i}(-M)^i\|\mathbf{z}\|^{l+m-i}$$

This is done by the following R-function

```
momentcentered=function(l,m,k,M,a,b){
  # computes E[|Z|^l(|Z|-M)^m{a<|Z|<b}]
  # need for E[rho_t'(|Z|)^2] in alpha_t
  ss=0
```

```
  for (i in 0:m){
    ss=ss+choose(m,i)*((-M)^i)*momenttranslated(l+m-i,k,a,b)
  }
  return(ss)
}
```

The following R-function implements $\mathbb{E}_{\mathbf{0},\mathbf{I}_k}\left[\rho_t'(\|\mathbf{z}\|)^2\right]$.

```
psitranslatedsquared=function(k,M,c0){
  # computes the numerator E[rho_t'(|Z|)^2] appearing in the
  # limiting covariance of betahat
  # see Corollary 5

  # computes E[|Z|{|Z|<=M}]
  term1=moment(2,k,M)

  # computes expectation over [M,M+c0]
  term2=momenttranslated(2,k,M,M+c0)-
    (4/(c0^2))*momentcentered(2,2,k,M,M,M+c0)+
    (6/(c0^4))*momentcentered(2,4,k,M,M,M+c0)-
    (4/(c0^6))*momentcentered(2,6,k,M,M,M+c0)+
    (1/(c0^8))*momentcentered(2,8,k,M,M,M+c0)

  return(term1+term2)
}
```

Both R-functions can be combined to compute the constant

$$\frac{\mathbb{E}_{\mathbf{0},\mathbf{I}_k}\left[\rho_t'(\|\mathbf{z}\|)^2\right]}{k\alpha_t^2}$$

by means of the following R-function.

```
constbetahattranslated=function(k,M,c0){
  # determines the scalar in the limiting covariance
  # of betahat for the translated biweight
  # k=dimension of multivariate normal
  # M is chosen sqrt(qchisq(1-pi),df=k)-c0
  # where pi=asymptotic rejection rate
  # c0=pre-determined cut-off value for given BDP r
  return(psitranslatedsquared(k,M,c0)/(k*alphatranslated(k,M,c0)^2))
}
```

Recall that the constant

$$\lambda_t = \frac{\mathbb{E}_{\mathbf{0},\mathbf{I}_k}\left[\rho_t'(\|\mathbf{z}\|)^2\right]}{k\alpha_t^2}$$

represents the asymptotic efficiency of the regression S-estimator computed with the translated biweight. In dimension $k = 4$ and cut-off constants M and c set to have $\pi = 0.01$ and $r = 0.5$, the R-commands

```
k=4
r=0.5
p=0.01
c0=rhotranslatedconst(k,r,p,0.01,5)
M=sqrt(qchisq(1-p,df=k))-c0
constbetahattranslated(k,M,c0)
```

lead to $\lambda_t = 1.271367$. In comparison, the same constant for Tukey's biweight with cut-off constant set to have $r = 0.5$,

```
k=4
r=0.5
c0=rhoconst(k,r,0.01,5)
constbetahat(k,c0)
```

gives $\lambda = 1.250273$. This coincides with the remark in Rocke, 1996 [10], that the translated biweight leads to a much higher asymptotic rejection rate and a relatively small loss in efficiency.

4.5 Asymptotic covariance of the covariance CTBS estimator

When the data are generated from the distribution P of $(\mathbf{y}, \mathbf{X})$, which is such that $\mathbf{y} \mid \mathbf{X} \sim N(\mathbf{X}\boldsymbol{\beta}, \boldsymbol{\Sigma})$, where $\boldsymbol{\Sigma} = \mathbf{V}(\boldsymbol{\theta}) = \sum_{j=1}^{l} \theta_j \mathbf{L}_j$, then according to Corollary 5, the asymptotic covariance of $\sqrt{n}(\boldsymbol{\theta}_n - \boldsymbol{\theta}(P))$ is given by

$$2\sigma_{1t}\Big(\mathbf{L}^T\left(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1}\right)\mathbf{L}\Big)^{-1} + \sigma_{2t}\boldsymbol{\theta}(P)\boldsymbol{\theta}(P)^T,$$

and $\sqrt{n}(\mathrm{vec}(\mathbf{C}_n) - \mathrm{vec}(\boldsymbol{\Sigma}))$ is asymptotically normal with mean zero and covariance matrix

$$2\sigma_{1t}\mathbf{L}\Big(\mathbf{L}^T\left(\boldsymbol{\Sigma}^{-1} \otimes \boldsymbol{\Sigma}^{-1}\right)\mathbf{L}\Big)^{-1}\mathbf{L}^T + \sigma_{2t}\mathrm{vec}(\boldsymbol{\Sigma})\mathrm{vec}(\boldsymbol{\Sigma})^T,$$

where

$$\sigma_{1t} = \frac{k(k+2)\mathbb{E}_{\mathbf{0},\mathbf{I}_k}\left[u_t(\|\mathbf{z}\|)^2\|\mathbf{z}\|^4\right]}{\Big(\mathbb{E}_{0,\mathbf{I}_k}\left[\rho_t''(\|\mathbf{z}\|)\|\mathbf{z}\|^2 + (k+1)\rho_t'(\|\mathbf{z}\|)\|\mathbf{z}\|\right]\Big)^2}$$

$$\sigma_{2t} = -\frac{2}{k}\sigma_1 + \frac{4\mathbb{E}_{\mathbf{0},\mathbf{I}_k}\left[(\rho_t(\|\mathbf{z}\|) - b_0)^2\right]}{\Big(\mathbb{E}_{\mathbf{0},\mathbf{I}_k}\left[\rho_t'(\|\mathbf{z}\|)\|\mathbf{z}\|\right]\Big)^2}$$

Let us start with σ_1. We have from Section 4.2 that $u_t(d) = \psi_t(d)/d$ and

$$u_t(d)^2 d^4 = \begin{cases} d^4 & , 0 \le d \le M \\ d^4\left(1 - \dfrac{(d-M)^2}{c^2}\right)^4 & , M \le d \le M+c \\ 0 & , d \ge M+c \end{cases}$$

$$= \begin{cases} d^4 & , 0 \le d \le M \\ d^4\left(1 - \dfrac{4(d-M)^2}{c^2} + \dfrac{6(d-M)^4}{c^4} - \dfrac{4(d-M)^6}{c^6} + \dfrac{(d-M)^8}{c^8}\right) & , M \le d \le M+c \\ 0 & , d \ge M+c \end{cases}$$

$$= \begin{cases} d^4 & , 0 \le d \le M \\ d^4 - \dfrac{4d^4(d-M)^2}{c^2} + \dfrac{6d^4(d-M)^4}{c^4} - \dfrac{4d^4(d-M)^6}{c^6} + \dfrac{d^4(d-M)^8}{c^8} & , M \le d \le M+c \\ 0 & , d \ge M+c \end{cases}$$

and

$$\rho_t''(d)d^2 = \begin{cases} d^2 & , 0 \le d < M \\ d^2\left(1 + M^4/c^4 - 2M^2/c^2\right) & \\ \quad + d^3(8M/c^2 - 8M^3/c^4) & \\ \quad + d^4(18M^2/c^4 - 6/c^2) & , M \le d \le M+c \\ \quad - 16Md^5/c^4 & \\ \quad + 5d^6/c^4 & \\ 0 & , d > M+c \end{cases}$$

whereas

$$\rho_t'(d)d = \begin{cases} d^2 & ,0 \le d < M \\ d^2\left(1 - \dfrac{(d-M)^2}{c^2}\right)^2 & ,M \le d \le M+c \\ 0 & ,d > M+c \end{cases}$$

$$= \begin{cases} d^2 & ,0 \le d < M \\ d^2(1 + M^4/c^4 - 2M^2/c^2) & \\ \quad + d^3(4M/c^2 - 4M^3/c^4) & \\ \quad + d^4(6M^2/c^4 - 2/c^2) & ,M \le d \le M+c \\ \quad - 4Md^5/c^4 & \\ \quad + d^6/c^4 & \\ 0 & ,d > M+c \end{cases}$$

This leads to the following R-function

```
sigma1t=function(k,M,c0){
  # determines the constant sigma1 in the scalar in the
  # limiting covariance of thetahat with translated biweight
  # k=dimension of multivariate normal
  # M= first parameter of translated biweight
  # c0= first parameter of translated biweight

  # Compute E[u_t(d)^2d^4]
  term1=moment(4,k,M)+
    momenttranslated(4,k,M,M+c0)-
    momentcentered(4,2,k,M,M,M+c0)*(4/(c0^2))+
    momentcentered(4,4,k,M,M,M+c0)*(6/(c0^4))-
    momentcentered(4,6,k,M,M,M+c0)*(4/(c0^6))+
    momentcentered(4,8,k,M,M,M+c0)/(c0^8)

  # Compute E[rho_t''(d)d^2]
  term2=moment(2,k,M)+
    momenttranslated(2,k,M,M+c0)*(1+M^4/(c0^4)-2*M^2/(c0^2))+
    momenttranslated(3,k,M,M+c0)*(8*M/(c0^2)-8*M^3/(c0^4))+
    momenttranslated(4,k,M,M+c0)*(18*M^2/c0^4-6/(c0^2))-
    momenttranslated(5,k,M,M+c0)*16*M/(c0^4)+
    momenttranslated(6,k,M,M+c0)*5/(c0^4)

  # Compute E[rho_t'(d)d]
  term3=moment(2,k,M)+
    momenttranslated(2,k,M,M+c0)*(1+M^4/(c0^4)-2*M^2/(c0^2))+
    momenttranslated(3,k,M,M+c0)*(4*M/(c0^2)-4*M^3/(c0^4))+
    momenttranslated(4,k,M,M+c0)*(6*M^2/c0^4-2/(c0^2))-
    momenttranslated(5,k,M,M+c0)*4*M/(c0^4)+
    momenttranslated(6,k,M,M+c0)/(c0^4)

  num=k*(k+2)*term1
  denum=(term2+(k+1)*term3)^2
  return(num/denum)
}
```

We continue with

$$\sigma_{2t} = -\frac{2}{k}\sigma_{1t} + \frac{4\mathbb{E}_{\mathbf{0},\mathbf{I}_k}[(\rho_t(\|\mathbf{z}\|) - b_0)^2]}{\left(\mathbb{E}_{\mathbf{0},\mathbf{I}_k}\left[\rho_t'(\|\mathbf{z}\|)\|\mathbf{z}\|\right]\right)^2}.$$

Note that we have chosen

$$b_{0t} = \mathbb{E}_{\mathbf{0},\mathbf{I}_k}\rho_t(\|\mathbf{z}\|).$$

This means that

$$\mathbb{E}_{\mathbf{0},\mathbf{I}_k}[(\rho_t(\|\mathbf{z}\|) - b_0)^2] = \mathbb{E}_{\mathbf{0},\mathbf{I}_k}[\rho_t(\|\mathbf{z}\|)^2] - (\mathbb{E}_{\mathbf{0},\mathbf{I}_k}\rho(\|\mathbf{z}\|))^2.$$

We have

$$\rho_t(y)^2 = \begin{cases} d^4/4 & ,0 \le d \le M \\ a_{1t}^2 + a_{2t}^2 + a_{3t}^2 d^4 + a_{4t}^2 d^6 + a_{5t}^2 d^8 + a_{6t}^2 d^{10} + a_{7t}^2 d^{12} & \\ \quad + 2a_{1t}a_{2t} + 2a_{1t}a_{3t}d^2 + 2a_{1t}a_{4t}d^3 + 2a_{1t}a_{5t}d^4 + 2a_{1t}a_{6t}d^5 + 2a_{1t}a_{7t}d^6 & \\ \quad + 2a_{2t}a_{3t}d^2 + 2a_{2t}a_{4t}d^3 + 2a_{2t}a_{5t}d^4 + 2a_{2t}a_{6t}d^5 + 2a_{2t}a_{7t}d^6 & \\ \quad + 2a_{3t}a_{4t}d^5 + 2a_{3t}a_{5t}d^6 + 2a_{3t}a_{6t}d^7 + 2a_{3t}a_{7t}d^8 & ,M \le d \le M+c \\ \quad + 2a_{4t}a_{5t}d^7 + 2a_{4t}a_{6t}d^8 + 2a_{4t}a_{7t}d^9 & \\ \quad + 2a_{5t}a_{6t}d^9 + 2a_{5t}a_{7t}d^{10} & \\ \quad + 2a_{6t}a_{7t}d^{11} & \\ a_{8t}^2 & ,d \ge M+c. \end{cases}$$

```
sigma2t=function(k,M,c0){
  # determines the constant sigma2 in the scalar in the
  # limiting covariance of thetahat of the translated biweight
  # k=dimension of multivariate normal
  # M=first parameter of translated biweight
  # c0=second parameter of translated biweight

  a1t=M^2/2
  a2t=-M^2*(M^4-5*M^2*c0^2+15*c0^4)/(30*c0^4)
  a3t=1/2+M^4/(2*c0^4)-M^2/c0^2
  a4t=4*M/(3*c0^2)-4*M^3/(3*c0^4)
  a5t=3*M^2/(2*c0^4)-1/(2*c0^2)
  a6t=-4*M/(5*c0^4)
  a7t=1/(6*c0^4)
  a8t=M^2/2+c0*(5*c0+16*M)/30

  # Compute E[rho_t(d)^2]
  term1=moment(4,k,M)/4+
    a1t^2*momenttranslated(0,k,M,M+c0)+
    a2t^2*momenttranslated(0,k,M,M+c0)+
    a3t^2*momenttranslated(4,k,M,M+c0)+
    a4t^2*momenttranslated(6,k,M,M+c0)+
    a5t^2*momenttranslated(8,k,M,M+c0)+
    a6t^2*momenttranslated(10,k,M,M+c0)+
    a7t^2*momenttranslated(12,k,M,M+c0)+
    2*a1t*a2t*momenttranslated(0,k,M,M+c0)+
    2*a1t*a3t*momenttranslated(2,k,M,M+c0)+
    2*a1t*a4t*momenttranslated(3,k,M,M+c0)+
    2*a1t*a5t*momenttranslated(4,k,M,M+c0)+
    2*a1t*a6t*momenttranslated(5,k,M,M+c0)+
```

```
    2*a1t*a7t*momenttranslated(6,k,M,M+c0)+
    2*a2t*a3t*momenttranslated(2,k,M,M+c0)+
    2*a2t*a4t*momenttranslated(3,k,M,M+c0)+
    2*a2t*a5t*momenttranslated(4,k,M,M+c0)+
    2*a2t*a6t*momenttranslated(5,k,M,M+c0)+
    2*a2t*a7t*momenttranslated(6,k,M,M+c0)+
    2*a3t*a4t*momenttranslated(5,k,M,M+c0)+
    2*a3t*a5t*momenttranslated(6,k,M,M+c0)+
    2*a3t*a6t*momenttranslated(7,k,M,M+c0)+
    2*a3t*a7t*momenttranslated(8,k,M,M+c0)+
    2*a4t*a5t*momenttranslated(7,k,M,M+c0)+
    2*a4t*a6t*momenttranslated(8,k,M,M+c0)+
    2*a4t*a7t*momenttranslated(9,k,M,M+c0)+
    2*a5t*a6t*momenttranslated(9,k,M,M+c0)+
    2*a5t*a7t*momenttranslated(10,k,M,M+c0)+
    2*a6t*a7t*momenttranslated(11,k,M,M+c0)+
    a8t^2*(1-moment(0,k,M+c0))

  # Compute E[rho_t'(d)d]
  term2=moment(2,k,M)+
    momenttranslated(2,k,M,M+c0)*(1+M^4/(c0^4)-2*M^2/(c0^2))+
    momenttranslated(3,k,M,M+c0)*(4*M/(c0^2)-4*M^3/(c0^4))+
    momenttranslated(4,k,M,M+c0)*(6*M^2/c0^4-2/(c0^2))-
    momenttranslated(5,k,M,M+c0)*4*M/(c0^4)+
    momenttranslated(6,k,M,M+c0)/(c0^4)

  # Compute E[rho_t(d)]
  term3=expecrhotranslated(k,M,c0)

  num=4*(term1-(term3)^2)
  denum=(term2)^2
  return(-(2/k)*sigma1t(k,M,c0)+num/denum)
}
```

Both functions can be combined in Tyler's efficiency scalar η. As mentioned in [7], $\eta = 2\sigma_1 + \sigma_2$, when $k = 1$ and $\eta = \sigma_1$, when $k \geq 2$.

```
etat=function(k,M,c0){
  # computes the efficiency index of Tyler for translated biweight
  ifelse(k==1,1,0)*(2*sigma1t(k,M,c0)+sigma2t(k,M,c0))+
    ifelse(k>1,1,0)*sigma1t(k,M,c0)
}
```

4.6 Example in Copt and Victoria-Feser, 2016 [2]

Let us return to the example in Copt and Victoria-Feser [2]. In this case, the dimension is $k = 4$. The asymptotic rejection rate was chosen to be $\pi = 0.01$ and the breakdown point $\epsilon = 0.5$. The R-commands

```
k=4
p=0.01
r=0.5
c0t=rhotranslatedconst(k,r,p,0.01,5)
M=sqrt(qchisq(1-p,df=k))-c0t
```

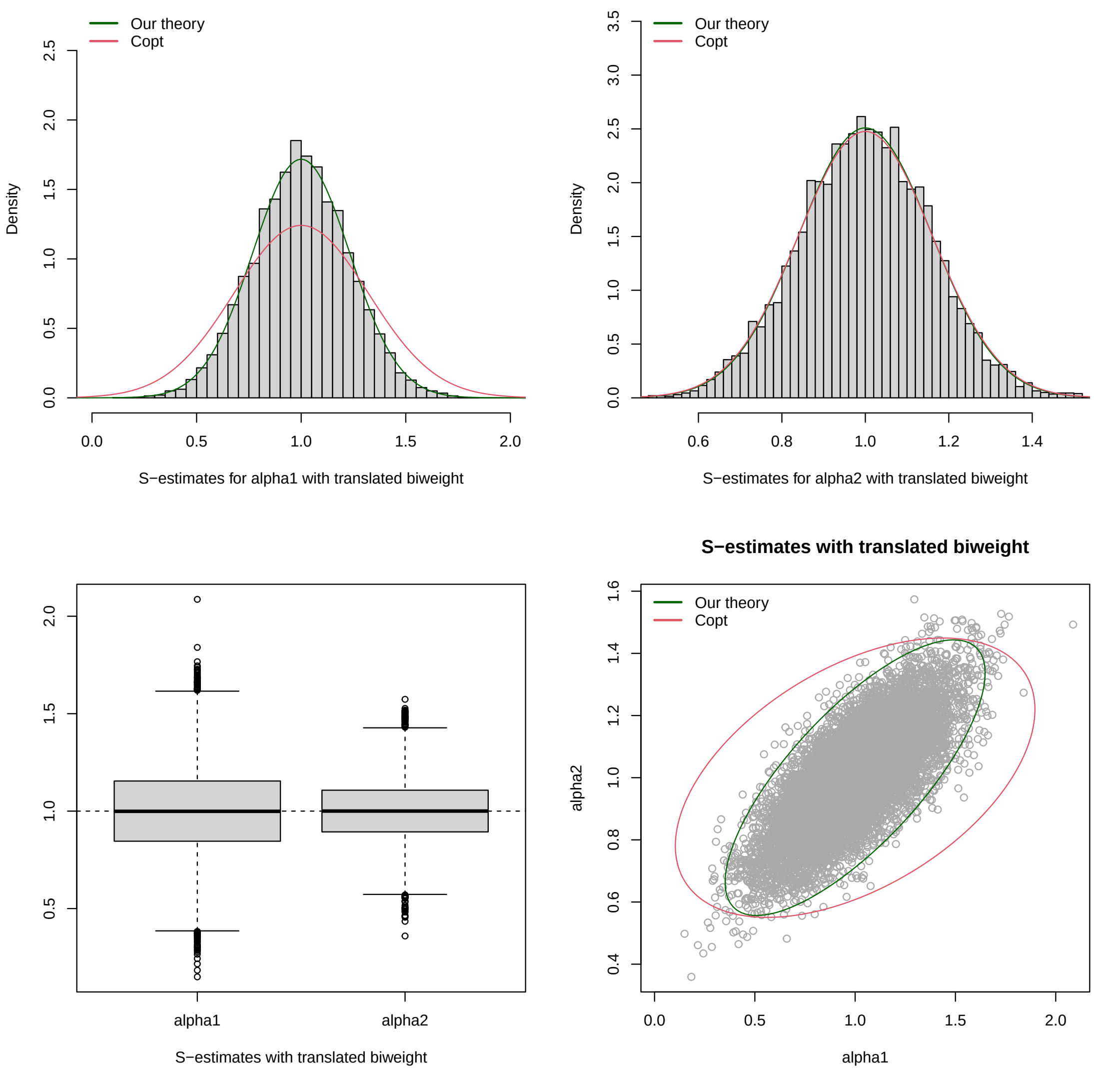

Figure 5: Empirical and asymptotic distributions of $\boldsymbol{\beta}_n$ computed with translated biweight.

lead to $c = 2.262801$ and $M = 1.380920$.

The asymptotic variances for $\widehat{\boldsymbol{\beta}}_n$ are coded

```
# Expression from our theory
varbetatranslated=constbetahattranslated(k,M,c0t)*
  solve(t(X)%*%solve(Vtheta)%*%X)

# Expression from Copt & Victoria-Feser (2006)
varbetaCopttranslated=constbetahattranslated(k,M,c0t)*
  solve(t(X)%*%X) %*% t(X) %*% Vtheta %*% X %*% solve(t(X)%*%X)
```

This leads to

$$\texttt{varbetatranslated} = \begin{pmatrix} 5.398737 & 2.804305 \\ 2.804305 & 2.528433 \end{pmatrix}$$

and

$$\texttt{varbetaCopttranslated} = \begin{pmatrix} 10.329855 & 2.529796 \\ 2.529796 & 2.595812 \end{pmatrix}.$$

As expected, this does not differ much from the asymptotic variances found with Tukey's biweight

$$\texttt{varbeta} = \begin{pmatrix} 5.309162 & 2.757777 \\ 2.757777 & 2.486482 \end{pmatrix}$$

and

$$\texttt{varbetaCopt} = \begin{pmatrix} 10.158464 & 2.487822 \\ 2.487822 & 2.552743 \end{pmatrix}.$$

The results of a simulation with 10 000 replications of the iteration described in Figure 1 conducted with the translated biweight, can be seen in Figure 5. Again, we can see a clear match between the empirical (co)variances with our theoretical results (in green), which once more deviates completely from the predictions by Copt and Victoria-Feser [2] (in red).

