RoBoost-PLSR :
A new Robust PLS regression method

Maxime Metz; Florent Abdelghafour; Jean-Michel Roger; Matthieu Lesnoff

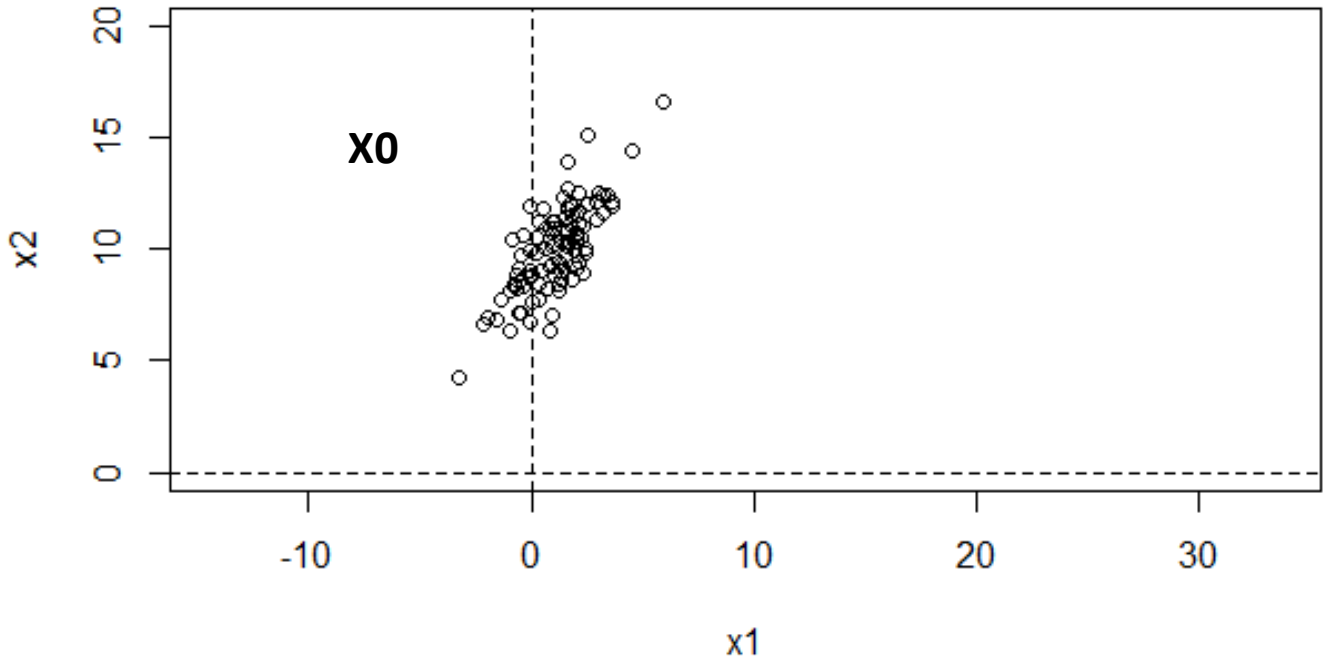
Joint Research Unit
ITAP Technologies & methods for the agriculture of tomorrow
INRAE - Montpellier SupAgro

Summary

1. Introduction
   a. Robustness
   b. Robust PLS methods
2. Theory
   a. RoBoost-PLS regression
   b. Pros and Cons
3. Materials and Methods
   a. Methods
   b. Data
4. Results and discussions
5. Conclusion

