**To do with Valérie 24/11/23**

1. Look at the results we have and confirm the comments below.
2. Compare more precisely CCM and ICM for error contamination.
3. Run simulations with x contamination with larger values for alphac, and with error contamination with larger values of mec (not possible to take mec=-1280 and 5%, the code breaks down).
4. Try nrep=250 to check the stability of the results.
5. Try simulations with k=8,
6. Generate randomly the second column of the X matrix with different values for the different cases and keep it fixed in the repetitions. Concerning the X contamination, we can either keep the contamination as it is known (multiplication by alphac) or use a shift (see for Agostinelli and Yohai bottom of page 1770).
7. Try to compare the code with Claudio with Monte Carlo simulations (see remark below).

**Comments on simulation results with Rik and Valérie 22/11/23**

1. **CCM**

1- CCM, rc=0 with contamination on the error with proportion of outliers varying from 0% to 30%

1. For mec=-40, behavior of the robust beta estimators not very different from MLE and deviate for larger contamination. The best estimate is ctau with comparable estimates as MLE. The MM estimates are better than S estimates. For theta, S and MLE are similar and cTAU is a bit worse.
2. For mec=-80, all robust estimators are better than MLE with S the best. All estimates deviate for larger contamination. The MM estimate are between the ctau and MLE estimates. For theta1, 2 and 3, S is better than MLE which is better than cTAU. For theta4, S is comparable to MLE and cTAU is much better.
3. For mec=-160, the S and cTAU are comparable for beta while MLE deviates a lot when the contamination increases. MM deteriorates for 30% of contamination while it was comparable with S and cTAU for less contamination. For all thetas, MLE is bad as soon as there is contamination. S and cTAU are quite comparable except for theta 1 and theta 4 especially for large contamination (20% for theta 1 and 5% for theta 4) where cTAU is better than S.

**Conclusion: the larger is the size of the contamination the better the cTAU estimate behaves compared to the other estimates while the S estimate is better than cTAU when the size is not very large.**

2- CCM, rc=0 with contamination on the second component random effect with proportion of outliers varying from 0% to 30%. Note that we do not expect changes on the beta1 when the contamination increases.

a- when mbc2=-12.5 and b=-25, all beta estimators behave the same and the results are bad for beta2 with contamination. For theta the estimates also behave quite similarly with increasing bias for theta1 and theta3 with the contamination. Results are not bad for theta 2 and theta 4 for all estimates.

b- when mbc2=-50, the robust estimators of beta2 are stable up to 10% and S is a bit better while the MLE is biased for 5 and 10% contamination already. For 30% all estimates are biased a lot. S estimates are quite better. For theta1, all estimates do not differ a lot and become quite biased when the contamination is larger or equal than 20%. All estimators are quite stable for theta2 (except cTAU for 10%). The robust estimates are stable for theta3 up to 10% while there are large deviations for larger contamination. MLE is already bad for 5% contamination. For theta4, the MLE is stable and good but the S and cTAU have strange behavior (large bias for contamination around 5-10%).

**Conclusion: all estimates comparable when the size of the contamination is not too large with bad results for all estimates for beta2, moderatley bad for theta1 and theta3 and quite good for theta4. When the size of the contamination becomes large, the MLE estimates becomes quite bad compared to the robust beta estimates while the MLE is better for theta4.**

3- CCM, rc=0 with contamination on the X matrix with proportion of outliers varying from 0% to 30%.

1. When alphac=2, for beta1, MLE, S and MM are comparable and deviate a bit when the contamination increases (but the bias does not increase with the contamination) while cTAU is good and stable. For beta2, the conclusions are the same as for beta1 but with much more deviation for MLE, MM and S when the contamination increases. For thetas, the estimates are quite good and comparable.
2. When alphac=5 and 10, the conclusions differ from the ones with alphac=2 in the sense that the behavior of the estimates is more extreme. MLE, MM and S behave the same way while cTAU behaves very well. For theta, results are quite stable but deteriorate for theta3 (and theta2) with MLE and S which are comparable.

**Conclusion: for X contamination, S and MM are not very good while cTAU is very good.**

Results are similar with sometimes larger boxplots for CCM and rc=1.

1. **ICM**

1- ICM with contamination on the error with proportion of outliers varying from 0% to 30%

1. For mec=-40, behavior of the robust beta estimators not very different from MLE and deviate for larger contamination only for beta1 (deviates for beta2 but less). For theta, S, ctau and MLE are similar and are ok for theta1, 2 , 3 but deviate for theta4.
2. For mec=-80, same as -40 but with more extreme deviations.
3. For mec=-160, ~~the S and cTAU are comparable for beta while MLE deviates a lot when the contamination increases~~. ~~MM and S deteriorate for 20% of contamination and all of them for 30% while results were comparable with S, MM and cTAU for less contamination.~~ Again the results are not bad for beta2 and deteriorate for beta1. S estimates for beta 1 look better for large contamination. For thetas, MLE is good except for theta4. For theta2 and theta3, the robust estimates are quite good. For theta 4 the MLE and S estimates deteriorate with the contamination and also cTAU for 30%.