The asymptotic variances for $\widehat{\boldsymbol{\theta}}_n$ are coded

```
# Expression from our theory
varthetatranslated=(2*sigma1t(k,M,c0t)*
            solve(t(L)%*%(solve(Vtheta)%x%solve(Vtheta))%*%L)+
            sigma2t(k,M,c0t)*theta%*%t(theta))

# Expression from Copt & Victoria-Feser (2006)
varthetaCopttranslated=2*sigma1t(k,M,c0t)*
  solve((t(L)%*%L))%*%t(L)%*%(Vtheta%x%Vtheta)%*%L%*%solve((t(L)%*%L))+
  sigma2t(k,M,c0t)*theta%*%t(theta)
```

This leads to

$$\texttt{varthetatranslated} = \begin{pmatrix} 4.7012730 & -0.4686337 \\ -0.4686337 & 0.8238429 \end{pmatrix}$$

and

$$\texttt{varthetaCopttranslated} = \begin{pmatrix} 5.4730131 & -0.4050425 \\ -0.4050425 & 2.9713930 \end{pmatrix}.$$

As expected, this does not differ much from the asymptotic variances found with Tukey's biweight

$$\texttt{vartheta} = \begin{pmatrix} 4.3569335 & -0.3881412 \\ -0.3881412 & 0.7981275 \end{pmatrix}$$

and

$$\texttt{varthetaCopt} = \begin{pmatrix} 5.0652567 & -0.3297755 \\ -0.3297755 & 2.7692047 \end{pmatrix}.$$

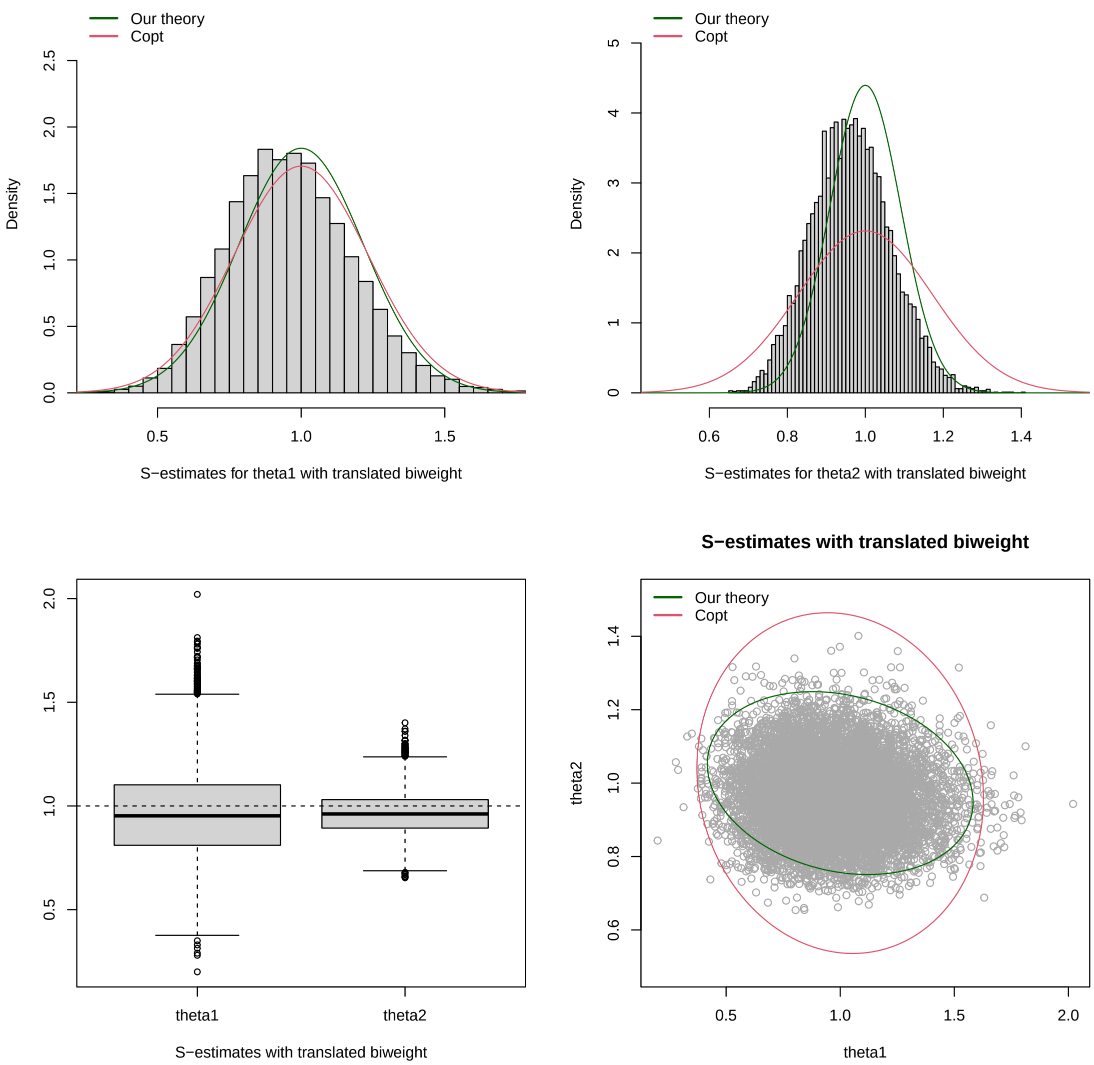

Figure 6: Empirical and asymptotic distributions of $\boldsymbol{\theta}_n$ computed with translated biweight.

The results of a simulation with 10 000 replications of the iteration described in Figure 1 conducted with the translated biweight, can be seen in Figure 6. Again, we can see a clear match between the empirical (co)variances with our theoretical results (in green), which once more deviates completely from the predictions by Copt and Victoria-Feser [2] (in red).

For some reason, the bias in the S-estimates for $\widehat{\theta}_{1n}$ and $\widehat{\theta}_{2n}$ computed with the translated biweight is more prominent than for those computed with Tukey's biweight (see Figure 3).

5 Behavior under contamination

After checking the theoretical results by means of a simulation, it is also of interest how the estimators are affected by contamination. This section discusses different possible sources of contamination and the effect on the S-estimates. As a basis I have taken the model from Copt and Victoria-Feser [2]:

$$\mathbf{y}_i = \mathbf{X}_i\boldsymbol{\beta} + \gamma_i\mathbf{Z} + \boldsymbol{\epsilon}_i, \quad i = 1, \ldots, n.$$

Here,

$$\mathbf{X}_i = [\mathbf{1}_4 \quad \mathbf{x}_i] \quad \text{and} \quad \boldsymbol{\beta} = \begin{pmatrix} 1 \\ 1 \end{pmatrix},$$

where $\mathbf{x}_1, \ldots, \mathbf{x}_n$ are independent with $\mathbf{x}_i \sim N(\mathbf{0}, \mathbf{I}_4)$, where $\beta_1, \ldots, \beta_n$ are independent with

$$\beta_i \sim N(0, \sigma_\beta^2) \quad \text{with} \quad \sigma_\beta^2 = 1, \quad \text{and} \quad \mathbf{Z} = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix},$$

and where $\boldsymbol{\epsilon}_1, \ldots, \boldsymbol{\epsilon}_n$ are independent with

$$\boldsymbol{\epsilon}_i \sim N(\mathbf{0}, \sigma_\epsilon^2\mathbf{I}_4), \quad \text{with} \quad \sigma_\epsilon^2 = 1.$$

Note that the design matrices $\mathbf{X}_i$ differ for each observation and are not orthogonalized. The parameters of interest are $\boldsymbol{\beta} = (\beta_1, \beta_2) = (1, 1)$ and $\boldsymbol{\theta} = (\sigma_\beta^2, \sigma_\epsilon^2) = (1, 1)$.

In order to investigate the effect on the S-estimates, I have performed different simulations generating samples of size $n = 100$ from a contaminated distribution. I have investigated different sources of contamination and different proportions, and in each simulation I did 100 replications. For the S-estimates, I used Tukey's biweight with cut-off constant $c = 4.096567$. This corresponds to breakdown point 0.5 in dimension $k = 4$. Furthermore, I use Rocke's translated biweight with cut-off constants $M = 1.38092$ and $c = 2.262801$. This corresponds to an asymptotic rejection rate $\pi = 0.01$ and breakdown point 0.5 in dimension $k = 4$. Both rho-functions and related derivatives can be seen in Figure 4.

5.1 Outliers in the x-space

I first considered the following (location) contamination in the $\mathbf{x}$-space

$$\mathbf{x}_i \sim (1-r)N(\mathbf{0}, \mathbf{I}_4) + rN(\mathbf{x}^*, \mathbf{I}_4), \quad \mathbf{x}^* = \begin{pmatrix} 10 \\ 0 \\ 0 \\ 0 \end{pmatrix},$$

where $r \in [0, (n-k)/(2n) = [0, 0.48)$ is the proportion of contamination. So, this type of contamination generates a large point mass in the first coordinate of the second column of each $\mathbf{X}_i$.

The effect on the estimates is more or less the same if we vary r from 0.1 to 0.45. All the S-estimates remain stable. This can be seen in Figure 7. Each boxplot represents 100 estimates based on a generated sample from the contaminated distribution.

An even better illustration can be seen in Figure 8. For the data from the last replication in the simulation run, it shows the 2-dimensional projections of $\mathbf{x}_1, \ldots, \mathbf{x}_n$. The contaminated points (red points) clearly stick out as outliers in the x_1-dimension. In each plot one can also see the 97,5% contour line corresponding to the uncontaminated distribution of $\mathbf{x}$ (green) and the one estimated by the MCD (red).

Finally, I looked at the plot proposed by Rousseeuw and van Zomeren, 1990 [12], which combines outlyingness in the $\mathbf{x}$-space with outlyingness in the $\mathbf{y}$-space. For the data from the last

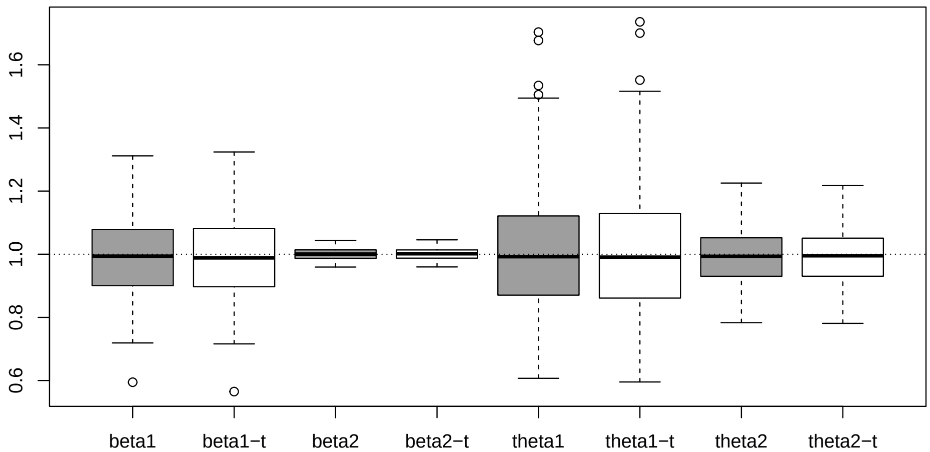

Figure 7: Boxplot of S-estimates; 45% contamination in $\mathbf{x}$-space.

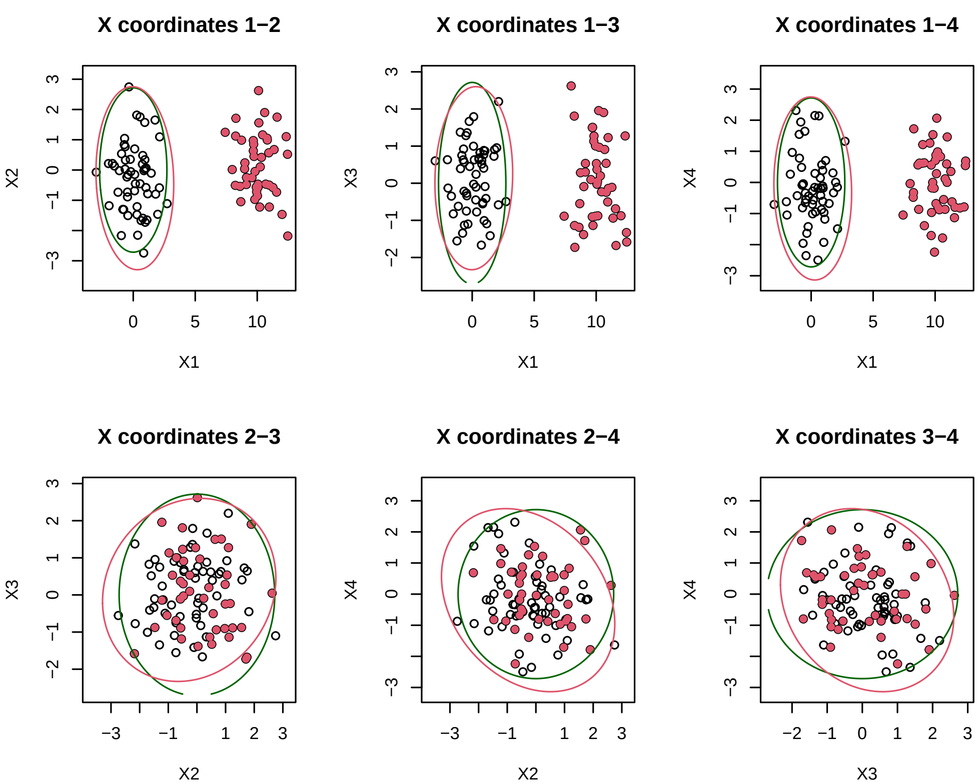

Figure 8: 2d projections of the $\mathbf{x}_i$; 45% contamination in $\mathbf{x}$-space.

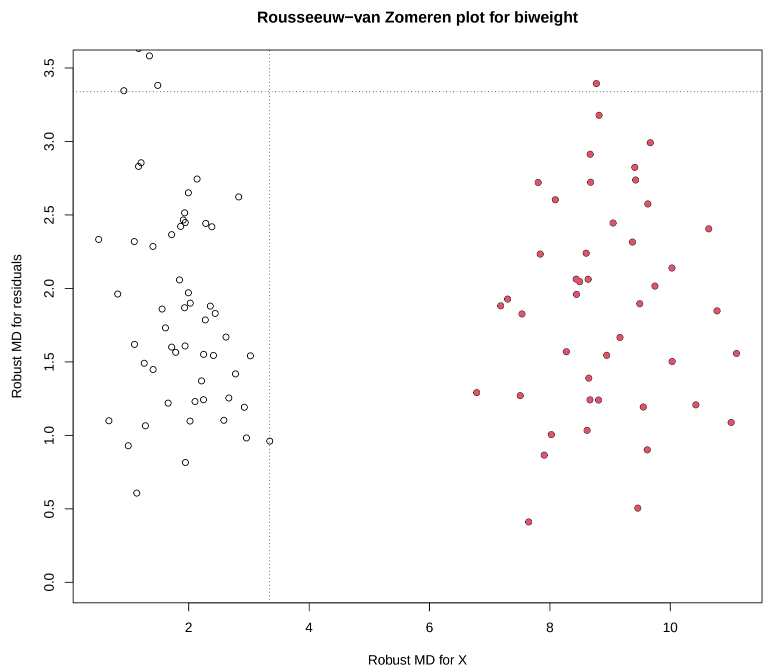

Figure 9: Rousseeuw-van Zomeren plot; 45% contamination in $\mathbf{x}$-space.

replication in the simulation run, it can be seen in see Figure 9. The plot is constructed as follows. On the horizontal axis one plots the robust Mahalanobis distances in the $\mathbf{x}$-space

$$\mathrm{MD}_{\mathbf{x},i} = \sqrt{(\mathbf{x}_i - \widehat{\boldsymbol{\mu}}_x)^T \widehat{\boldsymbol{\Sigma}}_x^{-1} (\mathbf{x}_i - \widehat{\boldsymbol{\mu}}_x)}$$

where $(\widehat{\boldsymbol{\mu}}_x, \widehat{\boldsymbol{\Sigma}}_x)$ denote the MCD estimates for location and covariance. The vertical dotted line is at the square root of the 97,5% quantile of the $\chi^2(4)$ distribution. On the vertical axis one plots the robust Mahalanobis distances of the residuals

$$\mathrm{MD}_{\mathbf{r},i} = \sqrt{\mathbf{r}_i^T \widehat{\mathbf{V}}^{-1} \mathbf{r}_i}$$

where

$$\mathbf{r}_i = \mathbf{y}_i - \mathbf{X}_i \widehat{\boldsymbol{\alpha}}, \quad i = 1, \ldots, n,$$

and $\widehat{\mathbf{V}} = \mathbf{V}(\widehat{\boldsymbol{\theta}})$ is the estimated covariance of $\mathbf{u}_i = \mathbf{y}_i - \mathbf{X}_i \boldsymbol{\beta}$. The horizontal dotted line is at the square root of the 97,5% quantile of the $\chi^2(4)$ distribution. As can be seen from the plot the contaminated points (in red) stick out on the horizontal axis, and are detected as outliers in the $\mathbf{x}$-space, but do not stick out on the vertical axis.

As expected the 50% breakdown S-estimates for $\boldsymbol{\beta}$ are not much affected, but it is surprising to see that even 45% contamination does not have a large effect on the estimates. Finally, I can report that contamination in all directions of the $\mathbf{x}$-space, i.e., $\mathbf{x}_i^* = (10, 10, 10, 10)^T$, leads to similar plots. Furthermore, scale contamination of the form

$$\mathbf{x}_i \sim (1-r)N(\mathbf{0}, \mathbf{I}_4) + rN(\mathbf{0}, \mathbf{R}), \quad \mathbf{R} = \begin{pmatrix} 11 & 0 & 0 & 0 \\ 0 & 11 & 0 & 0 \\ 0 & 0 & 11 & 0 \\ 0 & 0 & 0 & 11 \end{pmatrix}$$

lead to similar plots.

5.2 Outliers in the fixed effects

Another possible source of contamination are the fixed effects. I have investigated the effect of a location contamination

$$\beta_i \sim (1-r)N(0,1) + rN(10,1).$$

This will produce outlying observations $\mathbf{y}_i$. For $r = 0.2$ this is illustrated in Figure 10. For the data from the last replication, it shows that 2-dimensional projections of the $\mathbf{y}_i$ in the $\mathbf{y}$-space. Two 97,5% contour lines have been added to the plot: (i) one with center $\overline{\mathbf{X}}\boldsymbol{\alpha}$ and covariance $\mathbf{V}(\boldsymbol{\theta})$ (green) representing the theoretical distributions $N(\mathbf{X}_i\boldsymbol{\alpha}, \mathbf{V}(\boldsymbol{\theta}))$ of $\mathbf{y}_i \mid \mathbf{X}_i$, and (ii) one with estimated center $\overline{\mathbf{X}}\widehat{\boldsymbol{\alpha}}$ and estimated covariance $\mathbf{V}(\widehat{\boldsymbol{\theta}})$ (red), where

$$\overline{\mathbf{X}} = \frac{1}{n}\sum_{i=1}^{n} \mathbf{X}_i = [\mathbf{1}_4 \quad \overline{\mathbf{x}}]$$

Note that the estimated covariance is somewhat larger. This is due to the fact that the S-estimates for $\boldsymbol{\theta}$ are affected by the 20% contamination. Note that despite the larger values for $\widehat{\theta}_1$ and $\widehat{\theta}_2$, the direction of the contour lines remains unaffected. This can be explained by the fact that $\mathbf{V}(\boldsymbol{\theta})$ is linear and that the effect on both $\widehat{\theta}_1$ and $\widehat{\theta}_2$ is of the same magnitude.

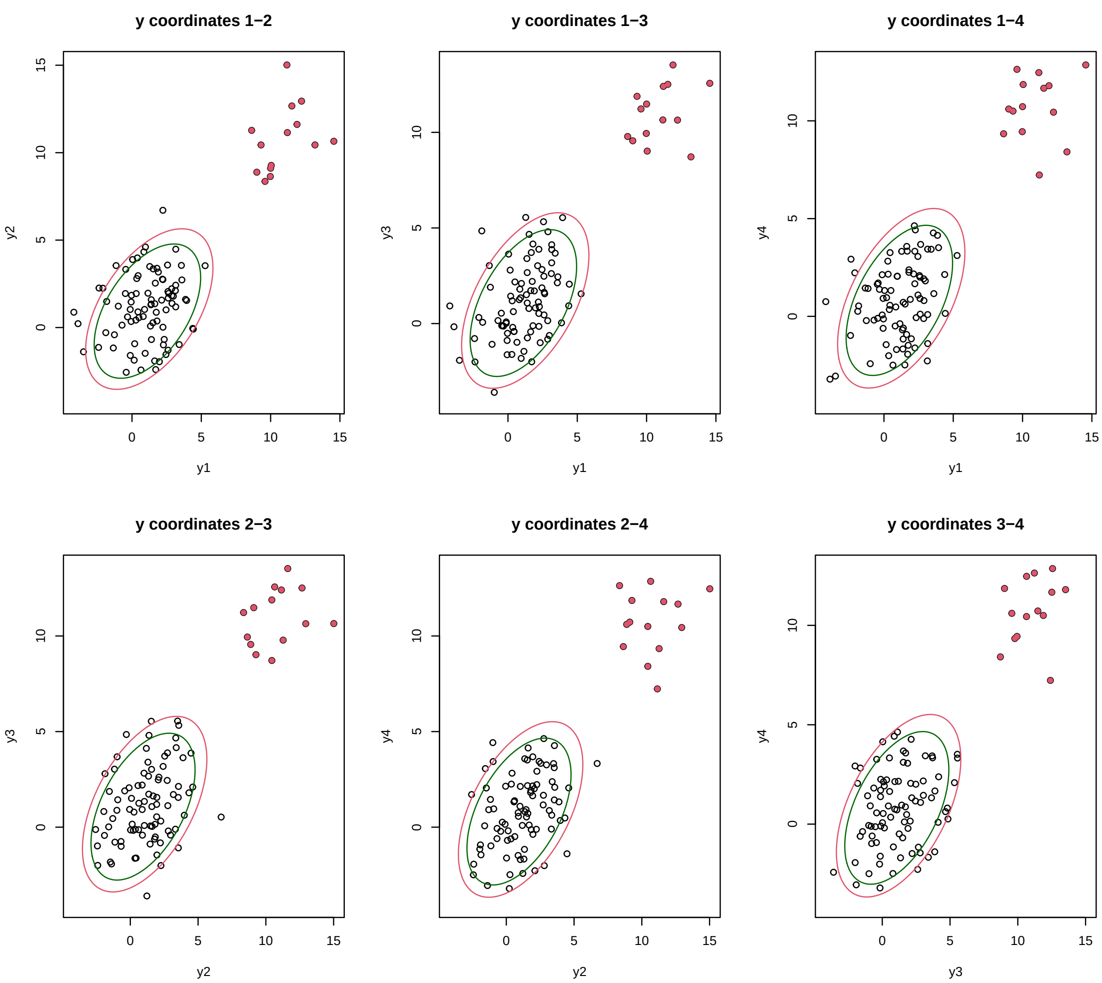

Figure 10: 2d projections of $\mathbf{y}_i$; 20% location contamination in γ_i.

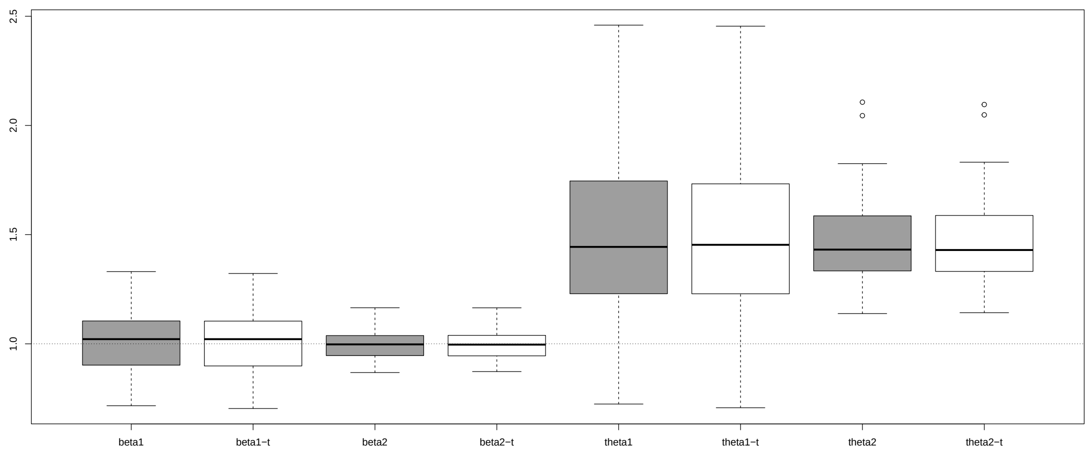

Figure 11: Boxplots of S-estimates; 20% location contamination in γ_i.

