README file

This README file describes which scripts can be used to run simulations. It also describes what R-functions are used in the scripts.

There are four R-scripts that one can use to run a simulation

1. Simulation\_setting\_model\_MCG.R
2. Simulation\_setting\_model\_MCG\_cTAU.R
3. Simulation\_setting\_model\_MCG\_contCCM.R
4. Simulation\_setting\_model\_MCG\_contCCM\_cTAU.R

Each of these simulation scripts can be run starting from an empty environment, since it sources other R-scripts that are needed. Before running the simulation script one has first to create particular subfolders as described in the simulation script. In this way the results of the simulation are stored in .Rdata files in the subfolders. In each simulation script the user can specify the contamination scenarios for which to run a simulation.

The first R-script runs a simulation according to the model in Mason, Cantoni and Ghisletta (2021) for the MLE, S- and MM-estimators.

The second R-script runs a simulation according to the model in Mason, Cantoni and Ghisletta (2021) for the MLE, S, MM, and the cTAU estimators from Agostinelli and Yohai (2016).

The third R-script runs a simulation according to the model in Mason, Cantoni and Ghisletta (2021) for the MLE, S- and MM-estimators, but uses the central contamination model (CCM) to generate the contamination.

The fourth R-script runs a simulation according to the model in Mason, Cantoni and Ghisletta (2021) for the MLE, S, MM, and the cTAU estimators from Agostinelli and Yohai (2016), but uses the central contamination model (CCM) to generate the contamination.

The reason to have separate scripts including the cTAU estimator is that the computation of these estimators with varComprob are very slow.

# Preparing scripts

1. asympt\_norm\_constants.R

This script computes the constants involved in the limiting covariances of the S-estimators. See, e.g., Theorem 3 in Rousseeuw & Yohai (1984), Corollary 5.1 in Lopuhaä (1989), or Corollary 9.2 in Lopuhaä, Gares & Ruiz-Gazen (2023).

Expectations are exact and are computed by means of the Gamma-function.

1. biweight\_functions.R

This script defines the biweight and the translated biweight rho functions and their derivatives. These function are needed in solving the fixed point equations satisfied by the MLE, S- and MM-estimators

# Data generating scripts

1. function\_data\_gen\_MCG.R

This script contains the code for the function data\_gen\_MCG. This function generates a single dataset according to the model in Mason, Cantoni & Ghisletta (2021) with contamination generated according to the ICM (independent contamination model or cellwise contamination) in the measurement error and according to CCM (central contamination model) in the random effects. In addition, the function also generates contamination in the design matrix of the fixed effects according to ICM.

The function returns the

* 1. y-observations in a matrix with n rows (individuals) and k columns (observations)
  2. a list X of the fixed effect design matrices for the individuals
  3. a list Z of the random effect design matrices for the individuals
  4. contamination probabilities
     1. pe (measurement error epsilon)
     2. pb (random effect b)
     3. px (fixed effect design matrix X)
  5. amount of contamination
     1. mec (shift in mean of measurement coordinate epsilon\_ij)
     2. mbc2 (shift in mean of random effect b2)
     3. alphac (multiplication of x2)
  6. number of outliers
     1. noei (number of contaminated individuals when pe>0)
     2. noe (number of contaminated observations when pe>0)
     3. nobi (number of contaminated individuals when pb>0)
     4. noxi (number of contaminated individuals when px>0)
     5. nox (number of contaminated observations when px>0)

1. function\_data\_gen\_MCG\_contCCM.R

This script contains the code for the function data\_gen\_MCG\_contCCM. This function generates datasets according to the model in Mason, Cantoni & Ghisletta (2021) with contamination generated according to the CCM (central contamination model) in the random effects, in the measurement error, and in the design matrix of the fixed effects.

The function returns the same objects as data\_gen\_MCG.

# Scripts to compute estimates

1. Robust\_lme.R

This script contains the code for the function Roblme. This function computes the MLE, the S- and MM-estimates for a single dataset, such as the ones in the output of the function data\_gen\_MCG\_contCCM.

The function returns

* 1. fixedeffectsS: a summary for the estimator for the fixed effects containing
     1. estimates
     2. standard errors
     3. SE ‘s are extracted from the estimated asymptotic variance, computed from the formulas in Corollary 9.2 in Lopuhaä, Gares and Ruiz-Gazen (2023), using plug-in.
     4. t-values
     5. p-values
  2. fixedeffectsMM: a summary for the MM-estimator (if it is computed) containing the same table.
  3. summarythetaS: a summary for the estimator for the covariance parameters containing the same table.
  4. varbetaShat: the estimated asymptotic variance for the estimator of the fixed effects, computed from the formulas in Corollary 9.2 in Lopuhaä, Gares and Ruiz-Gazen (2023), using plug-in.
  5. varbetaMMhat: the estimated asymptotic variance for the MM-estimator of the fixed effects (if it is computed)
  6. varthetahat: the estimated asymptotic variance for the estimator of the covariance parameters, computed from the formulas in Corollary 9.2 in Lopuhaä, Gares and Ruiz-Gazen (2023), using plug-in.
  7. w: weights from covMCD that is used as initial covariance estimator
  8. dis: the Mahalanobis distances
  9. numbers of iterations
     1. iterS for the S-estimator
     2. iterM for the MM-estimator

1. function\_MLESMM\_estimates\_MCG.R

This script contains the code for the function MLESMM\_estimates\_MCG. This function generates nrep datasets by means of data\_gen\_MCG. For each generated dataset the function computes the MLE, S and MM estimates with the function Robmle and saves the results in a list MLESMM containing

* 1. a separate list $MLE, $S and $MM for each estimator, containing
     1. $beta the estimators for the fixed effects
     2. $varbeta the estimated asymptotic variance for the estimators of the fixed effects
     3. $theta the estimators for the covariance parameters
     4. $vartheta the estimated asymptotic variance for the estimators of the covariance parameters.
  2. a dataframe $no\_outliers containing the number of outliers present in each iteration

1. function\_MLESMM\_estimates\_MCG\_contCCM.R

This script contains the code for the function MLESMM\_estimates\_MCG\_contCCM. This function generates nrep datasets by means of data\_gen\_MCG\_CCM. For each generated dataset the function computes the MLE, S and MM estimates with the function Robmle and saves the results in a list MLESMM\_CCM containing the same objects as MLESMM.

1. function\_MLESMMcTAU\_estimates\_MCG.R

This script contains the code for the function MLESMM\_estimates\_MCG. This function generates nrep datasets by means of data\_gen\_MCG. For each generated dataset the function computes the MLE, S and MM estimates with the function Robmle and the cTAU estimators with the function varComprob and saves the results in a list MLESMMcTAU containing the same objects as MLESMM.

1. function\_MLESMMcTAU\_estimates\_MCG\_contCCM.R

This script contains the code for the function MLESMM\_estimates\_MCG\_contCCM. This function generates nrep datasets by means of data\_gen\_MCG\_CCM. For each generated dataset the function computes the MLE, S and MM estimates with the function Robmle and the cTAU estimators with the function varComprob and saves the results in a list MLESMMcTAU\_CCM containing the same objects as MLESMM.

# Scripts for preparing graphs

1. boxplot\_epsilon\_contamination
2. ConfInterval\_CP\_epsilon\_contamination
3. ConfRegion\_CP\_epsilon\_contamination