2- ICM with contamination on the second component random effect with proportion of outliers varying from 0% to 30%. Note that we do not expect changes on the beta1 when the contamination increases.

When mbc2=-12.5, ~~-24~~ and -50, all beta estimators behave the same and the results are bad for beta2 with contamination. For thetas the estimates also behave quite similarly with increasing bias for theta1 and theta3 with the contamination. Results are not bad for theta 2 and theta 4 for all estimates. Same as CMM. For mbc2= -50, S estimates are a bit better for beta2 for large contamination.

3- ICM, rc=0 with contamination on the X matrix with proportion of outliers varying from 0% to 30%.

1. When alphac=2, for beta1, MLE, S and MM are comparable and deviate a bit when the contamination increases (but the bias does not increase with the contamination) while cTAU is good and stable. For beta2, the conclusions are the same as for beta1 but with much more deviation for MLE, MM and S when the contamination increases. For thetas, the estimates are quite good and comparable.
2. When alphac=5 and 10, the conclusions differ from the ones with alphac=2 in the sense that the behavior of the estimates is more extreme. MLE, MM and S behave the same way while cTAU behaves very well. For theta, results are quite stable but deteriorate for theta3 **and theta2** with MLE and S which are comparable.

**General conclusions: Results are quite similar between the 3 contamination frameworks (ICMn CCM with rc=1 and CCM with rc=0). It seems that the results for X contamination do not depend on the percentage of contamination.**

Remarks

* What is new compared to the other papers (Agostinelli and Yohai, AY, on cTAU and Mason et al., CG, with the same model for the data):
  + In AY the contamination is in y and in x together and not separately and not in the random effects. Only 10% contamination for ICM and CCM (they claim the results are the same for 5 and 15%). They consider dimension k=12.
  + In MCG, there are only confidence intervals and coverage probabilities and no contamination on X.
* For the comparison between Claudio S implementation and our implementation, Rik sent me some programs on the 23rd of November. On the small example we have from the function

varComprob(y ~ time, groups = groups, data = Dataset, varcov = K,

control = varComprob.control(lower = c(0,0,-Inf), method = "S", psi = "bisquare"))

> model.S$vcov.beta

[,1] [,2]

[1,] 6.684722 -1.077610

[2,] -1.077610 1.170561

While using our function Robmle we get

> summaryS$varbetaShat

          [,1]      [,2]

[1,]  8.479521 -1.348764

[2,] -1.348764  1.483060

We compared the S-estimator beta estimates on our simulation scenarios and in fact not only the variances of the beta differ but it seems that the betas also differ and in fact we can compute the Monte Carlo variances from our simulations and we find 6.8 for Claudio beta estimates and 8.5 for our beta estimates… To be confirmed.

Comparison more precisely CCM and ICM for error contamination

|  |  |
| --- | --- |
|  | CCM / ICM |
| contamination on the error | 1. For mec=-40, in CCM, estimates deviate for beta 2 for large contamination and do not deviate for ICM. In CCM estimates deviate for theta 1 2 3 but less for ICM. For CCM, estimates for theta 4 deviate positively instead they deviate negatively for ICM 2. For mec=-80, in CCM, estimates deviate for beta 2 for large contamination and do not deviate for ICM. In CCM estimates deviate for theta 1 2 3 4 for large contamination and in ICM estimates deviate only for theta 4 3. For mec=-160, in CCM estimates of beta 1 deviate only for MLE and for ctau and MM for very large contamination. In ICM estimates of beta1 deviates also for MLE and for other estimates from moderate contamination. In ICM ctau estimate looks still good for moderate contamination. In CCM estimates of beta2 deviate for MLE and other estimates for very large contamination. In ICM, estimates do not deviate.   In CCM, estimates deviate for theta1 2 3 for MLE and all estimates deviate for theta4 except ctau. In ICM all estimates deviate only for theta4. |
| contamination on the second component random effect | 1. when mbc2=-12.5 and mbc2=-50, similar behavior of estimates in CCM and ICM |
| contamination on the X matrix | 1. When alphac=2, similar behavior of estimates in CCM and ICM. 2. When alphac=5, 10 similar behavior of estimates in CCM and ICM. Except for theta1 and 2, estimates deviate more for S and MLE in ICM |

Generate randomly the second column of the X matrix with different values for the different cases and keep it fixed in the repetitions. Concerning the X contamination, we can either keep the contamination as it is known (multiplication by alphac) or use a shift (see for Agostinelli and Yohai bottom of page 1770).

New entry of the different functions.

Xa=TRUE/FALSE.

Xa == TRUE X2i follow a normal distribution with mean (0,…0) and variance-covariance Ip

Xshiftall=TRUE/FALSE

Xshiftall==TRUE if X2i contaminated so X2i follow a normal distribution with mean (mux,….,mux) and variance-covariance Ip

Xshiftall==FALSE if X2i contaminated so X2i[1] follow a normal distribution with mean mux and variance-covariance 1

Mux : the value of the shift =1,5,10,50

Try to compare the code with Claudio with Monte Carlo simulations (see remark below).

Code ready

Senarios ?