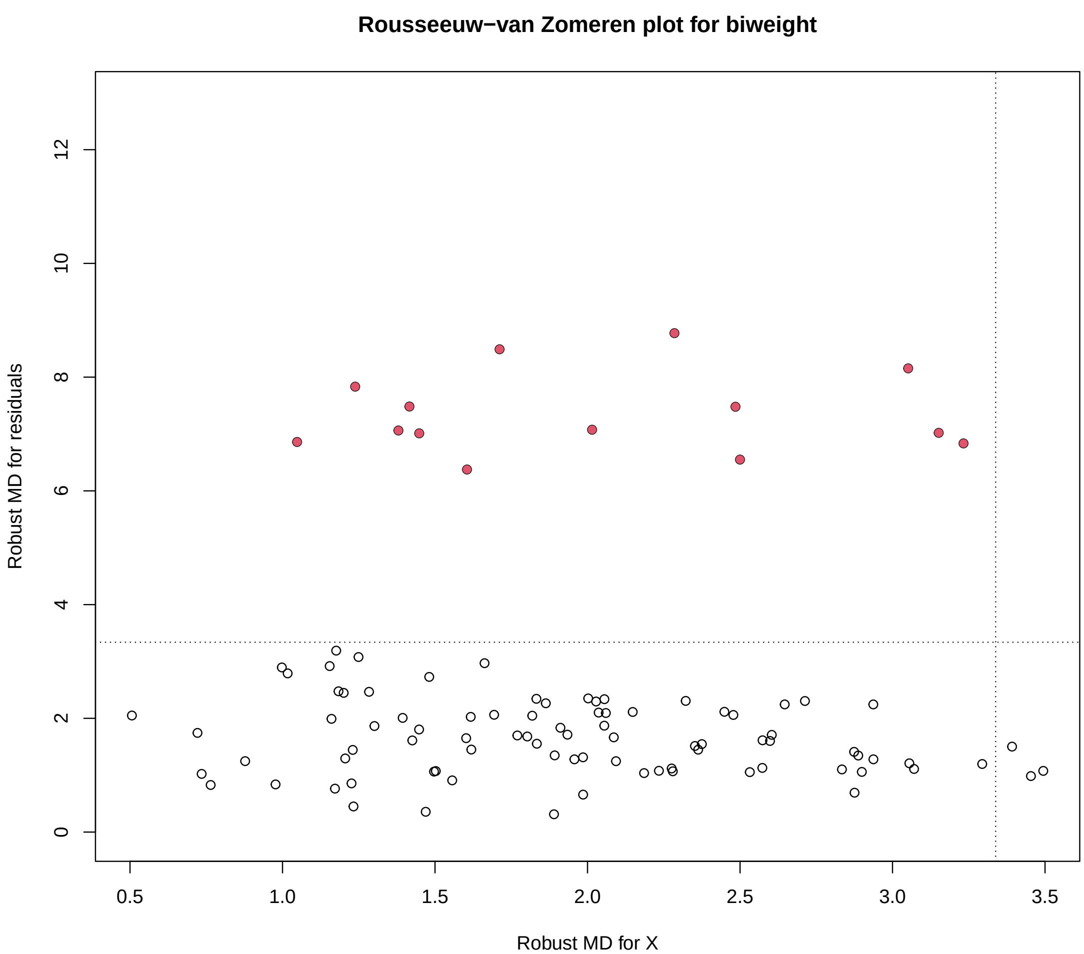

Figure 12: Rousseeuw-van Zomeren plot; 20% location contamination in γ_i.

The fact that the effect on the S-estimates for $\boldsymbol{\theta}$ is of the same magnitude is also illustrated in Figure 11. Note that the S-estimates for $\boldsymbol{\beta}$ remain unaffected.

Finally, because of the small effect on the S-estimates for $\boldsymbol{\theta}$ the contaminated points (in red) are still outlying in the $\mathbf{y}$-space. For this reason they are still detected by the Rousseeuw-van Zomeren plot, as being outlying points in the $\mathbf{y}$-space and not in the $\mathbf{x}$-space. This can be seen in Figure 12.

Increasing the amount of (location) contamination to 30%-40% leads to both an effect on the S-estimate for β_1 as well as to extreme large values for $\widehat{\theta}_1 = \widehat{\sigma}^2_\beta$. This yields large outliers in the $\mathbf{y}$-space. As expected, this will also introduce a masking effect in the Rousseeuw-van Zomeren plot, in which the outliers are no longer identified.

I also considered (scale) contamination in the variance of the fixed effects:

$$\beta_i \sim (1-r)N(0,1) + rN(0,11).$$

One would expect that the contaminated points will stick out much less than in Figure 10. Indeed, this is the case as one can see in Figure 13. Also note that the direction of the estimated contour lines (red) is a bit different than the one based on the theoretical covariance (green). This can be explained by the fact that with this type of contamination, only the S-estimate for $\theta_1 = \sigma^2_\beta$ is affected.

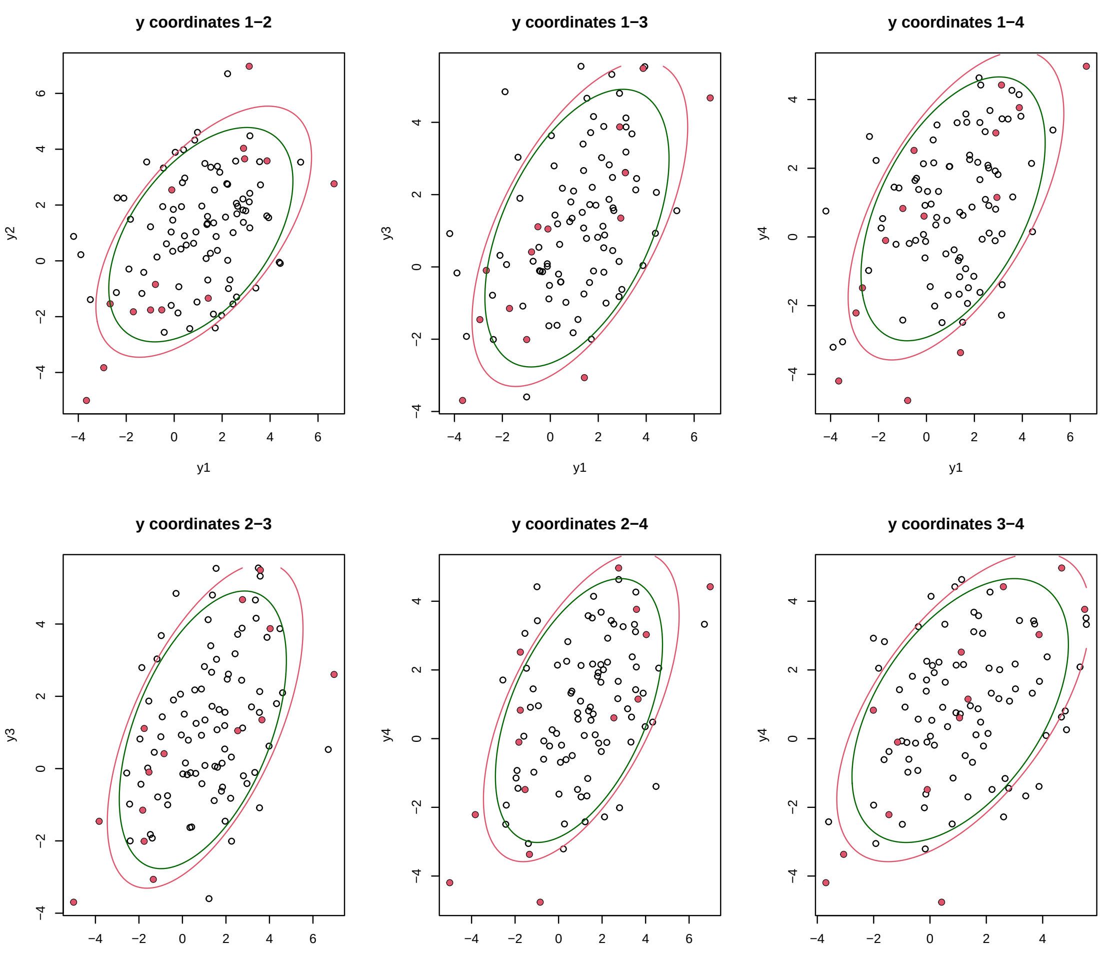

Figure 13: 2d projections of $\mathbf{y}_i$; 20% scale contamination in γ_i.

The different effect on the S-estimates for $\theta_1 = \sigma^2_\beta$ and $\theta_2 = \sigma^2_\epsilon$ is illustrated in Figure 14. Note that with 20% of this type of contamination, the S-estimates for $\boldsymbol{\beta}$ are unaffected.

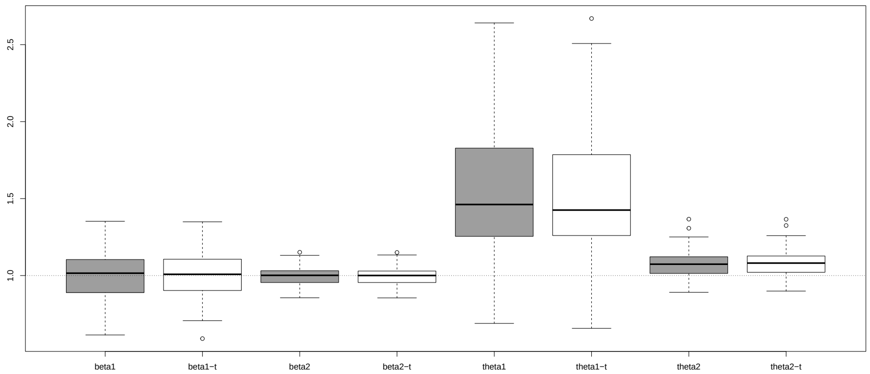

Figure 14: Boxplots of S-estimates; 20% scale contamination in γ_i.

Finally, as can also be seen from Figure 13, contamination in the variance of the random effects leads to a masking effect. This is illustrated in Figure 15, in which the contaminated points (in red) are identified as outliers.

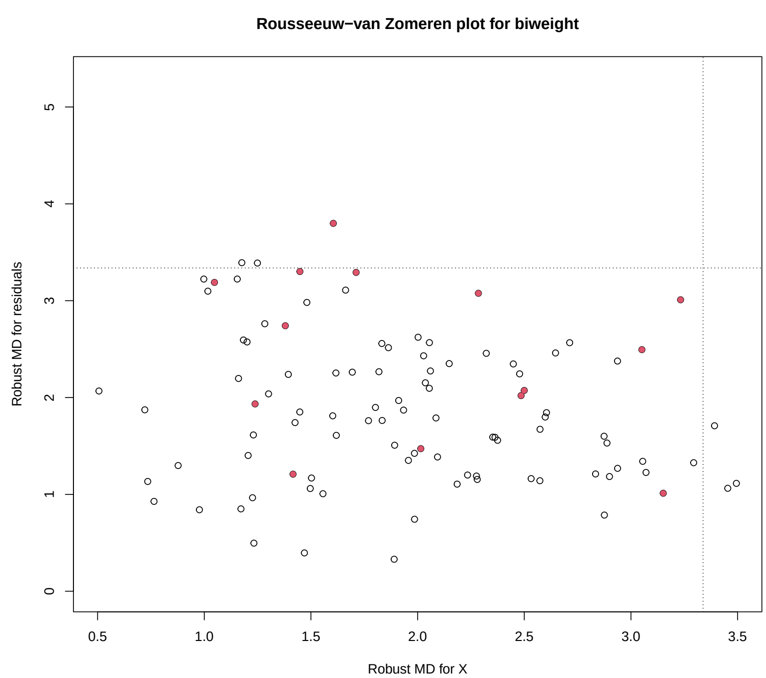

Figure 15: Rousseeuw-van Zomeren plot; 20% scale contamination in γ_i.

Larger amounts of scale contamination up to 30%-40% leads to similar plots.

5.3 Outliers in the measurement error

The last source of contamination I investigated is in the measurement error. I first investigated location contamination

$$\boldsymbol{\epsilon}_i \sim (1-r)N(\mathbf{0}, \mathbf{I}_4) + rN(\boldsymbol{\epsilon}^*, \mathbf{I}_4), \quad \boldsymbol{\epsilon}^* = \begin{pmatrix} 10 \\ 0 \\ 0 \\ 0 \end{pmatrix},$$

The effect is similar to the location contamination in the $\mathbf{x}$-space and in the random effects, see Figures 8 and 10. The effect on the observations is illustrated in Figure 16. Note that the estimated contour lines (red) indicate that the both $\widehat{\theta}_1$ and $\widehat{\theta}_2$ are affected. As the direction of the contour lines are comparable to the ones based on the theoretical covariance, this suggests that $\widehat{\theta}_1$ and $\widehat{\theta}_2$ are both too large by the same order of magnitude.

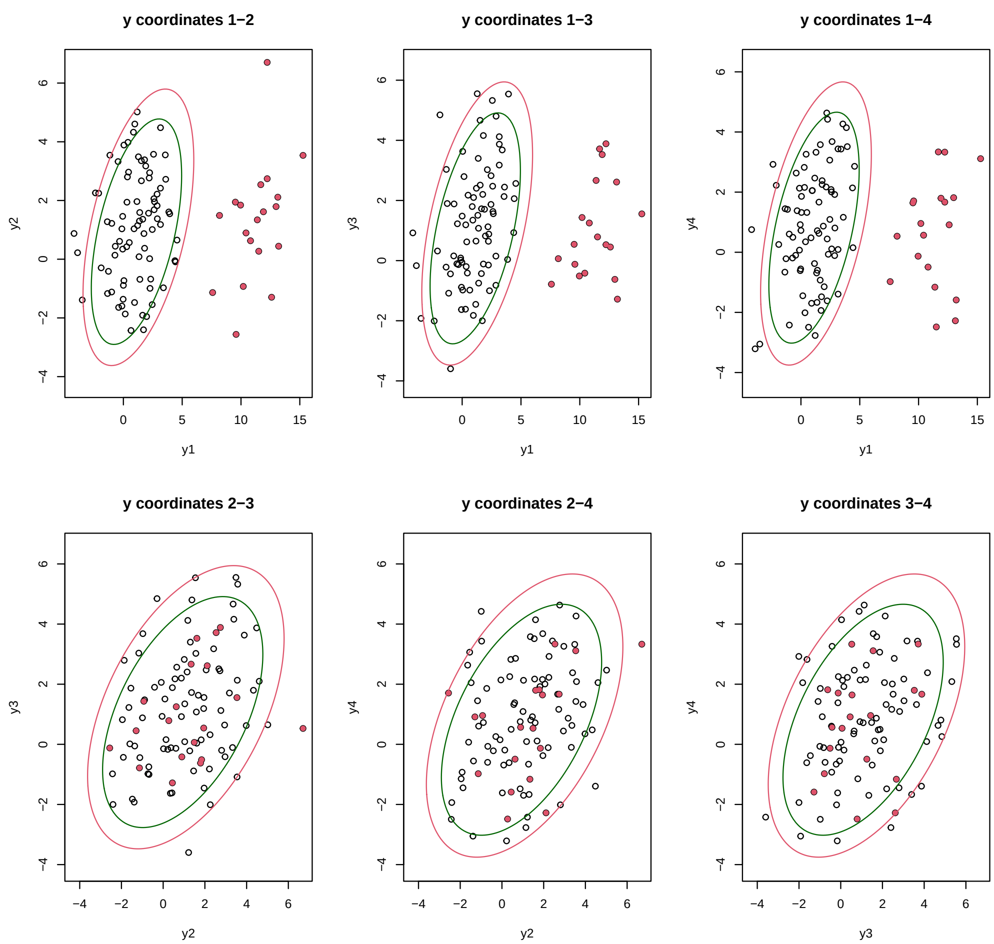

Figure 16: 2d projections of $\mathbf{y}_i$; 20% location contamination in measurement error.

This is indeed visible in Figure 17. Note the S-estimates for $\boldsymbol{\beta}$ are unaffected by 20% of this type of contamination. The effect on the S-estimates is comparable that of 20% of location contamination, see Figures 16 and 17.

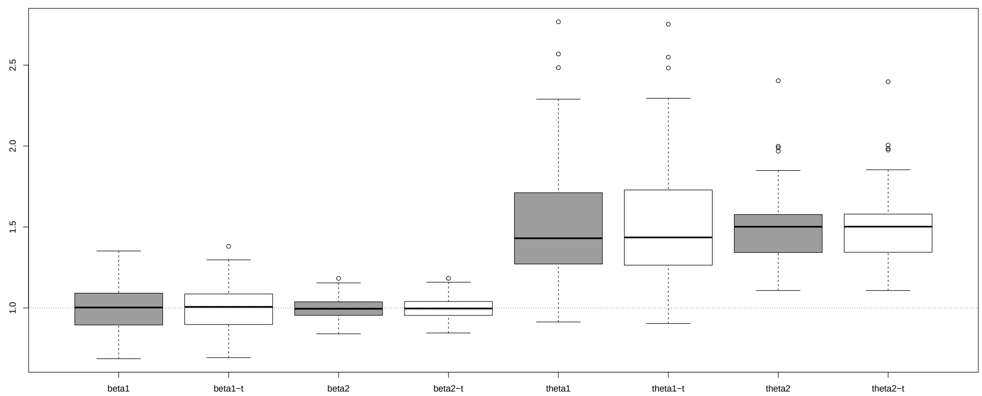

Figure 17: Boxplots of S-estimates; 20% location contamination in measurements.

As expected, with 20% contamination the contaminated points (red) are detected by the Rousseeuw-van Zomeren plot, as being outliers in the **y**-space. see Figure 18.

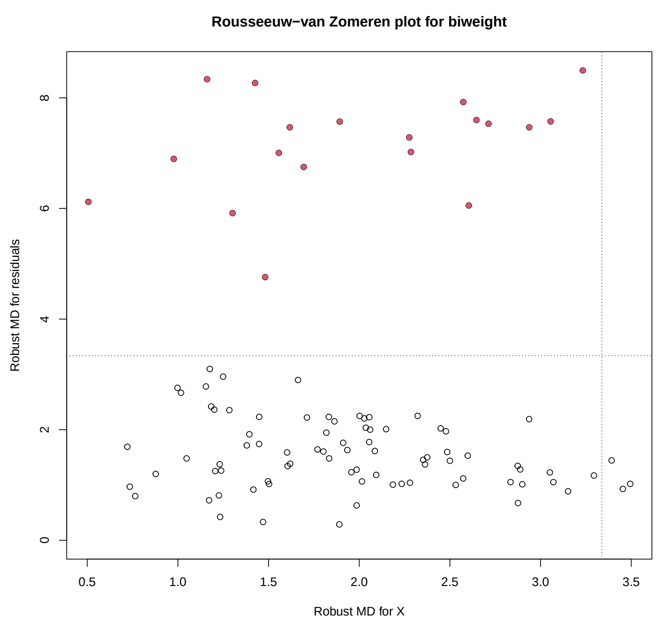

Figure 18: Rousseeuw-van Zomeren plot; 20% location contamination in measurements.

Larger amounts of 30%-40% of contamination lead to much larger values of $\widehat{\theta}_1$ and $\widehat{\theta}_2$, in particular for $\widehat{\theta}_2$. This causes a slight masking effect in the Rousseeuw-van Zomeren plot.

I also investigated scale contamination

$$\boldsymbol{\epsilon}_i \sim (1-r)N(\mathbf{0}, \mathbf{I}_4) + rN(\mathbf{0}, \mathbf{R}), \quad \mathbf{R} = \begin{pmatrix} 11 & 0 & 0 & 0 \\ 0 & 11 & 0 & 0 \\ 0 & 0 & 11 & 0 \\ 0 & 0 & 0 & 11 \end{pmatrix}$$

This leads to even larger values of $\widehat{\theta}_1$ and $\widehat{\theta}_2$, in particular $\widehat{\theta}_2 = \widehat{\sigma}^2_{\epsilon}$. As a consequence, this also produces a slight masking effect, in which a lot of contaminated point are not detected as outliers in the $\mathbf{y}$-space.

6 Resistance data

This dataset has been discussed in Copt, 2004 [3] and in Heritier *et al*, 2009 [5]. The data are taken from Table 3 in Berry [1]. Copt [3] used the data divided by 100 (see the remark on page 98). The model used for this dataset is

$$y_{ij} = \mu + \lambda_j + \gamma_i + \epsilon_{ij}, \quad i = 1, \dots, n;\ j = 1, \dots, 5,$$

where $\mu + \lambda_j$ is the expected resistance at the j-th electrode and the γ_i are random effects.

Different possible contrasts. There are different ways to express 6 parameters μ, $\lambda_1 \dots, \lambda_5$ into expected resistances at 5 electrodes. One of them is the so-called *"sum to zero"* contrast, for which

$$\begin{aligned}
y_{i1} &= \mu + \lambda_1 + \gamma_i + \epsilon_{i1}, \\
y_{i2} &= \mu + \lambda_2 + \gamma_i + \epsilon_{i2}, \\
y_{i3} &= \mu + \lambda_3 + \gamma_i + \epsilon_{i3}, \\
y_{i4} &= \mu + \lambda_4 + \gamma_i + \epsilon_{i4}, \\
y_{i5} &= \mu + \lambda_5 + \gamma_i + \epsilon_{i5} = \mu - \lambda_1 - \cdots - \lambda_4 + \gamma_i + \epsilon_{i5}.
\end{aligned}$$

With $\boldsymbol{\beta} = (\mu, \lambda_1, \lambda_2, \lambda_3, \lambda_4)$ the fixed effects term $\mathbf{X}\boldsymbol{\beta}$ is obtained with the following design matrix

$$\mathbf{X} = \begin{pmatrix} 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 \\ 1 & -1 & -1 & -1 & -1 \end{pmatrix} \tag{6.1}$$

With such a contrast, μ is interpreted as the overall expected resistance and λ_j, $j = 1, \dots, 5$, are interpreted as the deviation of the j-th electrode from the overall mean μ. An alternative is the so-called *"treatment"* contrast, for which

$$\begin{aligned}
y_{i1} &= \mu + \gamma_i + \epsilon_{i1}, \\
y_{i2} &= \mu + \lambda_2 + \gamma_i + \epsilon_{i2}, \\
y_{i3} &= \mu + \lambda_3 + \gamma_i + \epsilon_{i3}, \\
y_{i4} &= \mu + \lambda_4 + \gamma_i + \epsilon_{i4}, \\
y_{i5} &= \mu + \lambda_5 + \gamma_i + \epsilon_{i5}
\end{aligned}$$

With $\boldsymbol{\beta} = (\mu, \lambda_1, \lambda_2, \lambda_3, \lambda_4)$ the fixed effects term $\mathbf{X}\boldsymbol{\beta}$ is obtained with the following design matrix

$$\mathbf{X} = \begin{pmatrix} 1 & 0 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 1 & 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 1 & 0 \\ 1 & 0 & 0 & 0 & 1 \end{pmatrix} \tag{6.2}$$

With such a contrast, μ is interpreted as the expected resistance on E1 and λ_j, $j = 1, \dots, 4$, are interpreted as the deviation of the j-th electrode from E1.

Apart from the chosen design matrix, the model can be formulated as

$$\mathbf{y}_i = \mathbf{X}\boldsymbol{\beta} + \gamma_i \mathbf{Z} + \boldsymbol{\epsilon}_i, \quad i = 1, \dots, n$$

where

$$\mathbf{Z} = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}.$$

The measurements y_{ij}, $j = 1, \ldots, 5$, are resistance measurements at five different electrodes. It is assumed that the random effects are independent such that $\gamma_i \sim N(0, \sigma_\gamma^2)$ and that the measurement errors are independent, such that $\boldsymbol{\epsilon}_i \sim N(\mathbf{0}, \sigma_\epsilon^2 \mathbf{I}_5)$. Therefore, using the notation in Copt [3] and Heritier *et al* [5], the parameter vector is $\boldsymbol{\xi} = (\boldsymbol{\beta}, \boldsymbol{\theta})$, with $\boldsymbol{\beta} = (\mu, \lambda_1, \ldots, \lambda_4)$ and $\boldsymbol{\theta} = (\sigma_\gamma^2, \sigma_\epsilon^2)$.

The rho function. To analyze the data, the translated biweight $\rho_t(\,\cdot\,; M, c)$ can be used, where the cut-off constants are chosen such that the asymptotic rejection rate is $\pi = 0.01$ and the breakdown point is $\epsilon^* = 0.5$, see the remark on page 53 in Copt [3]. The resulting estimates are referred to as the CTBS estimates.

However, by taking $M = 0$ one obtains Tukey's biweight $\rho_B(\,\cdot\,; c) = \rho_t(\,\cdot\,; 0, c)$. The resulting estimates are referred to as the CBS estimates.

Finally, when $M \to \infty$, it can be seen that $\rho_t(d; M, c) \to d^2/2$, the rho-function that corresponds to the ML estimates. Therefore, we can obtain the ML estimates by using $\rho_t(d; M, c)$, with $M = 10\,000$.

Rocke's adaption. Finally, it may be that Copt [3] also uses Rocke's correction of the Mahalanobis distances

$$d_i = \sqrt{(\mathbf{y}_i - \mathbf{X}\widehat{\boldsymbol{\beta}})^T \mathbf{V}(\widehat{\boldsymbol{\beta}})^{-1} (\mathbf{y}_i - \mathbf{X}\widehat{\boldsymbol{\beta}})}$$

by replacing d_i by d_i/k, where

$$k = \frac{d_{[q]}}{\sqrt{(\chi_k^2)^{-1}(q/(n+1))}}, \quad q = \lfloor (n+k+1)/2 \rfloor,$$

see the remark on page 43. However, it is unclear whether Copt [3] actually uses the correction, or only refers to Rocke [10]. By the way, the correction has nothing to do with maintaining the right breakdown point, and was made by Rocke about the bi-flat rho function and not the translated biweight.

6.1 Results in Copt [3] and Heritier *et al* [5]

Chapter 8 in Copt [3] reports the analysis by the ML and CTBS estimators reports the analysis and Chapter 4 in Heritier *et al* [5] by the REML and CBS-MM estimates. I have tried to reproduce the results.

Note that the ML estimates correspond to a rho-function $\rho(d) = d^2/2$, with corresponding derivative $u(d) = 1$. Also note that $\rho_{ML}(d) = d^2/2$ is the limit of the translated biweight $\rho_t(d; M, c)$, as $M \to \infty$. This means that I can use my own code for the translated biweight with large M, say $M = 10\,000$, to compute the ML estimates. Re-scaling is essential here, because the S-constraint for the MLE becomes

$$\frac{1}{n}\sum_{i=1}^{n} u(d_i) d_i^2 = \frac{1}{n}\sum_{i=1}^{n} d_i^2 = b_0 = \mathbb{E}_{\mathbf{0}, \mathbf{I}_k} \|\mathbf{z}\|^2 = k$$

This means that the ratio

$$\frac{k}{\displaystyle\frac{1}{n}\sum_{i=1}^{n} u(d_i) d_i^2}$$

that appears in the iteration, is equal to 1. This makes the fixed point equation the same as the one for the MLE. My code is stable in the sense that starting from different initial values, the algorithm always converges to the same ML estimates.

ML estimates. Copt [3] reports the ML estimates in Table 8.1. They can be found running

```
options(contrasts=c("contr.sum","cont.poly"))
electrode=Electrode
electrode[,1]=electrode[,1]/100
names(electrode)
resistance.lme=lme(fixed=Resis~Elec,data=electrode,
                   random=list(Subject=pdIdent(~1)),method="ML")
```

Note that this corresponds to $\mathbf{X}$ from (6.1). I find the same ML estimates.

```
            betaMLErik betaMLEcopt betaMLElme
(Intercept)   2.030500        2.03   2.030500
Elec1        -0.213625       -0.21  -0.213625
Elec2         0.842625        0.84   0.842625
Elec3         0.549500        0.55   0.549500
Elec4        -0.526750       -0.53  -0.526750
```

and

```
  thetaMLErik thetaMLEcopt thetaMLE
1    1.329343        1.329 1.329342
2    2.098005        2.098 2.098005
```

REML estimates. Heritier *et al* [5] reports the REML estimates in Table 4.1. The estimates can be obtained with running

```
options(contrasts=c("contr.sum","cont.poly"))
electrode=Electrode
electrode[,1]=electrode[,1]/100
names(electrode)
resistance.lme=lme(fixed=Resis~Elec,data=electrode,
                   random=list(Subject=pdIdent(~1)),method="REML")
```

CTBS estimates. First note that Copt [3] does not re-scale the Mahalanobis distances, such that the S-constraint is satisfied. As a consequence, the resulting CTBS estimate is sensitive to the starting point and to whether Rocke's adaptation is used or not. The closest I get to the estimates reported by Copt [3] is

1. use $\mathbf{X}$ from (6.1)
2. not to use Rocke's adaptation
3. to start from location-covariance estimates $\boldsymbol{\mu}_0$ and $\mathbf{V}_0$ from `rogkmiss` and transfer these to initial values for $\boldsymbol{\beta}^{(0)} = (\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\boldsymbol{\mu}_0$ and $\boldsymbol{\theta}^{(0)}$ by solving $\mathbf{V}(\boldsymbol{\theta}^{(0)}) = \mathbf{V}_0$.

I find

```
  betaCTBSrik betaCTBScopt
1  1.62120099        1.628
2 -0.06937868       -0.068
3  0.51450295        0.512
4  0.14809870        0.142
5 -0.16822480       -0.158
```

and

```
 thetaCTBSrik thetaCTBSCopt
1     1.451052         1.087
2     2.180104         0.711
```

CBS-MM estimates. The results in Heritier *et al* [5] can be obtained by

1. do not re-scale to satisfy the S-constraint
2. use Rocke's adaptation
3. use $\mathbf{X}$ from (6.1)
4. The initial values seem to been taken from the location-covariance estimates $\boldsymbol{\mu}_0$ and $\mathbf{V}_0$ from `rogkmiss`. The covariance $\mathbf{V}_0$ is taken as initial value for $\mathbf{V}(\boldsymbol{\theta}^{(0)})$ and the initial value $\boldsymbol{\beta}^{(0)}$ is obtained from a 1-step iteration of the regression fixed point equation.

Using Rocke's adaption, I find CBS estimates for $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$. It turns out that

```
  SQRTthetaCBSrik SQRTthetaCBSHeritier
1     0.8425040746                0.842
2     0.7610013472                0.761
```

The CBS estimates for the covariance parameters, match with the ones in Table 4.3 in Heritier *et al* [5]. To perform the MM-step, we fix the CBS covariance estimate, and perform a second CBS estimate to update the estimates for $\boldsymbol{\beta}$, where we change the cut-off value to obtain higher efficiency. Heritier *et al* [5] matches cut-off value c_1 with 95% efficiency. If find $c_1 = 6.096263$ and

```
     betaCBSrik  betaCBSMMrik betaCBSMMHeritier
1  1.3491254582  1.4397141328             1.440
2 -0.1935244065 -0.1609962740            -0.161
3  0.3218545600  0.4034062650             0.403
4  0.2868180874  0.2433801149             0.243
5 -0.1547274102 -0.1694137797            -0.169
```

This matches perfectly with the CBS-MM estimates reported in Table 4.3 in Heritier *et al* [5].

6.2 CTBS and CBS-MM estimates that satisfy the S-constraint.

I computed the CTBS estimates with

1. re-scaling of the Mahalanobis distances, such that they satisfy the S-constraint.
2. the $\mathbf{X}$ from (6.1).
3. the translated biweight with $M = 2.017548$ and $c = 1.866557$.

Starting values for $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$ based on several (robust) multivariate location-scale estimates (i.e., regular OGK, weighted OGK, `rogkmiss`, MCD, or the sample mean and sample covariance, in combination with both options to transfer to initial values) all lead to the same CTBS estimates. Moreover, starting from the estimates reported in Copt [3], I also arrive at the same CTBS estimates.

I find

```
 betaCTBSrescale betaCTBScopt
1     1.4068525948        1.628
2    -0.2033510329       -0.068
3     0.3612523499        0.512
4     0.2790165792        0.142
5    -0.1747425887       -0.158
```

and

```
  thetaCTBSrescale thetaCTBSCopt
1      0.8277895843         1.087
2      0.7270973018         0.711
```

We can compare the determinants of both solutions, after re-scaling them to satisfy the S-constraint. I find

```
  detCTBSresistanceRESCALE detCTBSresistanceCopt
1              1.360023179           1.810413459
```

I conclude that the solution in Copt [3] is not the best CTBS solution to the S-minimization problem.

CBS-MM estimates. I used

1. re-scaling of the Mahalanobis distances, such that they satisfy the S-constraint.
2. the $\mathbf{X}$ from (6.1).
3. the translated biweight with $M = 0$ and $c = 4.652023$ (which is the same as the biweight with $c = 4.652023$).

Starting values for $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$ based on several multivariate location-scale estimates all lead to the same CBS estimates. Afterwards, I computed the MM-estimates for $\boldsymbol{\beta}$ with the biweight function with $c_1 = 6.096263$ and

1. did not use re-scaling
2. or Rocke's adaptation.

For the CBS estimates and CBS-MM estimates, I find

```
  betaCBSrescale betaCBSMMrescale betaCBSMMHeritier
1   1.4037235857     1.4352455330             1.440
2  -0.1762969575    -0.1617255441            -0.161
3   0.3788562273     0.4033765468             0.403
4   0.2624506883     0.2448399600             0.243
5  -0.1690997801    -0.1720360220            -0.169
```

and

```
  SQRTthetaCBSrescale SQRTthetaCBSHeritier
1        0.9056086861                0.842
2        0.8922838601                0.761
```

Although the CBS-estimates for $\boldsymbol{\theta}$ differ, I end up with (more or less) the same CBS-MM estimates for $\boldsymbol{\beta}$ as in Heritier *et al* [5]. If I compare the determinants of the CBS estimators of the first step, after re-scaling them to satisfy the S-constraint, I find

```
  detCBSresistanceRik detCBSresistanceHeritier
1         7.933499019              8.644552483
```

I conclude that the CBS estimator of Heritier is not the optimal solution to the S-minimization problem.

6.3 Standard errors

The next step is to investigate whether the differences with Copt [3] and Heritier *et al* [5] lead to different conclusions about the significance of the parameters. For this we need the standard errors of the estimates. The standard errors of the estimates can be derived from the expressions of the asymptotic variances.

The asymptotic covariance of $\sqrt{n}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})$ is given by

$$\gamma \left(\mathbb{E} \left[\mathbf{X}^T \boldsymbol{\Sigma}^{-1} \mathbf{X} \right] \right)^{-1}, \quad \gamma = \frac{\mathbb{E}_{\mathbf{0},\mathbf{I}_k} \left[\rho'(\|\mathbf{z}\|)^2 \right]}{k\alpha^2}$$

where

$$\alpha = \mathbb{E}_{\mathbf{0},\mathbf{I}_k}\left[\left(1-\frac{1}{k}\right)\frac{\rho'(\|\mathbf{z}\|)}{\|\mathbf{z}\|}+\frac{1}{k}\rho''(\|\mathbf{z}\|)\right].$$

Due to the simple structure of $\mathbf{X}$, $\mathbf{Z}$, and the covariance $\sigma_\epsilon^2\mathbf{I}_5$ of the measurement error, the expressions in Copt *et al* [2] lead to the same asymptotic variances. The standard errors are found by dividing the asymptotic variances by n and then take the square roots of the elements on the main diagonal.

ML estimates. For the ML estimates, I find $\gamma = 1$, and for the asymptotic covariance matrix of $\sqrt{n}(\widehat{\boldsymbol{\beta}}_{ML} - \boldsymbol{\beta})$, I find

```
        [,1]    [,2]    [,3]    [,4]    [,5]
[1,] 1.74894  0.0000  0.0000  0.0000  0.0000
[2,] 0.00000  1.6784 -0.4196 -0.4196 -0.4196
[3,] 0.00000 -0.4196  1.6784 -0.4196 -0.4196
[4,] 0.00000 -0.4196 -0.4196  1.6784 -0.4196
[5,] 0.00000 -0.4196 -0.4196 -0.4196  1.6784
```

The SE's can be found by diving the main diagonal by $n = 16$ and take the square root. Copt [3] reports the SE's in Table 8.1 and Heritier *et al* [5] in Table 4.1. This gives

```
  SEbetaMLErik SEbetaRMLEHeritier SEbetaMLEcopt
1    0.3306191              0.341        0.3306
2    0.3238831              0.334        0.3239
3    0.3238831              0.334        0.3239
4    0.3238831              0.334        0.3239
5    0.3238831              0.334        0.3239
```

The values match with those reported in Copt [3] and deviate somewhat from Heritier *et al* [5]. The latter may be due to the fact that our ML and REML estimates for $\boldsymbol{\theta}$ differ.

CTBS estimates. For the regression CTBS estimator I find $\gamma = 1.198852$ and asymptotic covariance

```
        [,1]     [,2]     [,3]     [,4]     [,5]
[1,] 1.16674  0.00000  0.00000  0.00000  0.00000
[2,] 0.00000  0.69735 -0.17434 -0.17434 -0.17434
[3,] 0.00000 -0.17434  0.69735 -0.17434 -0.17434
[4,] 0.00000 -0.17434 -0.17434  0.69735 -0.17434
[5,] 0.00000 -0.17434 -0.17434 -0.17434  0.69735
```

Again, the SE's can be found by diving the main diagonal by $n = 16$ and take the square root. Copt [3] reports the SE's in Table 8.1 This gives

```
  SEbetCTBSrescale SEbetaCTBScopt
1     0.2700386320         0.2638
2     0.2087680771         0.2553
3     0.2087680771         0.2553
4     0.2087680771         0.2553
5     0.2087680771         0.2553
```

The differences with Copt [3] may be attributed to re-scaling.

CBS-MM estimates. For the CBS-MM estimates I find $\gamma = 1.052632$ and asymptotic covariance

```
1.03091  0.00000  0.00000  0.00000  0.00000
0.00000  0.67046 -0.16761 -0.16761 -0.16761
0.00000 -0.16761  0.67046 -0.16761 -0.16761
0.00000 -0.16761 -0.16761  0.67046 -0.16761
0.00000 -0.16761 -0.16761 -0.16761  0.67046
```

Again, the SE's can be found by diving the main diagonal by $n = 16$ and take the square root. Heritier *et al* [5] in Table 4.3. This gives

```
  SEbetaCBSMMrescale SEbetaCBSMMheritier
1       0.2538339328               0.233
2       0.2047039697               0.175
3       0.2047039697               0.175
4       0.2047039697               0.175
5       0.2047039697               0.175
```

6.4 Summary

We can combine our ML, CTBS, and CBS-MM estimates and their standard errors and compute the corresponding standard normal two-sided p-value.

ML estimators. For the MLE I get

```
        MLE SE.MLE       TMLE PvalMLE
1  2.030500  0.331  6.1415087  0.0000
2 -0.213625  0.324 -0.6595744  0.5095
3  0.842625  0.324  2.6016331  0.0093
4  0.549500  0.324  1.6965998  0.0898
5 -0.526750  0.324 -1.6263584  0.1039
```

If I compare this with Table 4.1 or 4.3 in Heritier *et al* [5], I find the same conclusions at 5% significance level.

CTBS estimators. For the CTBS estimates I find

```
        CTBS SE.CTBS      TCTBS PvalCTBS
1  1.4068524   0.270  5.2098114   0.0000
2 -0.2033518   0.209 -0.9740561   0.3300
3  0.3612510   0.209  1.7303943   0.0836
4  0.2790176   0.209  1.3364960   0.1814
5 -0.1747419   0.209 -0.8370145   0.4026
```

If I compare this with Table 8-1 in Copt [3], the p-values are quite different. I find the same conclusions, except for $\beta_3 = \lambda_2$ which I find not significant at 5% level and which is has p-value 0.0449 in Copt [3].

CBS-MM estimators. For the CBS-MM estimators¡ I find

```
       CBSMM SE.CBSMM     TCBSMM PvalCBSMM
1  1.4352456    0.254  5.6542648    0.0000
2 -0.1617255    0.205 -0.7900461    0.4295
3  0.4033765    0.205  1.9705365    0.0488
4  0.2448399    0.205  1.1960686    0.2317
5 -0.1720360    0.205 -0.8404137    0.4007
```

This gives a difference with the CTBS estimator, in the sense that here $\beta_3 = \lambda_2$ is significant at 5% level. Therefore, I find the same conclusions as in Table 4.3 in Heritier *et al* [5]

6.5 Data exploration

The parameter $\beta_1 = \mu$ is the expected response on electrode E1, and it is no surprise that it is significantly from zero. The difference in p-values with Heritier *et al* [5] is due to the smaller standard errors. This is a consequence of the smaller CBS estimates for $\boldsymbol{\theta}$, which is a consequence of not re-scaling to satisfy the S-constraint, when computing the CBS estimate.

The plot of the robust Mahalanobis distances identifies subject 2 and 15 as clear outliers, see Figure 19. Based on my estimates, I conclude there is (almost) no effect of the electrode type. This coincides with the REML analysis in Heritier *et al* [5] after removal of subject 15 and with the conclusion in Berry [1] after a log-transform of the data.

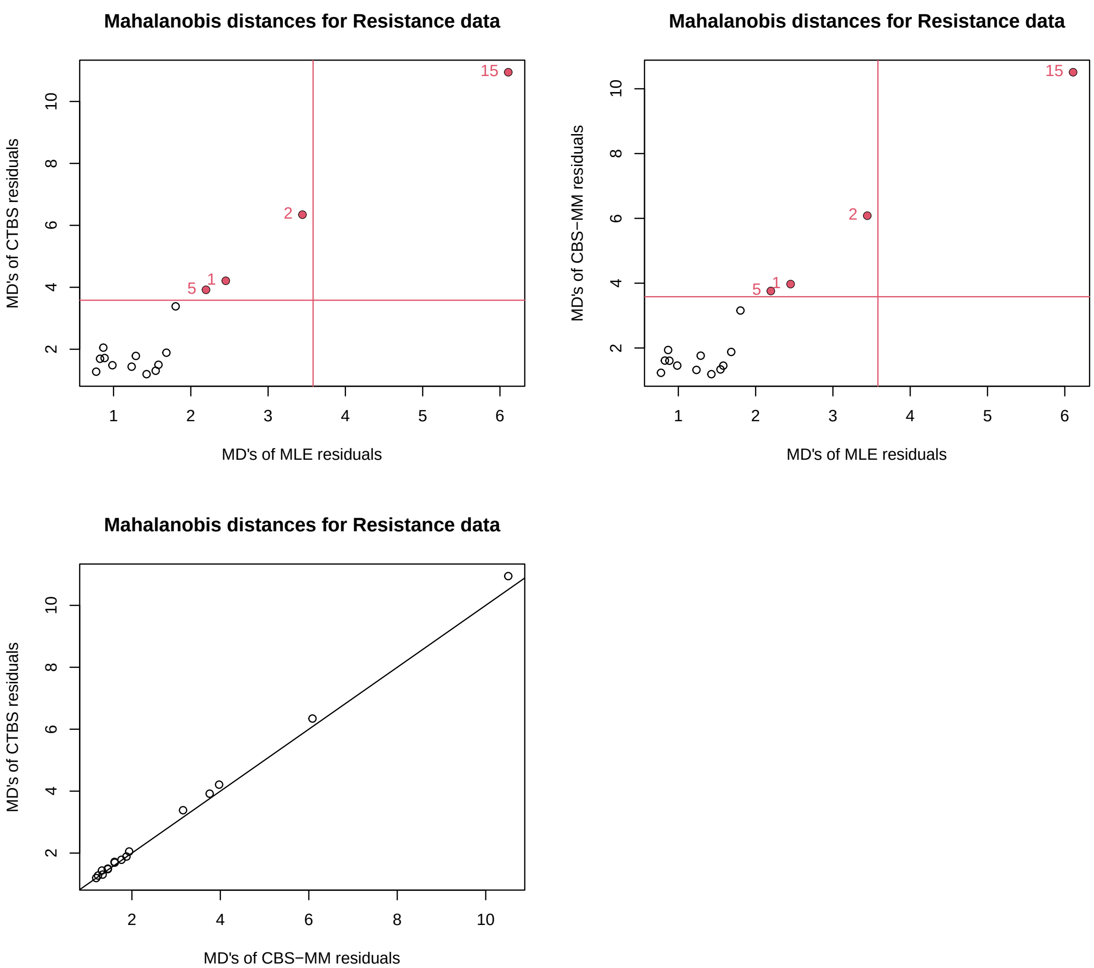

Figure 19: Mahalanobis distances for the resistance data

The boxplots of the responses at the 5 electrodes with and without observation 2 and 15 can be seen in Figure 20. This seems to confirm no effect of electrode type after removal of the outlying observations.

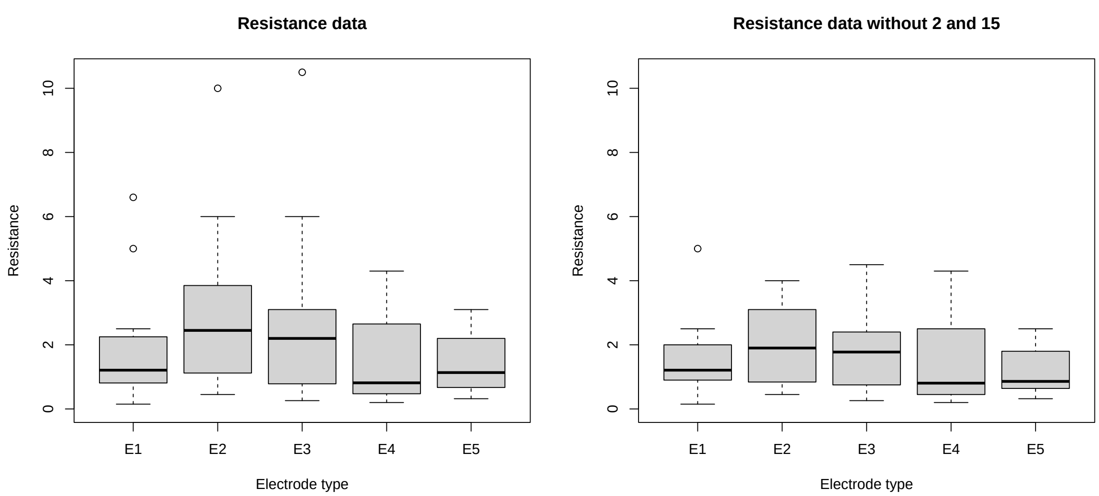

Figure 20: Boxplot responses with and without subjects 2 and 15.

7 Semantic Priming data

Consider the example in Section 4.2.3 in [6] about semantic and associative priming in picture naming. The model is given by

$$y_{ijk} = \mu + \lambda_j + \gamma_k + (\lambda\gamma)_{jk} + s_i + (\lambda s)_{ij} + (\gamma s)_{ik} + \epsilon_{ijk}, \quad i = 1,\dots,n;\ j = 1,2;\ k = 1,2,3,$$

with λ_j, $j = 1,2$, the fixed effect for the factor delay (long/short), γ_k, $k = 1,2,3$, the fixed effect for the factor condition (neutral/related/unrelated), and $(\lambda\gamma)_{jk}$, $j = 1,2$, $k = 1,2,3$ the fixed effects for the interaction between the factor delay and condition. This yields 12 parameters, for which are different ways or expressing them in 6 parameters.

Sum to zero contrast. One of those is the so-called "sum to zero" contrast, which means

$$\begin{aligned}
\lambda_1 + \lambda_2 &= 0 \Rightarrow \lambda_2 = -\lambda_1 \\
\gamma_1 + \gamma_2 + \gamma_3 &= 0 \Rightarrow \gamma_3 = -\gamma_1 - \gamma_2 \\
(\lambda\gamma)_{11} + (\lambda\gamma)_{12} + (\lambda\gamma)_{13} &= 0 \Rightarrow (\lambda\gamma)_{13} = -(\lambda\gamma)_{11} - (\lambda\gamma)_{12} \\
(\lambda\gamma)_{21} + (\lambda\gamma)_{22} + (\lambda\gamma)_{23} &= 0 \\
(\lambda\gamma)_{11} + (\lambda\gamma)_{21} &= 0 \Rightarrow (\lambda\gamma)_{21} = -(\lambda\gamma)_{11} \\
(\lambda\gamma)_{12} + (\lambda\gamma)_{22} &= 0 \Rightarrow (\lambda\gamma)_{22} = -(\lambda\gamma)_{12} \\
(\lambda\gamma)_{13} + (\lambda\gamma)_{23} &= 0 \Rightarrow (\lambda\gamma)_{23} = -(\lambda\gamma)_{13} = (\lambda\gamma)_{11} + (\lambda\gamma)_{12}
\end{aligned}$$

These contrasts are used by Heritier *et al* [5] (see remark on bottom of page 107). In this way, we are left with 6 parameters $\mu, \lambda_1, \gamma_1, \gamma_2, (\lambda\gamma)_{11}$, and $(\lambda\gamma)_{12}$. This leads to the following model for the observations

$$\begin{aligned}
y_{i11} &= \mu + \lambda_1 + \gamma_1 + (\lambda\gamma)_{11} + s_i + (\lambda s)_{i1} + (\gamma s)_{i1} + \epsilon_{i11} \\
y_{i12} &= \mu + \lambda_1 + \gamma_2 + (\lambda\gamma)_{12} + s_i + (\lambda s)_{i1} + (\gamma s)_{i2} + \epsilon_{i12} \\
y_{i13} &= \mu + \lambda_1 + \gamma_3 + (\lambda\gamma)_{13} + s_i + (\lambda s)_{i1} + (\gamma s)_{i3} + \epsilon_{i13} \\
&= \mu + \lambda_1 + (-\gamma_1 - \gamma_2) + (-(\lambda\gamma)_{11} - (\lambda\gamma)_{12}) + s_i + (\lambda s)_{i1} + (\gamma s)_{i3} + \epsilon_{i13} \\
y_{i21} &= \mu + \lambda_2 + \gamma_1 + (\lambda\gamma)_{21} + s_i + (\lambda s)_{i2} + (\gamma s)_{i1} + \epsilon_{i21} \\
&= \mu - \lambda_1 + \gamma_1 + (\lambda\gamma)_{21} + s_i + (\lambda s)_{i2} + (\gamma s)_{i1} + \epsilon_{i21} \\
y_{i22} &= \mu + \lambda_2 + \gamma_2 + (\lambda\gamma)_{22} + s_i + (\lambda s)_{i2} + (\gamma s)_{i2} + \epsilon_{i22} \\
&= \mu - \lambda_1 + \gamma_2 - (\lambda\gamma)_{12} + s_i + (\lambda s)_{i2} + (\gamma s)_{i2} + \epsilon_{i22} \\
y_{i23} &= \mu + \lambda_2 + \gamma_3 + (\lambda\gamma)_{23} + s_i + (\lambda s)_{i2} + (\gamma s)_{i3} + \epsilon_{i23} \\
&= \mu - \lambda_1 + (-\gamma_1 - \gamma_2) + (\lambda\gamma)_{11} + (\lambda\gamma)_{12} + s_i + (\lambda s)_{i2} + (\gamma s)_{i3} + \epsilon_{i23}
\end{aligned}$$

In relation with vector of fixed effects parameters $\boldsymbol{\beta} = (\mu, \lambda_1, \gamma_1, \gamma_2, (\lambda\gamma)_{11}, (\lambda\gamma)_{12})$, this corresponds to design matrix

$$\mathbf{X} = \begin{pmatrix} 1 & 1 & 1 & 0 & 1 & 0 \\ 1 & 1 & 0 & 1 & 0 & 1 \\ 1 & 1 & -1 & -1 & -1 & -1 \\ 1 & -1 & 1 & 0 & -1 & 0 \\ 1 & -1 & 0 & 1 & 0 & -1 \\ 1 & -1 & -1 & -1 & 1 & 1 \end{pmatrix}.$$

In this case

$$\mathbf{y}_i = \mathbf{X}\boldsymbol{\beta} + \begin{bmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{bmatrix} s_i + \begin{bmatrix} 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} (\lambda s)_{i1} \\ (\lambda s)_{i2} \end{bmatrix} + \begin{bmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \\ 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} (\gamma s)_{i1} \\ (\gamma s)_{i2} \\ (\gamma s)_{i3} \end{bmatrix} + \epsilon_i, \quad i = 1,\dots,21.$$

Furthermore, there are three covariance parameters σ_s^2, $\sigma_{\lambda s}^2$, and $\sigma_{\gamma s}^2$, that correspond to the random effects, $s_i \sim N(0, \sigma_s^2)$, $(\boldsymbol{\lambda} s)_i \sim N(0, \sigma_{\lambda s}^2 \mathbf{I}_2)$, $(\boldsymbol{\gamma} s)_i \sim N(0, \sigma_{\gamma s}^2 \mathbf{I}_3)$, and one covariance parameter σ_ϵ^2 for the measurement error $\epsilon_i \sim N(0, \sigma_\epsilon^2 \mathbf{I}_6)$. This means that the covariance parameters are $\boldsymbol{\theta} = (\sigma_s^2, \sigma_{\lambda s}^2, \sigma_{\gamma s}^2, \sigma_\epsilon^2)$.

Moreover, the design matrices for the random effects s_i, $(\boldsymbol{\lambda} s)_i$ and $(\boldsymbol{\gamma} s)_i$ are given by

$$\mathbf{Z}_1 = \mathbf{e}_6, \quad \mathbf{Z}_2 = \begin{bmatrix} \mathbf{e}_3 & \mathbf{0}_3 \\ \mathbf{0}_3 & \mathbf{e}_3 \end{bmatrix}, \quad \mathbf{Z}_3 = \begin{bmatrix} \mathbf{I}_3 \\ \mathbf{I}_3 \end{bmatrix}.$$

This leads to

$$\begin{aligned}
\mathbf{L}_1 &= \mathbf{Z}_1 \mathbf{Z}_1^T = \mathbf{e}_6 \mathbf{e}_6^T = \mathbf{J}_6 \\
\mathbf{L}_2 &= \mathbf{Z}_2 \mathbf{Z}_2^T = \begin{bmatrix} \mathbf{e}_3 & \mathbf{0}_3 \\ \mathbf{0}_3 & \mathbf{e}_3 \end{bmatrix} \begin{bmatrix} \mathbf{e}_3^T & \mathbf{0}_3^T \\ \mathbf{0}_3^T & \mathbf{e}_3^T \end{bmatrix} = \mathbf{I}_2 \otimes \mathbf{J}_3 \\
\mathbf{L}_3 &= \mathbf{Z}_3 \mathbf{Z}_3^T = \begin{bmatrix} \mathbf{I}_3 \\ \mathbf{I}_3 \end{bmatrix} \begin{bmatrix} \mathbf{I}_3 & \mathbf{I}_3 \end{bmatrix} = \mathbf{J}_2 \otimes \mathbf{I}_3 \\
\mathbf{L}_4 &= \mathbf{I}_6,
\end{aligned}$$

and therefore we have the following linear structure of the covariance matrix:

$$\begin{aligned}
\mathbf{V}(\boldsymbol{\theta}) &= \theta_1 \mathbf{L}_1 + \theta_2 \mathbf{L}_2 + \theta_3 \mathbf{L}_3 + \theta_4 \mathbf{L}_4 \\
&= \sigma_s^2 \mathbf{J}_6 + \sigma_{\lambda s}^2 (\mathbf{I}_2 \otimes \mathbf{J}_3) + \sigma_{\gamma s}^2 (\mathbf{J}_2 \otimes \mathbf{I}_3) + \sigma_\epsilon^2 \mathbf{I}_6.
\end{aligned}$$

7.1 Results in Heritier *et al* [5]

We cannot compare our findings with Copt [3], for two reasons, (i) the data he uses are not available and (ii) he treats a different model with factors Object (hammer/saw/screwdriver) and Item (related/neutral/unrelated) for this topic. However, we can compare our findings with Heritier *et al* [5] who report the REML and CBS-MM estimators in Table 4.5.

Note that the dataframe `Semantic` used by Heritier and the function `lme`, has factor `Delay` with levels

```
"long"  "short"
```

and factor `Condition` with levels

```
"neutral"   "related"   "unrelated"
```

Furthermore, it has the following order of observations per subject

```
  Subject Delay Condition Resp
1     S49 short   related  456
2     S49  long   related  462
3     S49 short   neutral  499
4     S49  long   neutral  502
5     S49 short unrelated  496
6     S49  long unrelated  514
```

This means that the order is

$$y_{22}, \quad y_{12}, \quad y_{21}, \quad y_{11}, \quad y_{23}, \quad y_{13}$$

Hence, in order to compare our results with Heritier or with the values produced by the function `lme`, we have two options

1. Keep dataframe `Semantic` as it is and reorder the rows of the design matrices $\mathbf{X}$, $\mathbf{Z}_1$, $\mathbf{Z}_2$, and $\mathbf{Z}_3$ in our own code.

2. Keep our own code with the design matrices $\mathbf{X}$, $\mathbf{Z}_1$, $\mathbf{Z}_2$, and $\mathbf{Z}_3$ as they are and reorder the rows per subject in dataframe `Semantic`.

ML estimates. The ML estimates for $\boldsymbol{\beta}$ can be obtained with the code

```
options(contrasts=c("contr.sum","cont.poly"))
semantic.lme=lme(fixed=Resp~Delay*Condition,
          data=Semantic,
          random=list(Subject=pdBlocked(list(pdIdent(~1),
                              pdIdent(~Delay-1),
                              pdIdent(~Condition-1)))),
          method="ML")
```

With both options for matching our code with the order of the observations, I find

```
                   betaMLrik  betaMLlme
(Intercept)       633.436508 633.436508
Delay1            -18.071429 -18.071429
Condition1         18.563492  18.563492
Condition2        -51.222222 -51.222222
Delay1:Condition1  -3.690476  -3.690476
Delay1:Condition2  16.809524  16.809524
```

This matches with the values reported in Heritier *et al* [5]. However, for the estimates for the covariance parameters, I run into problems finding negative variance estimates. Heritier has a similar problem when computing the CBS-MM estimates, but the function `lme` apparently finds a way to avoid these problems. I find within 4 iterations

```
  thetaMLrik thetaMLlme
1  15505.291  14320.215
2  -2370.168      0.000
3   -359.986    825.097
4  12035.100   9664.933
```

This means we have negative estimates for variances $\theta_2 = \sigma^2_{\lambda s}$ and $\theta_3 = \sigma^2_{\gamma s}$. Negative variances may be an indication of a wrong model for the covariance matrix or it may be an indication that the value of the variance component is zero; especially if it is large and negative might well be suggestive that it is zero (see comments on page 61 in Searle [13]). One option is to replace negative variances by zero. If I do this within the iteration, I find the same values for 1 and 4, whereas then 2 and 3 are equal to zero.

REML estimates. The ML estimates for $\boldsymbol{\beta}$ can be obtained with the code

```
options(contrasts=c("contr.sum","cont.poly"))
semantic.lme=lme(fixed=Resp~Delay*Condition,
          data=Semantic,
          random=list(Subject=pdBlocked(list(pdIdent(~1),
                              pdIdent(~Delay-1),
                              pdIdent(~Condition-1)))),
          method="REML")
```

Heritier *et al* [5] report the REML estimates in Table 4.5. I find

```
                  betaREMLlme betaREMLheritier
(Intercept)        633.436508          633.436
Delay1             -18.071429          -18.071
Condition1          18.563492           18.563
Condition2         -51.222222          -51.220
Delay1:Condition1   -3.690476           -3.690
Delay1:Condition2   16.809524           16.809
```

and

```
  SQRTthetaREMLlme SQRTthetaREMLheritier
1          122.622               122.622
2            0.007                 0.006
3           29.434                29.433
4          100.738               100.730
```

This matches reasonably well with the values reported by Heritier.

CBS and CBS-MM estimates. I used the translated biweight with $M = 0$ and $c = 5.147689$ (which is the same as the biweight with $c = 5.147689$) and

1. I do not re-scale to satisfy the S-constraint
2. use Rocke's adaptation
3. Match the order of observations to the design matrix $\mathbf{X}$.
4. The initial values seem to been taken from the location-covariance estimates $\boldsymbol{\mu}_0$ and $\mathbf{V}_0$ from `rogkmiss`. The covariance $\mathbf{V}_0$ is taken as initial value for $\mathbf{V}(\boldsymbol{\theta}^{(0)})$ and the initial value $\boldsymbol{\beta}^{(0)}$ is obtained from a 1-step iteration of the regression fixed point equation.

This yields a negative value for $\widehat{\theta}_2$, as in Heritier *et al* [5]. Afterwards, I computed the MM-estimates for $\boldsymbol{\beta}$ with the biweight function with $c_1 = 6.356216$ and

1. Left $\widehat{\theta}_2 = -758.4122$ unchanged, which does not affect the positive definiteness of $\mathbf{V}(\widehat{\boldsymbol{\theta}})$
2. I do not use re-scaling
3. I used Rocke's adaptation.

For the CBS estimates and CBS-MM estimates, I find

```
  betaCBSrik betaCBSMMsemanticRik betaCBSMMHeritier
1  574.08644           586.420463           586.420
2  -17.42581           -17.875726           -17.876
3   16.25436            14.316692            14.317
4  -59.15527           -56.994476           -56.994
5   12.57753            12.706480            12.706
6    6.60549             8.844396             8.844
```

and

```
  SQRTthetaCBSrik SQRTthetaCBSHeritier
1        77.99054               77.991
2             NaN                   NA
3        27.19920               27.199
4        81.88500               81.885
```

This matches perfectly with the values reported in Table 4.5 in Heritier *et al* [5].

7.2 CBS-MM estimates that satisfy the S-constraint.

I used

1. re-scaling of the Mahalanobis distances, such that they satisfy the S-constraint.
2. Match the order of observations to the design matrix $\mathbf{X}$.
3. the translated biweight with $M = 0$ and $c = 5.147689$ (which is the same as the biweight with $c = 5.147689$).

This yields a negative value for $\widehat{\theta}_2$, as in Heritier *et al* [5]. Afterwards, I computed the MM-estimates for $\boldsymbol{\beta}$ with the biweight function with $c_1 = 6.356227$ and

1. left $\widehat{\theta}_2 = -787.8901$ unchanged, which does not affect the positive definiteness of $\mathbf{V}(\widehat{\boldsymbol{\theta}})$.
2. did not use re-scaling
3. did not use Rocke's adaptation

For the CBS estimates and CBS-MM estimates, I find

```
  betaCBSrescale betaCBSMMsemanticRESCALE betaCBSMMHeritier
1     577.770355               590.132801           586.420
2     -17.626675               -17.888090           -17.876
3      15.495349                14.170968            14.317
4     -58.783391               -55.931705           -56.994
5      12.909620                12.314851            12.706
6       7.162824                 9.615452             8.844
```

and

```
  SQRTthetaCBSrescale SQRTthetaCBSHeritier
1            83.10227               77.991
2                  NA                   NA
3            29.87492               27.199
4            86.22060               81.885
```

The estimates are quite similar and the differences may be attributed to not re-scaling and using Rocke's adaption by Heritier.

7.3 Standard Errors

The asymptotic covariance of $\sqrt{n}(\widehat{\boldsymbol{\beta}} - \boldsymbol{\beta})$ is given by

$$\gamma \left(\mathbb{E}\left[\mathbf{X}^T \boldsymbol{\Sigma}^{-1} \mathbf{X}\right]\right)^{-1}, \quad \gamma = \frac{\mathbb{E}_{\mathbf{0},\mathbf{I}_k}\left[\rho'(\|\mathbf{z}\|)^2\right]}{k\alpha^2}$$

where

$$\alpha = \mathbb{E}_{\mathbf{0},\mathbf{I}_k}\left[\left(1 - \frac{1}{k}\right)\frac{\rho'(\|\mathbf{z}\|)}{\|\mathbf{z}\|} + \frac{1}{k}\rho''(\|\mathbf{z}\|)\right].$$

Due to the simple structure of $\mathbf{X}$, $\mathbf{Z}$, and the covariance $\sigma_\epsilon^2 \mathbf{I}_6$ of the measurement error, the expressions in Copt *et al* [2] lead to the same asymptotic variances. The standard errors are found by dividing the asymptotic variances by n and then take the square roots of the elements on the main diagonal.

ML estimates. For the ML estimates, I find $\gamma = 1$. When I leave the negative estimates for θ_2 and θ_3 as they are, for the asymptotic covariance matrix of $\sqrt{n}(\widehat{\boldsymbol{\beta}}_{ML} - \boldsymbol{\beta})$, I find

```
16206.06   0.000     0.000     0.000     0.00     0.00
    0.00 820.766     0.000     0.000     0.00     0.00
    0.00   0.000  3771.709 -1885.855     0.00     0.00
    0.00   0.000 -1885.855  3771.709     0.00     0.00
    0.00   0.000     0.000     0.000  4011.70 -2005.85
    0.00   0.000     0.000     0.000 -2005.85  4011.70
```

The SE's can be found by diving the main diagonal by $n = 16$ and take the square root. Heritier *et al* [5] report the SE's in Table 4.5. This gives

```
  SEbetaMLrik SEbetaMLlme SEbetaREMLheritier
1   27.779799   27.779806             28.465
2    6.251728    8.758186              8.974
3   13.401686   13.401687             13.732
4   13.401686   13.401687             13.732
5   13.821481   12.385945             12.691
6   13.821481   12.385945             12.691
```

These values are comparable in order of magnitude. The difference with the values produced by `lme` may be due to the way it handles negative variances. The difference with Heritier *et al* [5] may be due to the fact that Heritier reports the REML estimates.

CBS-MM estimates. The expression for the limiting covariance of $\sqrt{n}(\widehat{\boldsymbol{\beta}}-\boldsymbol{\beta})$ remains the same (see (4.42) in Heritier *et al* [5]), but should be computed with cut-off constant $c_1 = 6.356216$. For the regression CBS estimator I find $\gamma = 1.052632$. When I leave the negative estimate for $\widehat{\theta}_2$ as it is I find asymptotic covariance

```
8472.15   0.00     0.000     0.000     0.000     0.000
   0.00 889.53     0.000     0.000     0.000     0.000
   0.00   0.00  3234.741 -1617.370     0.000     0.000
   0.00   0.00 -1617.370  3234.741     0.000     0.000
   0.00   0.00     0.000     0.000  2608.417 -1304.209
   0.00   0.00     0.000     0.000 -1304.209  2608.417
```

Again, the SE's can be found by diving the main diagonal by $n = 16$ and take the square root. Heritier *et al* [5] report the SE's in Table 4.5. This gives

```
  SEbetaCBSMMrescale SEbetaCBSMMheritier
1          20.085709              18.817
2           6.508346               6.082
3          12.411095              11.691
4          12.411095              11.691
5          11.144970              10.582
6          11.144970              10.582
```

Again, the values are comparable. The differences may be due to the difference in the estimates for $\boldsymbol{\theta}$.

7.4 Summary.

By putting everything together we find the following summary for the MLE

	MLE	SE.MLE	TMLE	PvalMLE	PvalHeritier
1	633.436508	27.780	22.8020552	0.0000	<10E-4
2	-18.071429	6.252	-2.8906296	0.0038	0.046
3	18.563492	13.402	1.3851609	0.1660	0.179
4	-51.222222	13.402	-3.8220729	0.0001	<10E-4
5	-3.690476	13.821	-0.2670102	0.7895	0.771
6	16.809524	13.821	1.2161884	0.2239	0.188

We find similar conclusions at 5% level as in Heritier *et al* [5]. However, the p-value for $\beta_2 = \lambda_1$ is more extreme.

For the CBS-MM estimator, I find

	CBSMM	SE.CBSMM	TCBS	PvalCBSMM	PvalHeritier
1	590.132801	20.086	29.3807308	0.0000	<10E-4
2	-17.888090	6.508	-2.7484847	0.0060	0.003
3	14.170968	12.411	1.1417984	0.2535	0.221
4	-55.931705	12.411	-4.5065892	0.0000	<10E-4
5	12.314851	11.145	1.1049694	0.2692	0.230
6	9.615452	11.145	0.8627616	0.3883	0.403

We find similar conclusions at 5% level as in Heritier *et al* [5]. This, the p-value for $\beta_2 = \lambda_1$ is less extreme.

7.5 Outlying observations

When computing the residuals based on my findings I find the Mahalanobis distances as in Figure 21. It is surprising that this plot identifies completely different items (19, 12 and perhaps 7

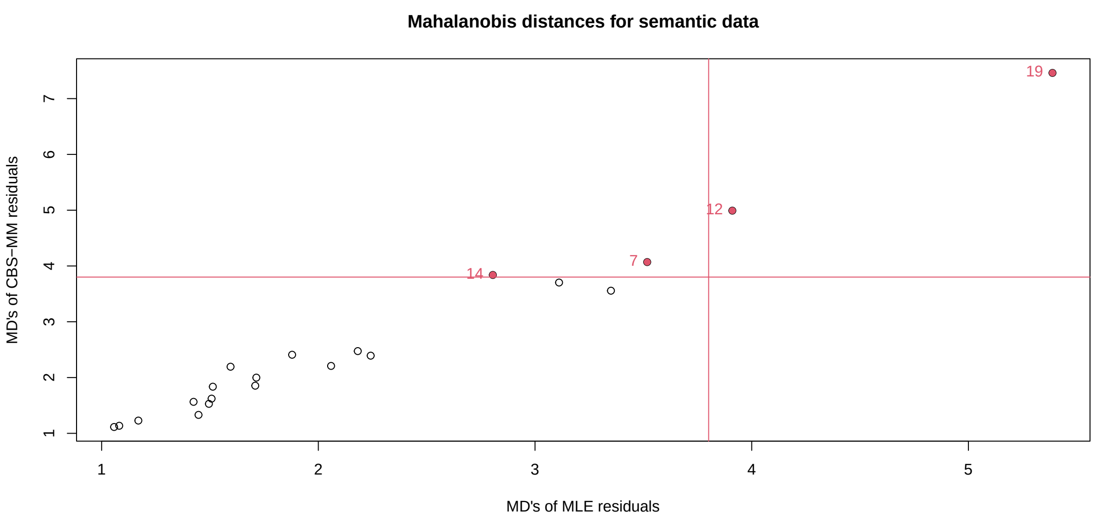

Figure 21: Mahalanobis distances Semantic Priming data

and 14) compared to the ones identified in Figure 4.6 in Heritier *et al* [5] (3, 8 and 16). However, looking at the data itself, see Figure 22, one can see that subjects 3, 8 and 16 do not correspond to outlying curves. An explanation may be that the reported negative variance estimates of the CBS estimator for $\boldsymbol{\theta}$ has destroyed the analysis?

On the other hand, in the interaction plots in Figure 23, we see that the curves for subject 19 (red) and 12 (green) clearly stick out, and somewhat less pronounced the curves for subjects 7 (light blue). and 14 (darkblue). Moreover, in the bottom-right interaction plot, is seems that there is an interaction effect between the factors DELAY and CONDITION.

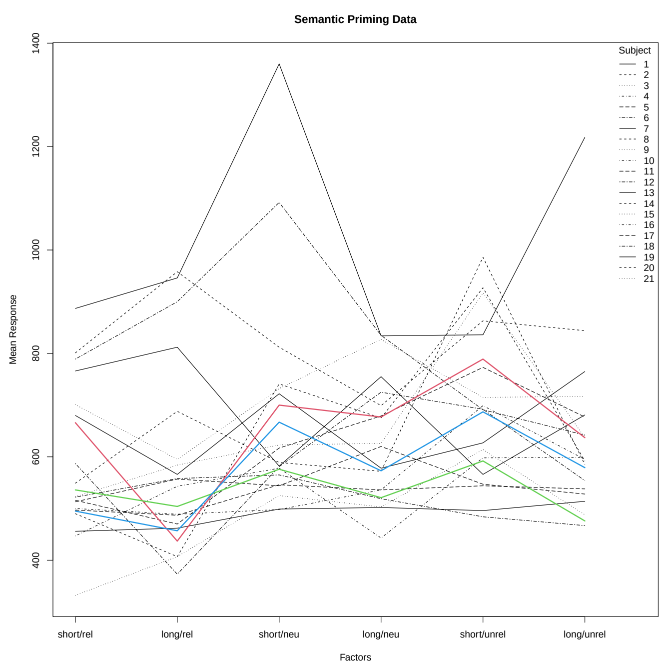

Figure 22: Semantic Priming Data

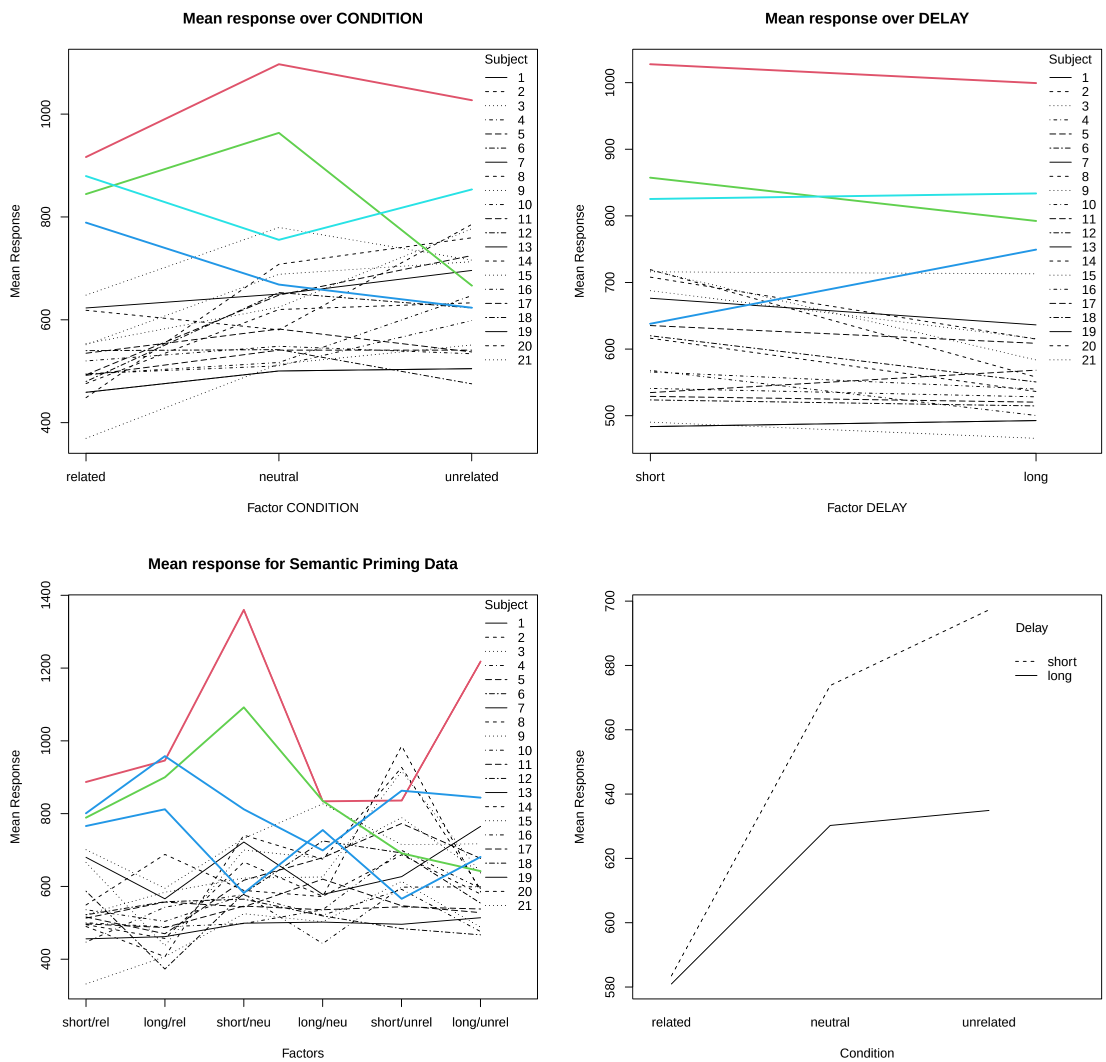

Figure 23: Semantic Priming Data

8 Orthodontic Growth Data

Consider the example in Section 4.2.4 in [6] about an orthodontic growth study, where a set of different measurements were collected from X-rays of 27 children's skulls (16 males and 11 females). The response variable is the distance in millimeters between the pituitary and the pterygomaxillary fissure, two points that can be easily located on the Xrays. The distance was measured at 8, 10, 12 and 14 years of age for each child.

In this case

$$\mathbf{y}_i = \begin{pmatrix} 1 & J_i & 8 & 8J_i \\ 1 & J_i & 10 & 10J_i \\ 1 & J_i & 12 & 12J_i \\ 1 & J_i & 14 & 14J_i \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_{0g} \\ \beta_1 \\ \beta_{1g} \end{pmatrix} + \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} \gamma_{0i} + \begin{pmatrix} 8 \\ 10 \\ 12 \\ 14 \end{pmatrix} \gamma_{1i} + \boldsymbol{\epsilon}_i, \quad i = 1, \dots, 27,$$

where $J_i = 0$, if the i-th subject is a boy and $J_i = 1$, if the i-th subject is a girl.

Here, the $\mathbf{X}_i$ are not the same for each individual, i.e.,

$$\mathbf{X}_i = \mathbf{X}_b = \begin{pmatrix} 1 & 0 & 8 & 0 \\ 1 & 0 & 10 & 0 \\ 1 & 0 & 12 & 0 \\ 1 & 0 & 14 & 0 \end{pmatrix}$$

for each subject that is a boy and

$$\mathbf{X}_i = \mathbf{X}_g = \begin{pmatrix} 1 & 1 & 8 & 8 \\ 1 & 1 & 10 & 10 \\ 1 & 1 & 12 & 12 \\ 1 & 1 & 14 & 14 \end{pmatrix}$$

for each subject that is a girl. The regression parameter is $\boldsymbol{\beta} = (\beta_0, \beta_{0g}, \beta_1, \beta_{1g})$.

There are two different covariance parameters corresponding to independent random effects, i.e., $\gamma_{0i} \sim N(0, \sigma^2_{\gamma_0})$ and $\gamma_{1i} \sim N(0, \sigma^2_{\gamma_1})$, and one for the measurement error $\boldsymbol{\epsilon}_i \sim N(\mathbf{0}, \sigma^2_\epsilon \mathbf{I}_4)$. Hence, the covariance parameter is $\boldsymbol{\theta} = (\sigma^2_{\gamma_0}, \sigma^2_{\gamma_1}, \sigma^2_\epsilon)$. Moreover, the design matrices of the random effects are

$$\mathbf{Z}_1 = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} \quad \text{and} \quad \mathbf{Z}_2 = \mathbf{t} = \begin{pmatrix} 8 \\ 10 \\ 12 \\ 14 \end{pmatrix}.$$

This leads to $\mathbf{L}_1 = \mathbf{Z}_1\mathbf{Z}_1^T$, $\mathbf{L}_2 = \mathbf{Z}_2\mathbf{Z}_2^T = \mathbf{t}\mathbf{t}^T$, and $\mathbf{L}_3 = \mathbf{I}_4$ and linear covariance structure

$$\begin{aligned} \mathbf{V}(\boldsymbol{\theta}) &= \theta_1\mathbf{L}_1 + \theta_2\mathbf{L}_2 + \theta_3\mathbf{L}_3 \\ &= \sigma^2_{\gamma_0}\mathbf{J}_6 + \sigma^2_{\gamma_1}\mathbf{t}\mathbf{t}^T + \sigma^2_\epsilon\mathbf{I}_4. \end{aligned}$$

8.1 Results in Copt [3] and Heritier *et al* [5]

The dataset has been analyzed by Copt [3], Heritier *et al* [5], and Demidenko [4]. Starting values will be obtained from (robust) estimators $\widehat{\boldsymbol{\mu}}$ and $\widehat{\mathbf{V}}$ for multivariate location and covariance. From this there are different options to obtain a starting value for $\boldsymbol{\beta}$.

1. Solving the fixed point equation for the regression parameter

$$\boldsymbol{\beta}^{(0)} = \left(\sum_{i=1}^n w_i \mathbf{X}_i^T \widehat{\mathbf{V}}^{-1} \mathbf{X}_i \right)^{-1} \sum_{i=1}^n w_i \mathbf{X}_i^T \widehat{\mathbf{V}}^{-1} \mathbf{y}_i$$

   where w_i are weights (0 or 1) obtained from the multivariate location-scatter estimates (for some robust estimates these are provided, otherwise take $w_i = 1$). Finally, take $\widehat{\mathbf{V}}$ as the starting value for $\mathbf{V}(\boldsymbol{\theta}^{(0)})$.

2. Note that $\mathbf{X}_b$ and $\mathbf{X}_g$ are non-singular, and cannot be used to invert $\widehat{\boldsymbol{\mu}}$ to a starting value $\boldsymbol{\beta}^{(0)}$. The starting value $\boldsymbol{\beta}^{(0)}$ can be determined from the location estimate as follows. Note that

$$\mathbf{X}_b\boldsymbol{\beta} = \begin{pmatrix} \beta_0 + 8\beta_1 \\ \beta_0 + 10\beta_1 \\ \beta_0 + 12\beta_1 \\ \beta_0 + 14\beta_1 \end{pmatrix} = \begin{pmatrix} 1 & 8 \\ 1 & 10 \\ 1 & 12 \\ 1 & 14 \end{pmatrix} \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix} = \mathbf{X}_0 \begin{pmatrix} \beta_0 \\ \beta_1 \end{pmatrix}$$

and

$$\mathbf{X}_g\boldsymbol{\beta} = \begin{pmatrix} \beta_0 + \beta_{0g} + 8(\beta_1 + \beta_{1g}) \\ \beta_0 + \beta_{0g} + 10(\beta_1 + \beta_{1g}) \\ \beta_0 + \beta_{0g} + 12(\beta_1 + \beta_{1g}) \\ \beta_0 + \beta_{0g} + 14(\beta_1 + \beta_{1g}) \end{pmatrix} = \begin{pmatrix} 1 & 8 \\ 1 & 10 \\ 1 & 12 \\ 1 & 14 \end{pmatrix} \begin{pmatrix} \beta_0 + \beta_{0g} \\ \beta_1 + \beta_{1g} \end{pmatrix} = \mathbf{X}_0 \begin{pmatrix} \beta_0 + \beta_{0g} \\ \beta_1 + \beta_{1g} \end{pmatrix}.$$

To determine the separate beta's, divide the data into two subsets, one with boys and one with girls. From the boys subset determine a multivariate location estimate $\widehat{\boldsymbol{\mu}}_b$, and similarly $\widehat{\boldsymbol{\mu}}_g$ from the subset of girls. Then take

$$\begin{pmatrix} \beta_0^{(0)} \\ \beta_1^{(0)} \end{pmatrix} = (\mathbf{X}_0^T\mathbf{X}_0)^{-1}\mathbf{X}_0^T\widehat{\boldsymbol{\mu}}_b$$

and

$$\begin{pmatrix} \beta_{0g}^{(0)} \\ \beta_{1g}^{(0)} \end{pmatrix} = (\mathbf{X}_0^T\mathbf{X}_0)^{-1}\mathbf{X}_0^T\widehat{\boldsymbol{\mu}}_g - \begin{pmatrix} \beta_0^{(0)} \\ \beta_1^{(0)} \end{pmatrix}.$$

Finally, take $\widehat{\mathbf{V}}$ as the starting value for $\mathbf{V}(\boldsymbol{\theta}^{(0)})$.

ML estimates. Heritier *et al* [5] do not report ML or REML estimators. Demidenko [4] models the data with a structured covariance that results from an AR(1) process. Therefore, we can only compare our ML estimates with Copt [3] and the output of the function `lme`.

Copt reports the ML estimates in Table 8-7. It seems that this can be obtained from the function `lme` with contrast SAS:

```
options(contrasts=c("contr.SAS","contr.poly"))
Dental=groupedData( distance ~ age | Subject,
                    data = as.data.frame( Dental ),
                    FUN = mean,
                    outer = ~ Sex,
                    labels = list( x = "Age",
                                   y = "Distance from pituitary to pterygomaxillary fissure" ),
                    units = list( x = "(yr)", y = "(mm)") )

DentalML.lme <- lme(distance~Sex*age, data = Dental,
                    random = list(Subject=pdDiag(~age)),method="ML")
```

With setting $M = 10000$ and re-scaling during the iteration with the translated biweight, I find the following for the ML and REML estimates

```
              betaMLrik betaMLCopt  betaMLlme betaREMLlme
(Intercept)   16.3406250      16.34 16.3406250  16.3406250
SexFemale      1.0321023       1.03  1.0321023   1.0321023
age            0.7843750       0.78  0.7843750   0.7843750
SexFemale:age -0.3048295      -0.30 -0.3048295  -0.3048295
```

and

```
   thetaMLrik thetaMLCopt  thetaMLlme thetaREMLlme
1 2.249224182       2.336 2.249223067  2.416802924
2 0.006757583       0.007 0.006757591  0.007746915
3 1.824211374       1.894 1.824212202  1.864595712
```

The ML estimates match perfectly with the output if `lme`. The ML estimates for $\boldsymbol{\beta}$ are the same as in Copt [3] and the ML estimates for $\boldsymbol{\theta}$ are quite close. To get the same results for ML as reported by `lme` one must choose `contr.SAS`. If one uses `contr.sum` our the estimates can be found as follows

```
est=fixed.effects(Dental.lme)
est[1]-est[2]
2*est[2]
est[3]-est[4]
2*est[4]
```

CTBS estimates. Copt [3] reports the CTBS estimates in Table 8.7. With the translated biweight with $M = 1.38092$ and $c = 2.262801$ and

1. without re-scaling during the iteration to satisfy the S-constraint,
2. not using Rocke's correction,
3. taking starting values from `rogkmiss`, retrieve $\boldsymbol{\beta}^{(0)}$ from 1-step iteration of regression fixed point and keep covariance matrix.

I find

```
  betaCTBSrik betaCTBSCopt
1  16.7979636        17.24
2   0.7615596         0.14
3   0.7069704         0.70
4  -0.2379409        -0.22
```

and

```
  thetaCTBSrik thetaCTBSCopt
1   1.85589466         2.812
2   0.01287646         0.012
3   1.00783288         1.032
```

The estimates are similar, but different. If I use Rocke's adaptation or obtain a starting value for $\boldsymbol{\beta}$ differently, I still get different estimates.

CBS-MM estimates. Heritier *et al* [5] reports the CBS-MM estimates in Table 4.10. I used the translated biweight with $M = 0$ and $c = 4.096567$ (which is the same as the biweight with $c = 4.096567$) and

1. no re-scaling during the iteration to satisfy the S-constraint,
2. using Rocke's correction,
3. taking starting values from `rogkmiss`, retrieve $\boldsymbol{\beta}^{(0)}$ from 1-step iteration of regression fixed point and keep covariance matrix.

I find CBS estimates for $\boldsymbol{\beta}$ and $\boldsymbol{\beta}$. It turns out that

```
  SQRTthetaCBSrik SQRTthetaCBSheritier
1       1.5841003                1.584
2       0.1149352                0.115
3       1.0402338                1.040
```

Afterwards, I computed the MM-estimates for $\boldsymbol{\beta}$ with the biweight function with $c_1 = 5.810343$ and

1. do not re-scale
2. use Rocke's adaptation

I find

```
  betaCBSrik betaCBSMMrik betaCBSMMheritier
1 17.1122953   17.3150622            17.395
2  0.5008058    0.1607241             0.080
3  0.6937364    0.6902533             0.581
4 -0.2342965   -0.2191102            -0.110
```

The estimates seem to differ from Heritier, but one can find our estimates be re-computing as before

```
est=betaCBSMMgrowthHeritier
est[1]-est[2]
2*est[2]
est[3]-est[4]
2*est[4]
```

Hence $\widehat{\boldsymbol{\beta}}_{rik} = \mathbf{A}\widehat{\boldsymbol{\beta}}_{heritier}$, where

$$\mathbf{A} = \begin{pmatrix} 1 & -1 & 0 & 0 \\ 0 & 2 & 0 & 0 \\ 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 2 \end{pmatrix} \quad \text{and} \quad \mathbf{A}^{-1} = \begin{pmatrix} 1 & 0.5 & 0 & 0 \\ 0 & 0.5 & 0 & 0 \\ 0 & 0 & 1 & 0.5 \\ 0 & 0 & 0 & 0.5 \end{pmatrix}$$

8.2 CTBS and CBS-MM estimates that satisfy the S-constraint.

I computed the CTBS and CBS-MM estimates with my own code. During the iteration to compute the S-estimates, I used

1. re-scaling of the Mahalanobis distances, such that they satisfy the S-constraint.
2. no Rocke's adaptation.

CTBS estimates. I used the translated biweight with $M = 1.38092$ and $c = 2.262801$. Starting values for $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$ based on several (robust) multivariate location-scale estimates (i.e., regular OGK, weighted OGK, `rogkmiss`, MCD, or the sample mean and sample covariance, in combination with both options to transfer to initial values) all lead to the same CTBS estimates. I find

```
  betaCTBSrescale betaCTBSCopt
1      16.9151528        17.24
2       0.6072116         0.14
3       0.7045369         0.70
4      -0.2336862        -0.22
```

and

```
  thetaCTBSrescale thetaCTBSCopt
1       2.21813187         2.812
2       0.01329487         0.012
3       1.05434677         1.032
```

The estimates are different, but similar in order of magnitude. The difference may be due to re-scaling to satisfy the S-constraint. We can compare the determinants of both solutions, after re-scaling them to satisfy the S-constraint. I find

```
  DetCTBSrescale DetCTBSCopt
1       22.11102    22.39952
```

I conclude that the CTBS estimator of Copt is not the optimal solution to the S-minimization problem.

CBS-MM estimates. I used the translated biweight with $M = 0$ and $c = 4.096567$ (which is the same as the biweight with $c = 4.096567$). Starting values for $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$ based on several multivariate location-scale estimates all lead to the same CBS estimates. Afterwards, I computed the MM-estimates for $\boldsymbol{\beta}$ with the biweight function with $c_1 = 6.017282$ and

1. no re-scaling
2. no Rocke's adaptation

For the CBS estimates and CBS-MM estimates, I find

```
  betaCBSrescale betaCBSMMrescale betaCBSMMheritier
1     17.0960329       17.3092961            17.395
2      0.5260926        0.1719783             0.080
3      0.6939917        0.6902045             0.581
4     -0.2352969       -0.2194966            -0.110
```

and

```
  SQRTthetaCBSrescale SQRTthetaCBSheritier
1           1.5536252                1.584
2           0.1135392                0.115
3           1.0261985                1.040
```

The estimates for $\boldsymbol{\theta}$ differ, but are similar. Differences may be due to my re-scaling. The estimates for $\boldsymbol{\beta}$ can be compared by first computing

```
est=betaCBSMMgrowthHeritier
est[1]-est[2]
2*est[2]
est[3]-est[4]
2*est[4]
```

This gives different estimates, but similar. Differences may be due to my re-scaling.

If I compare the determinants of the CBS estimators of the first step, after re-scaling them to satisfy the S-constraint, I find

```
  DetCBSrescale DetCBSHeritier
1      22.85376       22.85454
```

I conclude that the CBS estimator of Heritier is not the optimal solution to the S-minimization problem.

8.3 Standard Errors.

The asymptotic covariance is given by

$$\gamma \left(\mathbb{E}\left[\mathbf{X}^T \boldsymbol{\Sigma}^{-1} \mathbf{X} \right] \right)^{-1}, \quad \gamma = \frac{\mathbb{E}_{\mathbf{0},\mathbf{I}_k}\left[\rho'(\|\mathbf{z}\|)^2 \right]}{k\alpha^2}.$$

In this case the $\mathbf{X}_i$'s are different. This means we can estimate $\mathbb{E}\left[\mathbf{X}^T \boldsymbol{\Sigma}^{-1} \mathbf{X} \right]$ by

$$\frac{1}{n}\sum_{i=1}^{n} \mathbf{X}_i^T \mathbf{V}(\widehat{\boldsymbol{\theta}})^{-1} \mathbf{X}_i = \widehat{p}_b(\mathbf{X}_1^T \mathbf{V}(\widehat{\boldsymbol{\theta}})^{-1} \mathbf{X}_1) + \widehat{p}_g(\mathbf{X}_2^T \mathbf{V}(\widehat{\boldsymbol{\theta}})^{-1} \mathbf{X}_2)$$

where $\widehat{p}_b = 16/27$ and $\widehat{p}_g = 11/27$ are the estimated proportions of boys and girls in the sample.

The asymptotic variance from Copt and Victoria-Feser [2] is only given for situations where $\mathbf{X}_i = \mathbf{X}$, for all $i = 1, \ldots, n$, by

$$\frac{\mathbb{E}_{\mathbf{0},\mathbf{I}_k}\left[\rho'(\|\mathbf{z}\|)^2 \right]}{k\alpha^2} (\mathbf{X}^T\mathbf{X})^{-1} \mathbf{X}^T \boldsymbol{\Sigma} \mathbf{X} (\mathbf{X}^T\mathbf{X})^{-1},$$

which is the wrong expression for general $\mathbf{X}$. In the case of different $\mathbf{X}_i$, Copt [3] proposes to estimate $(\mathbf{X}^T\mathbf{X})^{-1} \mathbf{X}^T \boldsymbol{\Sigma} \mathbf{X} (\mathbf{X}^T\mathbf{X})^{-1}$ by

$$n \left(\sum_{i=1}^{n} \mathbf{X}_i^T \mathbf{X}_i \right)^{-1} \left(\sum_{i=1}^{n} \mathbf{X}_i^T \mathbf{V}(\widehat{\boldsymbol{\theta}}) \mathbf{X}_i \right) \left(\sum_{i=1}^{n} \mathbf{X}_i^T \mathbf{X}_i \right)^{-1}$$

which he (falsely) claims to be equal (????) to

$$n \left(\sum_{i=1}^{n} \mathbf{X}_i^T \mathbf{V}(\widehat{\boldsymbol{\theta}})^{-1} \mathbf{X}_i \right)^{-1}$$

see his page 78. In this way, he accidently arrives at the correct expression.

ML estimates. For the ML estimates, I find $\gamma = 1$, and for the asymptotic covariance matrix of $\sqrt{n}(\widehat{\boldsymbol{\beta}}_{ML} - \boldsymbol{\beta})$,

```
 23.189213 -23.189213 -1.6930962  1.6930962
-23.189213  56.918978  1.6930962 -4.1557815
 -1.693096   1.693096  0.1653213 -0.1653213
  1.693096  -4.155782 -0.1653213  0.4057885
```

The SE's can be found by diving the main diagonal by $n = 27$ and take the square root. Copt [3] reports the SE's in Table 8-7. Furthermore, we can obtain standard errors from the output of `lme`. This gives

```
  SEbetaMLErik SEbetaMLEcopt SEbetaMLElme
1   0.92674686         1.139    0.9408690
2   1.45193329         1.480    1.4740585
3   0.07824966         0.096    0.0794421
4   0.12259366         0.125    0.1244618
```

These values are different, but comparable in order of magnitude. Our values are most similar to the ones from `lme`. It is unclear how the standard errors have been obtained by Copt.

CTBS estimates. For the CTBS estimator I find $\gamma = 1.271367$, and for the asymptotic covariance matrix of $\sqrt{n}(\widehat{\boldsymbol{\beta}}_{ML} - \boldsymbol{\beta})$,

```
 19.009632 -19.009632 -1.2441157  1.2441157
-19.009632  46.660007  1.2441157 -3.0537386
 -1.244116   1.244116  0.1416247 -0.1416247
  1.244116  -3.053739 -0.1416247  0.3476242
```

The SE's can be found by diving the main diagonal by $n = 27$ and take the square root.

```
SEbetaCTBSrik SEbetaCTBSCopt
1     0.83908311           1.037
2     1.31459058           1.347
3     0.07242484           0.086
4     0.11346792           0.112
```

Our values are different, but comparable to the ones in Copt [3].

CBS-MM estimates. For the CBS-MM estimator, with $c_1 = 5.810343$, I find $\gamma = 1.05263$, and for the asymptotic covariance matrix of $\sqrt{n}(\widehat{\boldsymbol{\beta}}_{ML} - \boldsymbol{\beta})$,

```
 16.072403 -16.072403 -1.0288337  1.0288337
-16.072403  39.450443  1.0288337 -2.5253190
 -1.028834   1.028834  0.1164291 -0.1164291
  1.028834  -2.525319 -0.1164291  0.2857804
```

The SE's can be found by diving the main diagonal by $n = 27$ and take the square root.

```
  SEbetaCBSMMrik SEbetaCBSMMgrowthHeritier
1      0.77154014                     0.613
2      1.20877109                     0.613
3      0.06566725                     0.052
4      0.10288080                     0.052
```

These values deviate a lot from the ones in Heritier *et al* [5]. The reason why the standard errors for $\widehat{\beta}_0$ and $\widehat{\beta}_{0g}$ are the same and similarly for $\widehat{\beta}_1$ and $\widehat{\beta}_{1g}$, can be explained by the fact that $\widehat{\boldsymbol{\beta}}_{heritier} = \mathbf{A}^{-1}\widehat{\boldsymbol{\beta}}_{rik}$, where

$$\mathbf{A}^{-1} = \begin{pmatrix} 1 & 0.5 & 0 & 0 \\ 0 & 0.5 & 0 & 0 \\ 0 & 0 & 1 & 0.5 \\ 0 & 0 & 0 & 0.5 \end{pmatrix}$$

which means that $\mathrm{cov}(\widehat{\boldsymbol{\beta}}_{heritier}) = \mathbf{A}^{-1}\mathrm{cov}(\widehat{\boldsymbol{\beta}}_{rik})\mathbf{A}^{-T}$. When we apply this to our results, we find the following standard errors for $\widehat{\boldsymbol{\beta}}_{heritier}$.

```
0.60438581
0.60438581
0.05144043
0.05144043
```

This is similar to Heritier, but slightly different. The remaining differences may be due to our use of re-scaling.

8.4 Summary

Putting everything together, for the ML estimates I find the following summary

```
         MLE SE.MLE       TMLE PvalMLE PvalMLCopt
1 16.3406250  0.927 17.6322421  0.0000      0.000
2  1.0321023  1.452  0.7108469  0.4772      0.486
3  0.7843750  0.078 10.0240050  0.0000      0.000
4 -0.3048295  0.123 -2.4865034  0.0129      0.016
```

The p-values are similar to the ones in Copt [3]. The output fro the ML estimators from `lme` is

```
                  Value Std.Error DF   t-value p-value
(Intercept)   16.340625 0.9444007 79 17.302639  0.0000
SexFemale      1.032102 1.4795916 25  0.697559  0.4919
age            0.784375 0.0797403 79  9.836624  0.0000
SexFemale:age -0.304830 0.1249290 79 -2.440023  0.0169
```

We see that our standard errors are somewhat smaller leading to larger values of the Wald test statistic. The difference in p-values is partly explained by the fact that `lme` computes them with the t-distribution instead of the $N(0,1)$.

For the CTBS estimates I find

```
        CTBS SE.CTBS      TCTBS PvalCTBS PvalCTBScopt
1 16.9151528   0.839 20.1590917   0.0000       0.0000
2  0.6072116   1.315  0.4619017   0.6442       0.9172
3  0.7045369   0.072  9.7278346   0.0000       0.0000
4 -0.2336862   0.113 -2.0594917   0.0394       0.0495
```

The p-values are similar to the ones in Copt [3].

For the CBS-MM estimates I find

```
       CBSMM SE.CBSMM     TCBSMM PvalCBS PvalCBSheritier
1 17.3092961    0.772 22.4347219  0.0000          0.0000
2  0.1719783    1.209  0.1422753  0.8869          0.8962
3  0.6902045    0.066 10.5106309  0.0000          0.0000
4 -0.2194966    0.103 -2.1335034  0.0329          0.0344
```

Because $\widehat{\boldsymbol{\beta}}_{heritier} = \mathbf{A}^{-1}\widehat{\boldsymbol{\beta}}_{rik}$, we cannot compare our results directly with the ones in Heritier *et al* [5]. If we apply $\widehat{\boldsymbol{\beta}}_{heritier} = \mathbf{A}^{-1}\widehat{\boldsymbol{\beta}}_{rik}$ to our results, we find

```
       CTBSMM SE.CTBSMM     TCBSMM PvalCTBS
1 17.39321860     0.604 28.7783370   0.0000
2  0.08419434     0.604  0.1393056   0.8892
3  0.58073531     0.051 11.2894732   0.0000
4 -0.10955964     0.051 -2.1298354   0.0332
```

These p-values are very similar to the ones in Heritier *et al* [5] and the remaining difference may be explained by our use of re-scaling the CBS estimator.

All conclusions are in line with the model used in Demidenko [4] for the same dataset, where β_{1g} is taken zero, see model equation (4.122), and seems to be confirmed by his Figure 4.3, see also Figure 24.

The robust Mahalanobis distances for the residuals detect observation 20 and 24 as outliers, see Figure 25. This is in line with Figure 8-16 in Copt [3]. A closer look at the data seems to confirm these observations as outliers, see Figure 24 (20 in red, 24 in green). The analysis in Heritier *et al* [5] deviates completely from these findings. They look at the individual standardized residuals and identify female observations 10 and 11 as outliers, as well as boys observations 20, 21 and 24 (see their Figure 4.14).

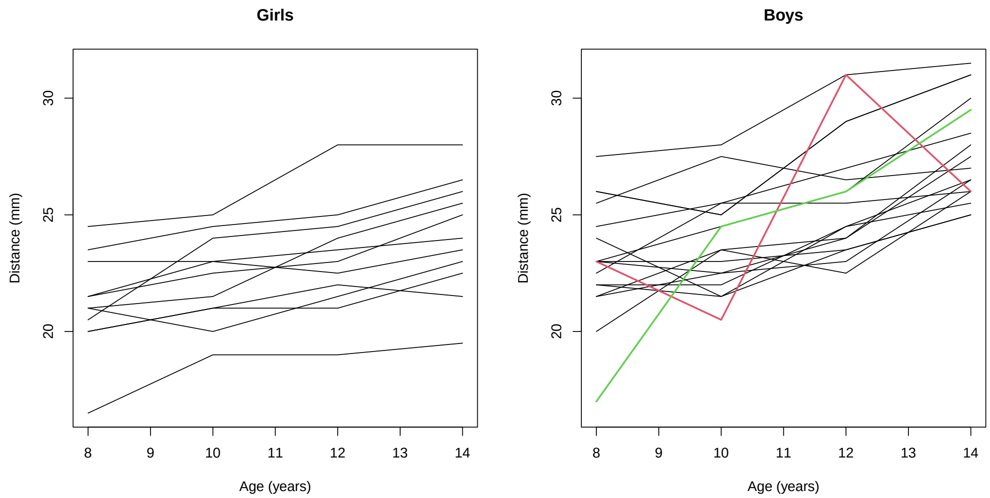

Figure 24: Growth data

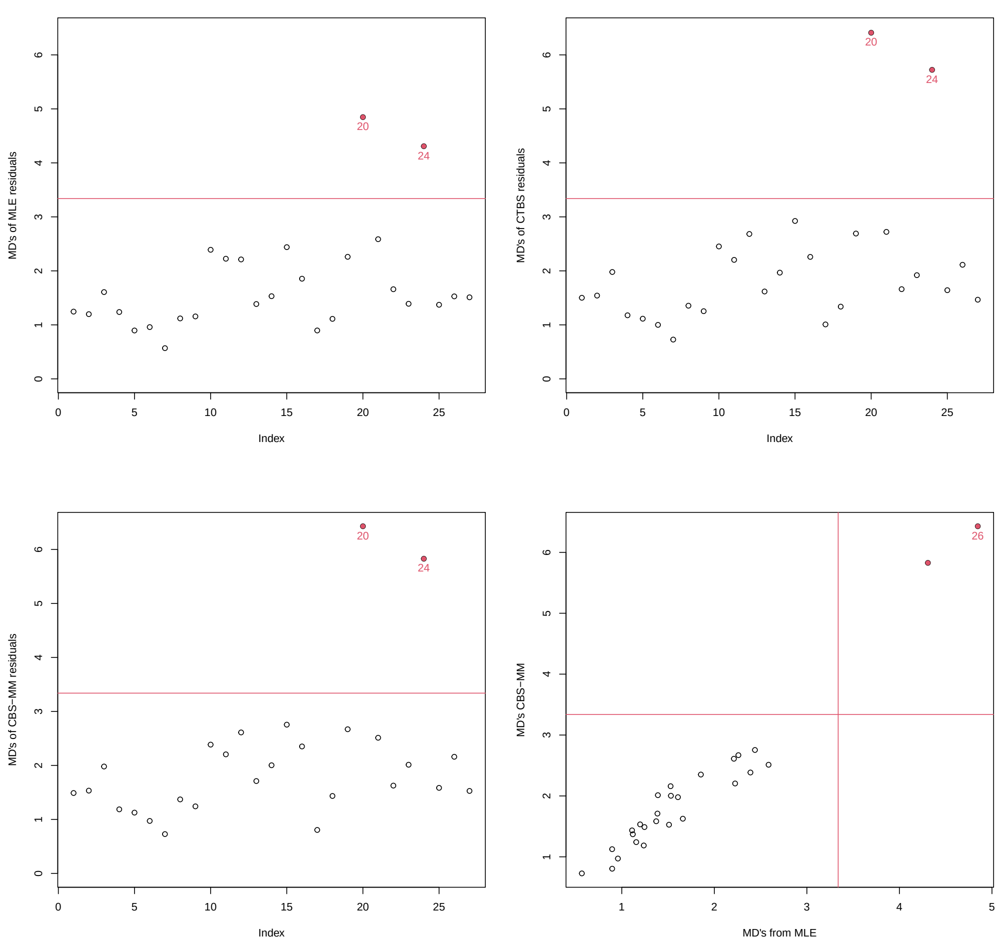

Figure 25: Robust MD's for Growth data

9 Metallic Oxide data.

This example is discussed in Section 8.3 in Copt [3] and in Section 4.7.1 in Heritier *et al* [5]. The data originates from a sampling study designed to explore the effects of process and measurement variation on the properties of lots of metallic oxides. Two samples were drawn from each lot. Duplicate analyses were then performed by each of two chemists, with a pair of chemists randomly selected for each sample.

The model can be written as

$$\mathbf{y}_i = \begin{pmatrix} 1 & J_i \\ 1 & J_i \\ 1 & J_i \\ 1 & J_i \\ 1 & J_i \\ 1 & J_i \\ 1 & J_i \\ 1 & J_i \end{pmatrix} \begin{pmatrix} \mu \\ \lambda \end{pmatrix} + \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{pmatrix} \gamma_i + \begin{pmatrix} 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 1 & 0 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \delta_{1i} \\ \delta_{2i} \end{pmatrix} + \begin{pmatrix} 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 \\ 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \xi_{11i} \\ \xi_{12i} \\ \xi_{21i} \\ \xi_{22i} \end{pmatrix} + \boldsymbol{\epsilon}_i$$

where

$$J_i = \begin{cases} 0, & \text{if the } i\text{-th measurement is of type 1} \\ 1, & \text{if the } i\text{-th measurement is of type 2.} \end{cases}$$

Here, the $\mathbf{X}_i$ are not the same for each subject, i.e.,

$$\mathbf{X}_i = \mathbf{X}_1 = [\mathbf{e}_8 \ \mathbf{0}]$$

if the i-th lot has metallic oxide type 1 and

$$\mathbf{X}_i = \mathbf{X}_2 = [\mathbf{e}_8 \ \mathbf{e}_8]$$

if the i-th lot has metallic oxide type 2. The regression parameter of fixed effects is $\boldsymbol{\beta} = (\mu, \lambda)$.

There are three different covariance parameters corresponding to the independent random effects, i.e.,

$$\gamma_i \sim N(0, \sigma_\gamma^2),$$

the random (intercept) effect for the i-th lot,

$$\boldsymbol{\delta}_i = (\delta_{i1}, \delta_{i2}) \sim N(\mathbf{0}, \sigma_\delta^2 \mathbf{I}_2),$$

the random effect for the i-th lot due to the sample (δ_{1i} for sample 1 drawn from the i-th lot and δ_{1i} for sample 2 drawn from the i-th lot), and

$$\boldsymbol{\xi}_i = (\xi_{11i}, \xi_{12i}, \xi_{21i}, \xi_{22i}) \sim N(\mathbf{0}, \sigma_\xi^2 \mathbf{I}_4),$$

the random effect for the i-th lot due to the chemist (ξ_{11i} for chemist 1 analyzing sample 1 drawn from the i-th lot, ξ_{12i} for chemist 2 analyzing sample 1 drawn from the i-th lot, ξ_{21i} for chemist 1 analyzing sample 2 drawn from the i-th lot, and ξ_{22i} for chemist 2 analyzing sample 2 drawn from the i-th lot) Hence, the covariance parameter is $\boldsymbol{\theta} = (\sigma_\gamma^2, \sigma_\delta^2, \sigma_\xi^2, \sigma_\epsilon^2)$.

This leads to

$$\begin{aligned} \mathbf{L}_1 &= \mathbf{Z}_1 \mathbf{Z}_1^T = \mathbf{J}_8 \\ \mathbf{L}_2 &= \mathbf{Z}_2 \mathbf{Z}_2^T = \mathbf{I}_2 \otimes \mathbf{J}_4 \\ \mathbf{L}_3 &= \mathbf{I}_4 \otimes \mathbf{J}_2 \\ \mathbf{L}_4 &= \mathbf{I}_8 \end{aligned}$$

and linear covariance structure

$$\begin{aligned} \mathbf{V}(\boldsymbol{\theta}) &= \theta_1 \mathbf{L}_1 + \theta_2 \mathbf{L}_2 + \theta_3 \mathbf{L}_3 + \theta_4 \mathbf{L}_4 \\ &= \sigma_\gamma^2 \mathbf{J}_8 + \sigma_\delta^2 \mathbf{I}_2 \otimes \mathbf{J}_4 + \sigma_\xi^2 \mathbf{I}_4 \otimes \mathbf{J}_2 + \sigma_\epsilon^2 \mathbf{I}_8. \end{aligned}$$

One may check that in the dataframe `Metal`, the order of the observations in the i-th lot

```
  Response Lots  Sample  Chemist  Type
1      4.1    1 Sample1 Chemist1 type1
2      4.0    1 Sample1 Chemist1 type1
3      4.3    1 Sample1 Chemist2 type1
4      4.3    1 Sample1 Chemist2 type1
5      4.1    1 Sample2 Chemist1 type1
6      4.0    1 Sample2 Chemist1 type1
7      4.1    1 Sample2 Chemist2 type1
8      4.1    1 Sample2 Chemist2 type1
```

corresponds with the chosen design matrices, so we do not have re-order the observations.

9.1 Results in Copt [3] and Heritier *et al* [5]

Copt [3] reports the ML estimates in Table 8.6. Heritier *et al* [5] do not report the ML of REML estimates. This means we can only compare with Copt and with the output of `lme`.

ML estimates. The ML estimates can be obtained from `lme` with the option `contr.treatment`

```
options(contrasts=c("contr.treatment","cont.poly"))
metallicML.lme=lme(fixed=Response~Type,
                   data=Metal,
                   random=list(Lots=pdBlocked(list(pdIdent(~1),
                                                   pdIdent(~Sample-1),
                                                   pdIdent(~Chemist-1)))),
                   method="ML")
```

With setting $M = 10000$ and re-scaling during the iteration with the translated biweight, I find the following for the ML estimates

```
            betaMLErik betaMLEcopt betaMLElme
(Intercept)    3.85625     3.85625    3.85625
Typetype2     -0.79375    -0.79375   -0.79375
```

and

```
  thetaMLErik thetaMLEcopt thetaMLElme
1  0.56535533        0.565  0.55517788
2  0.04330645        0.043  0.05733065
3  0.03189516        0.032  0.02035484
4  0.04322581        0.043  0.05091937
```

These estimates all coincide with the reported values in Table 8.6 in Copt [3]. The difference with the output of `lme` may be to using the wrong options.

CTBS estimates. Copt [3] reports the CTBS estimates in Table 8.7. With the translated biweight with $M = 3.341616$ and $c = 1.140597$ and

1. **with** re-scaling during the iteration to satisfy the S-constraint,
2. not using Rocke's correction,
3. taking starting values from `rogkmiss`, retrieve $\boldsymbol{\beta}^{(0)}$ from 1-step iteration of regression fixed point and keep covariance matrix.

I find

```
  betaCTBSrik betaCTBSCopt
1   3.9318992    3.9318973
2  -0.4017372   -0.4017356
```

and

```
  thetaCTBSrik thetaCTBSCopt
1   0.09706646         0.097
2   0.01194283         0.012
3   0.04028991         0.040
4   0.03555281         0.036
```

These estimates match perfectly with the ones reported in Table 8.6 in Copt [3]. If I do not re-scale I get similar values, but different.

CBS-MM. Heritier *et al* [5] reports the CBS-MM estimates in Table 4.9. I used the translated biweight with $M = 0$ and $c = 6.017282$ (which is the same as the biweight with $c = 6.017282$) and

1. no re-scaling during the iteration to satisfy the S-constraint,
2. using Rocke's correction,
3. taking starting values from `rogkmiss`, retrieve $\boldsymbol{\beta}^{(0)}$ from 1-step iteration of regression fixed point and keep covariance matrix.

I find CBS estimates for $\boldsymbol{\beta}$ and $\boldsymbol{\beta}$. It turns out that

```
  SQRTthetaCBSrik SQRTthetaCBSHeritier
1       0.3168768                0.317
2       0.1441247                0.144
3       0.1880997                0.188
4       0.1862799                0.186
```

Afterwards, I computed the MM-estimates for $\boldsymbol{\beta}$ with the biweight function with $c_1 = 6.818176$. I find

```
  betaCBSrik betaCBSMMrik betaCBSMMheritier
1  3.9205309    3.9095117             3.726
2 -0.3714336   -0.3672684             0.184
```

The estimates seem to differ from Heritier, but one can find our estimates by re-computing

```
est=betaCBSMMmetalHeritier
est[1]+est[2]
-2*est[2]
```

Hence $\widehat{\boldsymbol{\beta}}_{rik} = \mathbf{B}\widehat{\boldsymbol{\beta}}_{heritier}$, where

$$\mathbf{B} = \begin{pmatrix} 1 & 1 \\ 0 & -2 \end{pmatrix} \quad \text{and} \quad \mathbf{B}^{-1} = \begin{pmatrix} 1 & 0.5 \\ 0 & -0.5 \end{pmatrix}$$

9.2 CTBS and CBS-MM estimates that satisfy the S-constraint.

I computed the CTBS and CBS-MM estimates with my own code. During the iteration to compute the S-estimates, I used

1. re-scaling of the Mahalanobis distances, such that they satisfy the S-constraint.
2. no Rocke's adaptation.

CTBS estimates. I used the translated biweight with $M = 3.341616$ and $c = 1.140597$. Starting values for $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$ based on several (robust) multivariate location-scale estimates (i.e., regular OGK, weighted OGK, `rogkmiss`, MCD, or the sample mean and sample covariance, in combination with both options to transfer to initial values) all lead to the same CTBS estimates. I find

```
  betaCTBSrescale betaCTBSCopt
1       3.9318990    3.9318973
2      -0.4017371   -0.4017356
```

and

```
  thetaCTBSrescale thetaCTBSCopt
1       0.09706661         0.097
2       0.01194305         0.012
3       0.04028956         0.040
4       0.03555288         0.036
```

These estimates match perfectly with the ones reported in Copt [3].

CBS-MM estimates. I used the translated biweight with $M = 0$ and $c = 6.017282$ (which is the same as the biweight with $c = 6.017282$). Starting values for $\boldsymbol{\beta}$ and $\boldsymbol{\theta}$ based on several multivariate location-scale estimates all lead to the same CBS estimates. Afterwards, I computed the MM-estimates for $\boldsymbol{\beta}$ with the biweight function with $c_1 = 6.818171$ and

1. no re-scaling
2. no Rocke's adaptation

For the CBS estimates and CBS-MM estimates, I find

```
  betaCBSrescale betaCBSMMrescale betaCBSMMheritier
1      3.9198540        3.9087350             3.726
2     -0.3713213       -0.3669094             0.184
```

and

```
  SQRTthetaCBSrescale SQRTthetaCBSHeritier
1           0.3188276                0.317
2           0.1459191                0.144
3           0.1894468                0.188
4           0.1879771                0.186
```

The estimates for $\boldsymbol{\theta}$ differ, but are similar. Differences may be due to my re-scaling. The estimates for $\boldsymbol{\beta}$ can be compared by first computing

```
est=betaCBSMMmetalHeritier
est[1]+est[2]
-2*est[2]
```

which gives

```
    betaCBSMMmetalHeritierAdapted
                             3.91
                           -0.368
```

This is still different but very similar to Heritier. Remaining differences may be explained by our use of re-scaling.

9.3 Standard errors

The asymptotic covariance is

$$\gamma\left(\mathbb{E}\left[\mathbf{X}^T\mathbf{\Sigma}^{-1}\mathbf{X}\right]\right)^{-1}, \quad \gamma = \frac{\mathbb{E}_{\mathbf{0},\mathbf{I}_k}\left[\rho'(\|\mathbf{z}\|)^2\right]}{k\alpha^2}.$$

In this case the $\mathbf{X}_i$'s are different. This means we can estimate $\mathbb{E}\left[\mathbf{X}^T\mathbf{\Sigma}^{-1}\mathbf{X}\right]$ by

$$\frac{1}{n}\sum_{i=1}^{n}\mathbf{X}_i^T\mathbf{V}(\widehat{\boldsymbol{\theta}})^{-1}\mathbf{X}_i = \widehat{p}_1(\mathbf{X}_1^T\mathbf{V}(\widehat{\boldsymbol{\theta}})^{-1}\mathbf{X}_1) + \widehat{p}_2(\mathbf{X}_2^T\mathbf{V}(\widehat{\boldsymbol{\theta}})^{-1}\mathbf{X}_2)$$

where $\widehat{p}_1 = 18/31$ and $\widehat{p}_2 = 13/31$ are the estimated proportions of type 1 and type 2 in the sample.

The asymptotic variance from Copt and Victoria-Feser [2] is only given for situations where $\mathbf{X}_i = \mathbf{X}$, for all $i = 1, \dots, n$, by

$$\frac{\mathbb{E}_{\mathbf{0},\mathbf{I}_k}\left[\rho'(\|\mathbf{z}\|)^2\right]}{k\alpha^2}(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{\Sigma}\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1},$$

which is the wrong expression for general $\mathbf{X}$. In the case of different $\mathbf{X}_i$, Copt [3] proposes to estimate $(\mathbf{X}^T\mathbf{X})^{-1}\mathbf{X}^T\mathbf{\Sigma}\mathbf{X}(\mathbf{X}^T\mathbf{X})^{-1}$ by

$$n\left(\sum_{i=1}^{n}\mathbf{X}_i^T\mathbf{X}_i\right)^{-1}\left(\sum_{i=1}^{n}\mathbf{X}_i^T\mathbf{V}(\widehat{\boldsymbol{\theta}})\mathbf{X}_i\right)\left(\sum_{i=1}^{n}\mathbf{X}_i^T\mathbf{X}_i\right)^{-1}$$

which he (falsely) claims to be equal (????) to

$$n\left(\sum_{i=1}^{n}\mathbf{X}_i^T\mathbf{V}(\widehat{\boldsymbol{\theta}})^{-1}\mathbf{X}_i\right)^{-1}$$

see his page 78. In this way, he accidently arrives at the correct expression.

ML estimates. For the ML estimates I find $\gamma = 1$, and for the asymptotic covariance matrix of $\sqrt{n}(\widehat{\boldsymbol{\beta}}_{ML} - \boldsymbol{\beta})$,

```
 1.033997 -1.033997
-1.033997  2.465686
```

The SE's can be found by diving the main diagonal by $n = 31$ and take the square root. Copt [3] reports the SE's in Table 8-6. Furthermore, we can obtain standard errors from the output of `lme`. This gives

```
  SEbetaMLErik SEbetaMLEcopt SEbetaMLElme
1    0.1826328     0.1826328    0.1833738
2    0.2820253     0.2820253    0.2820253
```

This matches perfectly with Copt and is very similar to the output of `lme`.

CTBS estimates. For the CTBS estimator I find $\gamma = 1.093384$, and for the asymptotic covariance matrix of $\sqrt{n}(\widehat{\boldsymbol{\beta}}_{ML} - \boldsymbol{\beta})$,

```
 0.2213612 -0.2213612
-0.2213612  0.5278614
```

The SE's can be found by diving the main diagonal by $n = 27$ and take the square root.

```
  SEbetaCTBSrescale SEbetaCTBSCopt
1        0.08450257     0.09964693
2        0.13049056     0.15387677
```

Our values somewhat smaller, but comparable to the ones in Copt [3]. Differences may be due to our re-scaling of the CTBS estimator.

CBS-MM estimates. For the CBS-MM estimator, with $c_1 = 6.818171$, I find $\gamma = 1.05263$, and for the asymptotic covariance matrix of $\sqrt{n}(\widehat{\boldsymbol{\beta}}_{ML} - \boldsymbol{\beta})$,

```
 0.2278531 -0.2278531
-0.2278531  0.5433420
```

The SE's can be found by diving the main diagonal by $n = 31$ and take the square root.

```
  SEbetaCBSMMrescale SEbetaCBSMMHeritier
1         0.08573272               0.066
2         0.13239018               0.066
```

These values deviate a lot from the ones in Heritier *et al* [5]. The reason why the standard errors for $\widehat{\mu}$ and $\widehat{\lambda}$ are the same, can be explained by the fact that $\widehat{\boldsymbol{\beta}}_{heritier} = \mathbf{B}^{-1}\widehat{\boldsymbol{\beta}}_{rik}$, where

$$\mathbf{B}^{-1} = \begin{pmatrix} 1 & 0.5 \\ 0 & -0.5 \end{pmatrix}$$

which means that $\text{cov}(\widehat{\boldsymbol{\beta}}_{heritier}) = \mathbf{B}^{-1}\text{cov}(\widehat{\boldsymbol{\beta}}_{rik})\mathbf{B}^{-T}$. When we apply this to our results, we find the following standard errors for $\widehat{\boldsymbol{\beta}}_{heritier}$.

```
0.06619509
0.06619509
```

This is similar to Heritier, but slightly different, which matches perfectly.

9.4 Summary

Putting everything together, for the ML estimates I find the following summary for the ML estimates

```
       MLE SE.MLE      TMLE PvalMLE PvalMLCopt
1  3.85625  0.183 21.114768  0.0000      0.000
2 -0.79375  0.282 -2.814464  0.0049      0.005
```

whereas the output of `lme` gives

```
               Value Std.Error  DF   t-value p-value
(Intercept)  3.85625 0.1833738 217 21.029455  0.0000
Typetype2   -0.79375 0.2831694  29 -2.803092  0.0089
```

The difference with `lme` is due to the fact that `lme` uses the t-distribution to compute p-values.

For the CTBS estimators I find

```
        CTBS SE.CTBS     TCTBS PvalCTBS PvalCTBScopt
1  3.9318990   0.085 46.529932   0.0000        0.000
2 -0.4017371   0.130 -3.078668   0.0021        0.009
```

Our smaller p-value for λ is probably due to the smaller standard errors caused by re-scaling the CTBS estimator.

For the CBS-MM estimators I find

```
       CBSMM SE.CBSMM    TCBSMM PvalCBS PvalCBSheritier
1  3.9087350    0.086 45.592102  0.0000          0.0000
2 -0.3669094    0.132 -2.771425  0.0056          0.0053
```

Because $\widehat{\boldsymbol{\beta}}_{heritier} = \mathbf{B}^{-1}\widehat{\boldsymbol{\beta}}_{rik}$, we cannot compare our results directly with the ones in Heritier *et al* [5]. If we apply $\widehat{\boldsymbol{\beta}}_{heritier} = \mathbf{B}^{-1}\widehat{\boldsymbol{\beta}}_{rik}$ to our results, we find

```
      CBSMM SE.CBSMM    TCBSMM PvalCBSMM
1 3.7252803    0.066 56.277289    0.0000
2 0.1834547    0.066  2.771425    0.0056
```

These p-values are very similar to the ones in Heritier *et al* [5] and the remaining difference may be explained by our use of re-scaling the CBS estimator.

In all cases, the conclusion is the same, namely that both μ and λ are significantly different from zero. The robust Mahalanobis distances identify items 24 and 25 as outliers, and possibly items 12, 17 and 30, see Figure 26. This coincides with the results found in Copt [3] (see his Figure 8-13) and in Heritier *et al* [5] (see Figure 4.11). Although they conclude that observations 24 and 30 are outliers, and that 12 and 25 are influential. In a coplot of the data observations 24 (red) and 25 (green) clearly stick out, see Figure 27.

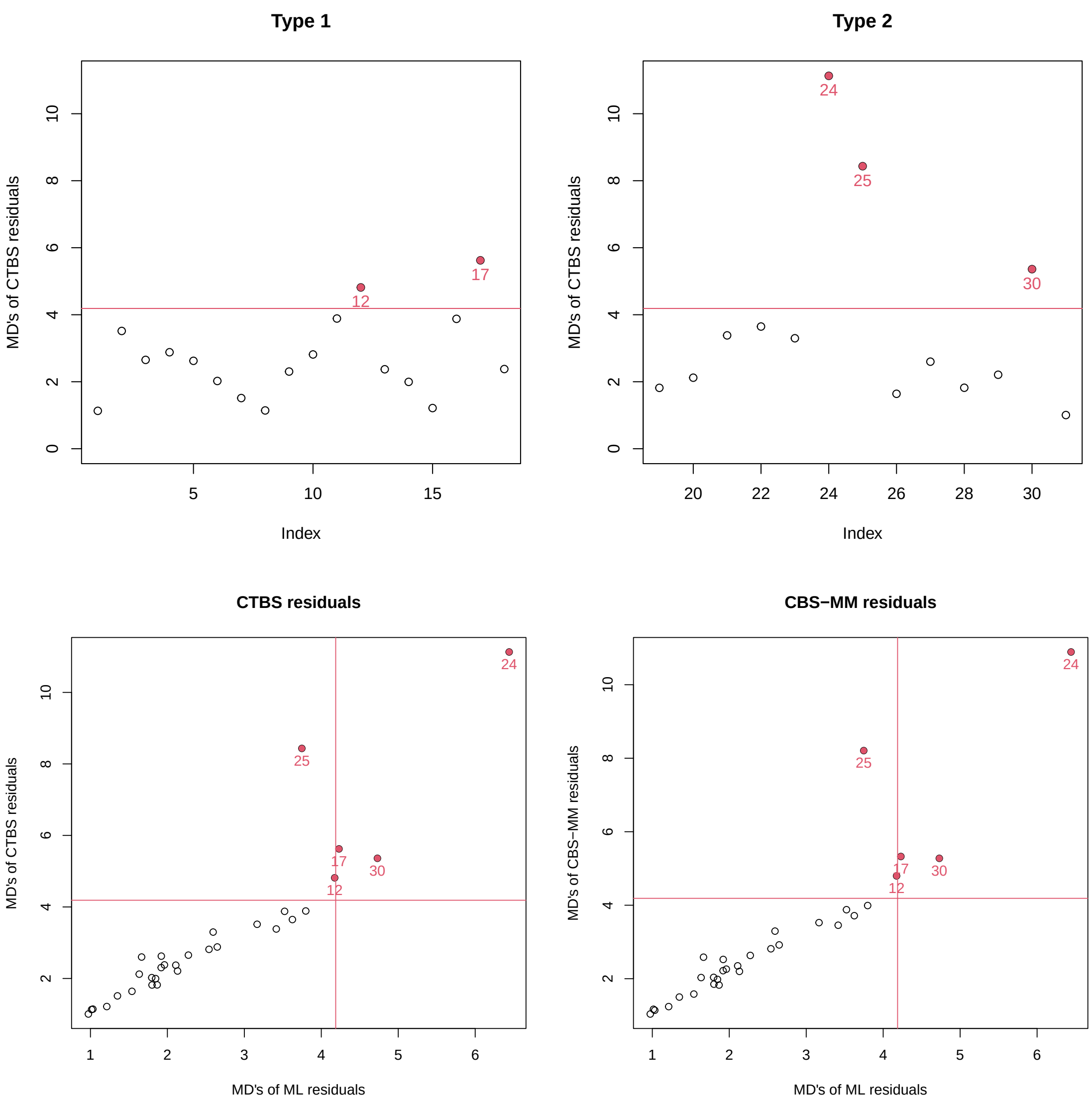

Figure 26: Robust MD's Metallic Oxide data

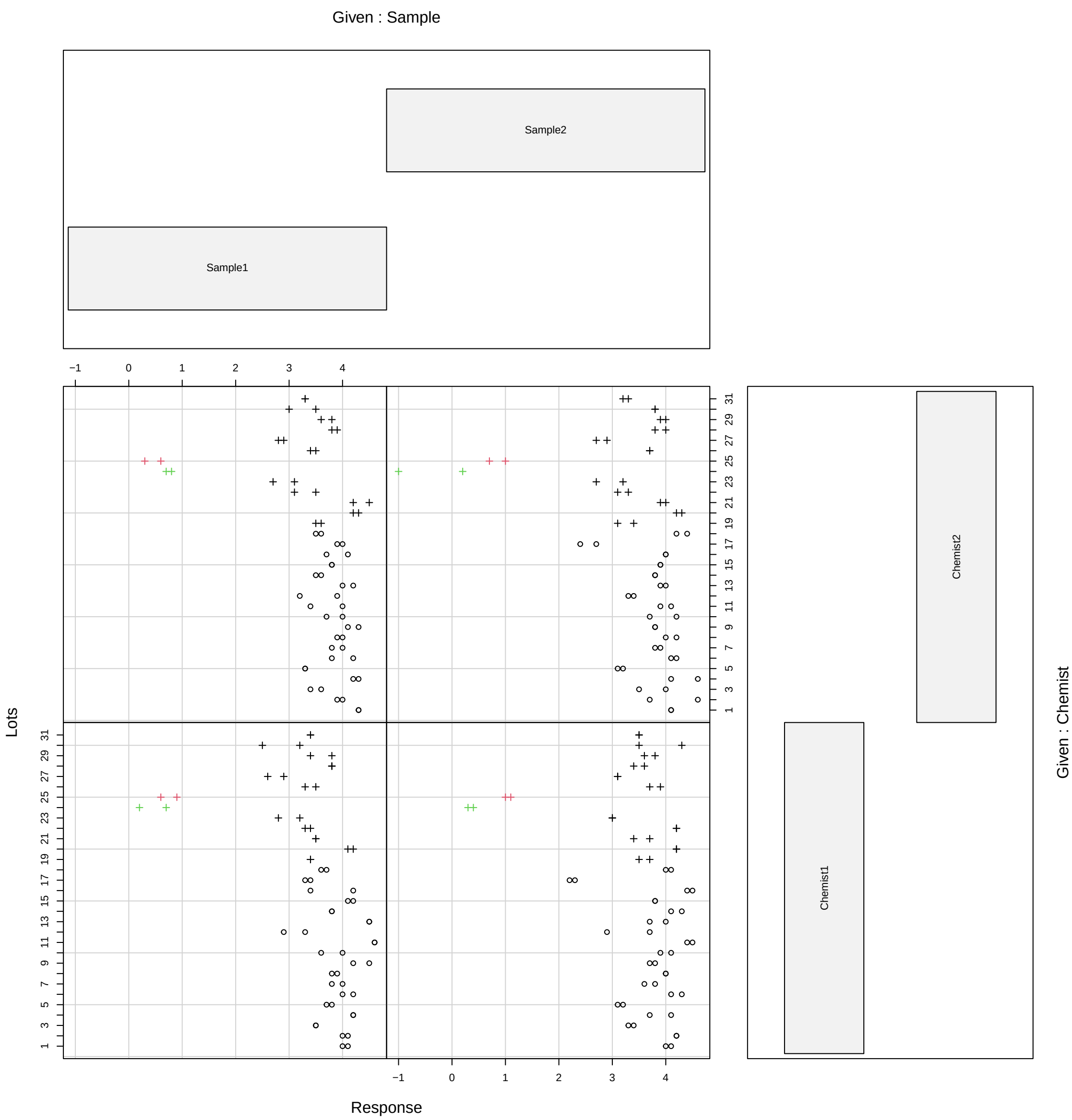

Figure 27: Metallic Oxide data

10 Calcium supplementation for bone gain

The model is an example of a linear growth curve model. If we restrict ourselves to 91 subjects that each have 5 measurements, we have a balanced model. The model is built from a first-stage model

$$\mathbf{y}_i = \mathbf{Z}_i \mathbf{a}_i + \boldsymbol{\epsilon}_i, \quad E(\boldsymbol{\epsilon}_i) = \mathbf{0}, \mathrm{cov}(\boldsymbol{\epsilon}_i) = \sigma^2 \mathbf{I}_k$$

and a second-stage model

$$\mathbf{a}_i \mathbf{A}_i \boldsymbol{\beta} + \mathbf{b}_i, \quad E(\mathbf{b}_i) = \mathbf{0}, \mathrm{cov}(\mathbf{b}_i) = \sigma^2 \mathbf{D}.$$

This leads to

$$\mathbf{y}_i = \mathbf{Z}_i \mathbf{A}_i \boldsymbol{\beta} + \mathbf{Z}_i \mathbf{b}_i + \boldsymbol{\epsilon}_i = \mathbf{X}_i \boldsymbol{\beta} + \mathbf{Z}_i \mathbf{b}_i + \boldsymbol{\epsilon}_i$$

a balanced linear mixed effects model, but with possibly different design matrices $\mathbf{Z}_i$ for the random effects $\mathbf{b}_i$. For the (balanced) calcium data, one has

$$\mathbf{Z}_i = \begin{pmatrix} t_{i1} & 1 \\ t_{i2} & 1 \\ t_{i3} & 1 \\ t_{i4} & 1 \\ t_{i5} & 1 \end{pmatrix} \quad \mathbf{A}_i = \begin{pmatrix} 1 & C_i & 0 \\ 0 & 0 & 1 \end{pmatrix} \quad \mathbf{X}_i = \mathbf{Z}_i \mathbf{A}_i = \begin{pmatrix} t_{i1} & t_{i1}C_i & 1 \\ t_{i2} & t_{i2}C_i & 1 \\ t_{i3} & t_{i3}C_i & 1 \\ t_{i4} & t_{i4}C_i & 1 \\ t_{i5} & t_{i5}C_i & 1 \end{pmatrix}$$

where $C_i = 0$ for the placebo group and $C_i = 1$ for the calcium group. Note that the vector of times $(t_{i1}, \dots, t_{i5})$ (in weeks) at which an individual is measured, can be different for the different individuals. This means we have different design matrices $\mathbf{Z}_i$ for the random effects.

Simplified analysis. We can further simplify the data, by recoding the times at which individuals are measured. The times $t_1, \dots, t_5$ are in weeks, t_1 varies between 0-15 weeks, t_2 between 26-43 weeks, t_3 between 50-66 weeks, t_4 between 76-96 weeks, and t_5 between 102-121 weeks. The time periods are disjoint and are on average about 4 weeks apart.

I propose to recode the $t_1, \dots, t_5$ simply into $1, 2, 3, 4, 5$, representing the time period in which the measurements took place. I also propose to assume that the random effects b_{i1} and b_{i2} are independent with possibly different variances. This leads to a linear mixed effects model, with

$$\mathbf{X}_i = \begin{pmatrix} 1 & C_i & 1 \\ 2 & 2C_i & 1 \\ 3 & 3C_i & 1 \\ 4 & 4C_i & 1 \\ 5 & 5C_i & 1 \end{pmatrix} \quad \mathbf{Z}_1 = \begin{pmatrix} 1 \\ 2 \\ 3 \\ 4 \\ 5 \end{pmatrix} \quad \mathbf{Z}_2 = \begin{pmatrix} 1 \\ 1 \\ 1 \\ 1 \\ 1 \end{pmatrix}$$

where $C_i = 0$ for the placebo group and $C_i = 1$ for the calcium group. This means we have two different 5×3 design matrices $\mathbf{X}_i$ corresponding to three fixed effects $\boldsymbol{\beta} = (\beta_1, \beta_2, \beta_3)$ and two fixed 5×1 design matrices $\mathbf{Z}_1$ and $\mathbf{Z}_2$ corresponding to two random effects $b_{i1} \sim N(0, \sigma_1^2)$ and $b_{i2} \sim N(0, \sigma_2^2)$. Finally, let us assume that $\boldsymbol{\epsilon}_i \sim N(\mathbf{0}, \sigma_\epsilon^2 \mathbf{I}^5)$. In this way, the covariance parameter is $\boldsymbol{\theta} = (\sigma_1^2, \sigma_2^2, \sigma_\epsilon^2)$.

Moreover, with $\mathbf{t} = (1, 2, 3, 4, 5)$,

$$\begin{aligned} \mathbf{L}_1 &= \mathbf{Z}_1 \mathbf{Z}_1^T = \mathbf{t}\mathbf{t}^T \\ \mathbf{L}_2 &= \mathbf{Z}_2 \mathbf{Z}_2^T = \mathbf{J}_5 \\ \mathbf{L}_3 &= \mathbf{I}_5 \end{aligned}$$

which gives a linear covariance structure

$$\begin{aligned} \mathbf{V}(\boldsymbol{\theta}) &= \theta_1 \mathbf{L}_1 + \theta_2 \mathbf{L}_2 + \theta_3 \mathbf{L}_3 \\ &= \sigma_1^2 \mathbf{t}\mathbf{t}^T + \sigma_2^2 \mathbf{J}_5 + \sigma_\epsilon^2 \mathbf{I}_5. \end{aligned}$$

ML estimates. With setting $M = 10000$ and re-scaling during the iteration with the translated biweight, I find the following for the ML estimates

```
         Beta          SE     Pval
1 0.022648582 0.001554870 0.000028
2 0.004408614 0.002231801 0.105200
3 0.851019780 0.009403138 0.000000
```

and

```
sqrt(theta)
0.006580065
0.062226518
0.011713433
```

We see that fixed effects β_1 and β_2 are very significant, whereas β_2 is not significant.

CTBS estimates. With the translated biweight with $M = 2.017548$ and $c = 1.866557$ and with re-scaling during the iteration to satisfy the S-constraint, I find similar estimates

```
         Beta          SE     Pval
1 0.022229388 0.001722339 0.000050
2 0.003412757 0.002473639 0.226203
3 0.848977556 0.010876270 0.000000
```

and

```
sqrt(theta)
0.00678869
0.06599474
0.01104843
```

and therefore also similar conclusion about the random effects.

Indeed, when comparing the Mahalanobis distances, they give very comparable results, see Figure 28. Items 96 (and 102) in de calcium group and item 23 in the placebo group are identified as outlying by both the ML and CTBS estimates.

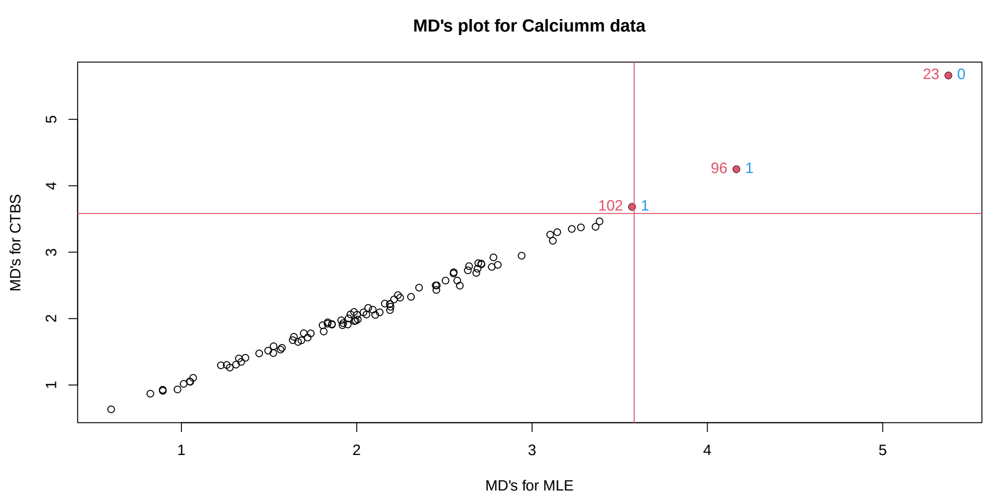

Figure 28: Mahalanobis distances for the calcium data

When looking at the data itself, it seems that items 23, 96 and 102, do not extremely deviate from the bulk of the data.

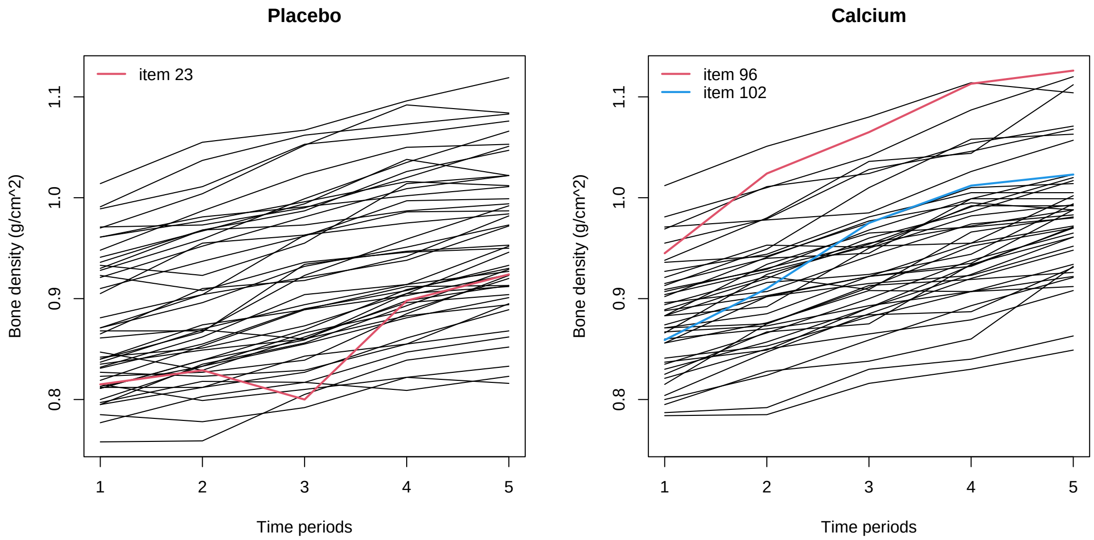

Figure 29: Calcium data

References

[1] M. V. Berry. Improved eigenvalue sums for inferring quantum billiard geometry. *J. Phys. A*, 20(9):2389–2403, 1987.

[2] S. Copt and M.-P. Victoria-Feser. High-breakdown inference for mixed linear models. *Journal of the American Statistical Association*, 101(473):292–300, 2006.

[3] Samuel Copt. *High breakdown inference for mixed linear models*. PhD thesis, Geneva, 2004.

[4] Eugene Demidenko. *Mixed models*. Wiley Series in Probability and Statistics. John Wiley & Sons, Inc., Hoboken, NJ, second edition, 2013. Theory and applications with R.

[5] Stephane Heritier, Eva Cantoni, Samuel Copt, and Maria-Pia Victoria-Feser. *Robust methods in biostatistics*. Wiley Series in Probability and Statistics. John Wiley & Sons, Ltd., Chichester, 2009.

[6] Stephane Heritier, Eva Cantoni, Samuel Copt, and Maria-Pia Victoria-Feser. *Robust methods in biostatistics*. Wiley Series in Probability and Statistics. John Wiley & Sons, Ltd., Chichester, 2009.

[7] Hendrik P. Lopuhaä. On the relation between S-estimators and M-estimators of multivariate location and covariance. *Ann. Statist.*, 17(4):1662–1683, 1989.

[8] Alice M. Richardson and Alan H. Welsh. Robust restricted maximum likelihood in mixed linear models. *Biometrics*, 51(4):1429–1439, 1995.

[9] D. M. Rocke and D. L. Woodruff. Computation of robust estimates of multivariate location and shape. *Statist. Neerlandica*, 47(1):27–42, 1993.

[10] David M. Rocke. Robustness properties of S-estimators of multivariate location and shape in high dimension. *Ann. Statist.*, 24(3):1327–1345, 1996.

[11] P. Rousseeuw and V. Yohai. Robust regression by means of S-estimators. In *Robust and nonlinear time series analysis (Heidelberg, 1983)*, volume 26 of *Lect. Notes Stat.*, pages 256–272. Springer, New York, 1984.

[12] Peter Rousseeuw and Bert van Zomeren. Unmasking multivariate outliers and leverage points. *JASA*.

[13] Shayle R. Searle, George Casella, and Charles E. McCulloch. *Variance components*. Wiley Series in Probability and Mathematical Statistics: Applied Probability and Statistics. John Wiley & Sons, Inc., New York, 1992. A Wiley-Interscience Publication.

[14] David E. Tyler. Some results on the existence, uniqueness, and computation of the M-estimates of multivariate location and scatter. *SIAM J. Sci. Statist. Comput.*, 9(2):354–362, 1988.

[15] Stefan Van Aelst and Gert Willems. Multivariate regression S-estimators for robust estimation and inference. *Statist. Sinica*, 15(4):981–1001, 2005.