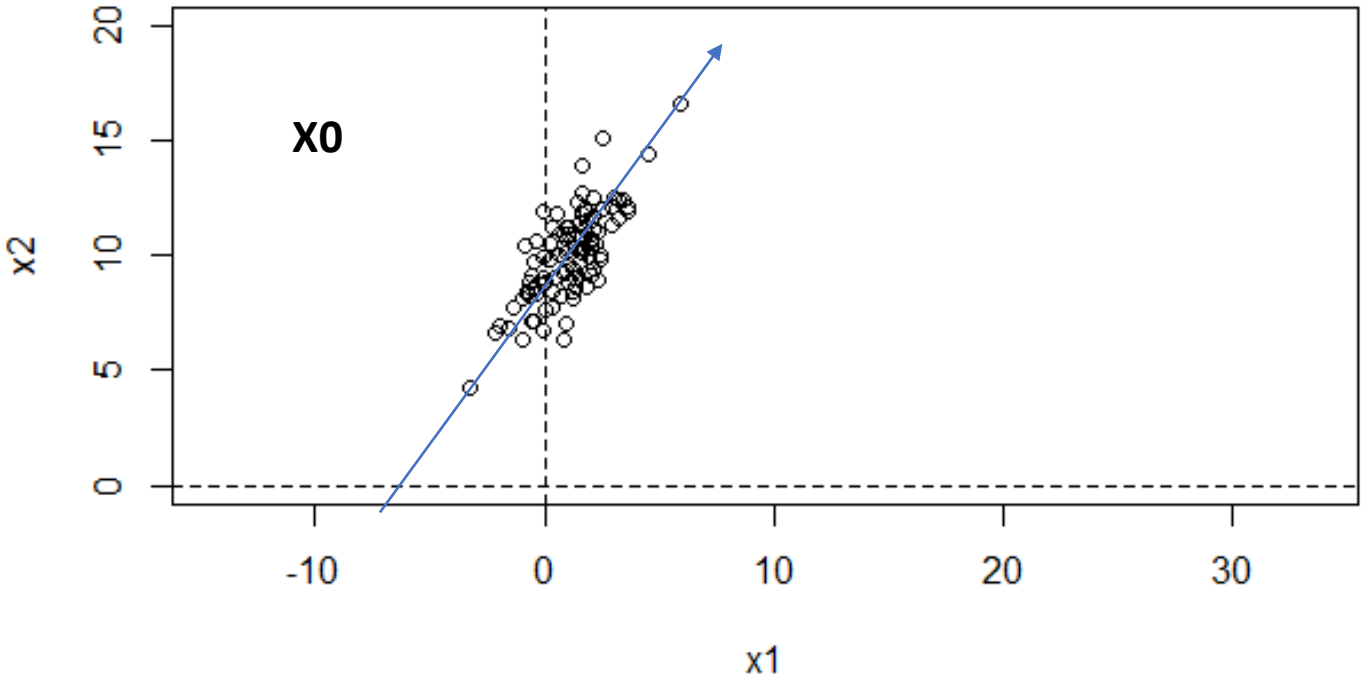
1. Introduction

a. Robustness



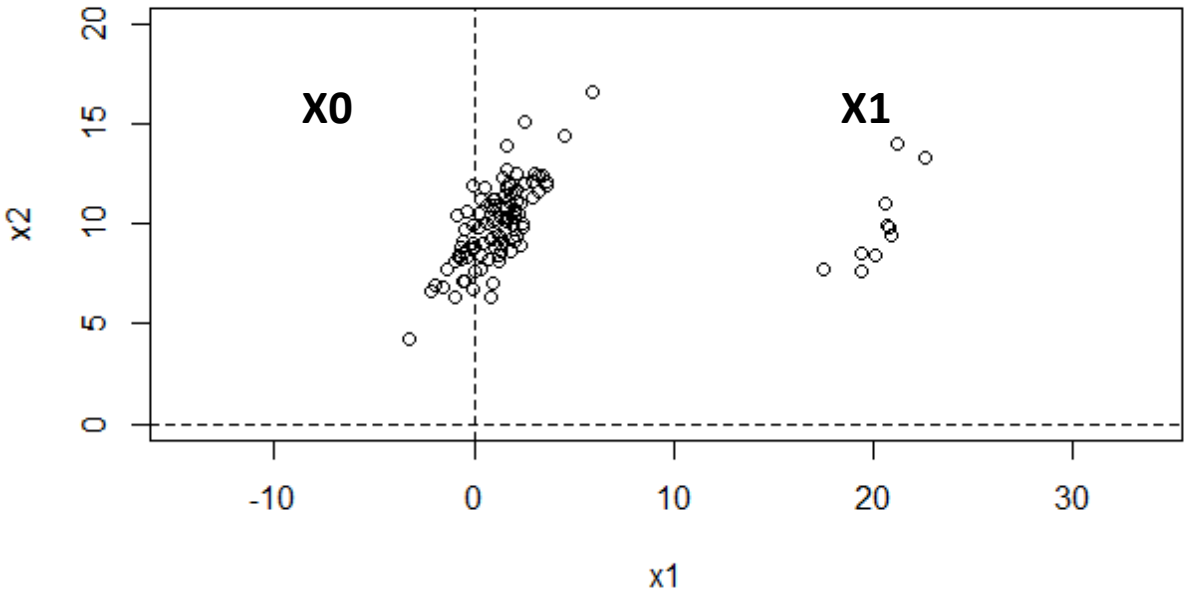
1. Introduction

a. Robustness



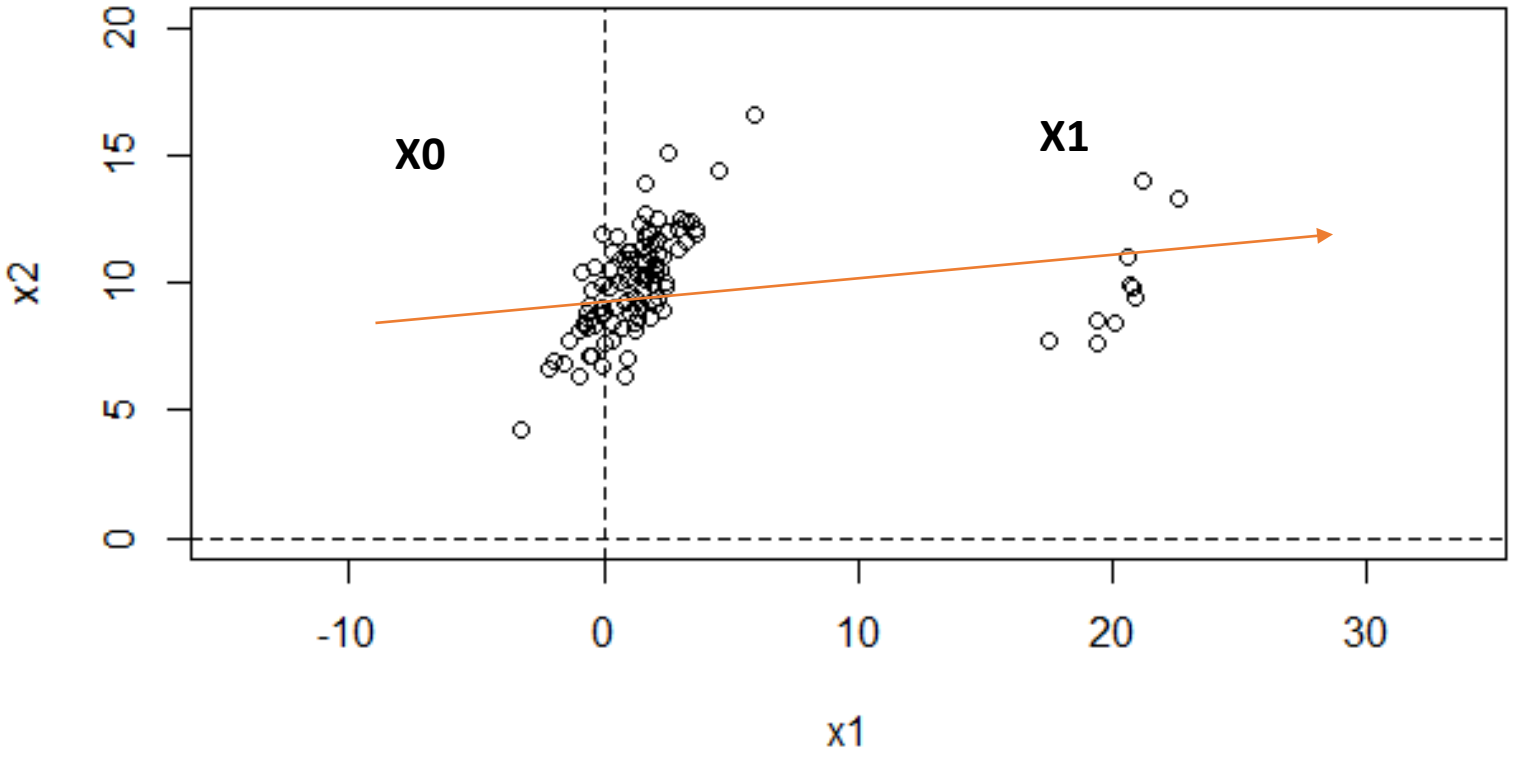
1. Introduction

a. Robustness



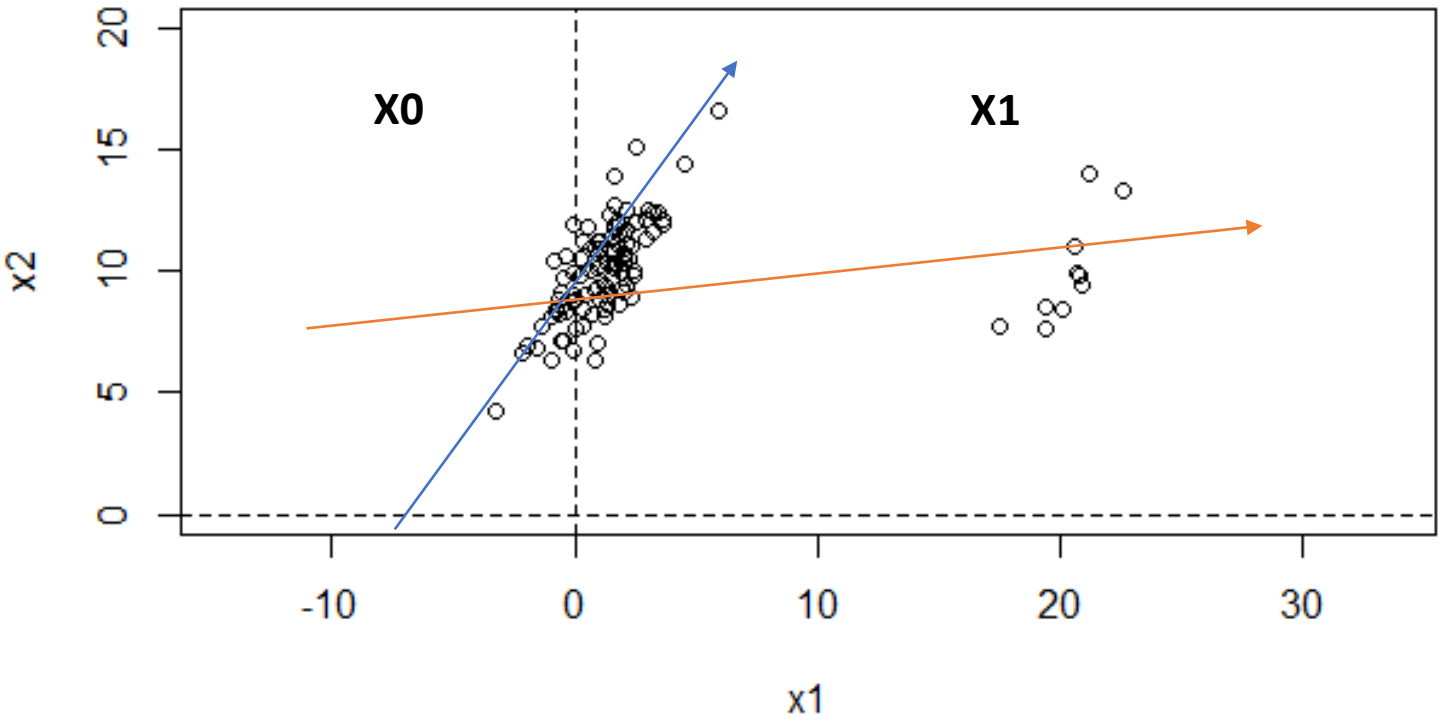
1. Introduction

a. Robustness



1. Introduction

a. Robustness



1. Introduction

a. Robustness

Main hypothesis in Robust methods :

- Robust methods assume that the largest mass of data is X0 .
- The learning database is polluted

Main difficulties in Robust methods :

- Find a good measurement to highlight the outliers (especially when estimating leverage points)

1. Introduction

b. Robust PLS methods in literature

Wakelinc et Macfie, « A Robust PLS Procedure »

Cummins , « Iteratively reweighted partial least squares: A performance analysis by monte carlo simulation »

Pell, « Multiple Outlier Detection for Multivariate Calibration Using Robust Statistical Techniques »

Griep et al., « Comparison of Semirobust and Robust Partial Least Squares Procedures »

Serneels et al., « Partial Robust M-Regression »;

Gil et Romera, « On Robust Partial Least Squares (PLS) Methods »

Møller, Frese, et Bro, « Robust Methods for Multivariate Data Analysis »

Hubert et Branden, « Robust Methods for Partial Least Squares Regression ».

1. Introduction

b. Robust PLS methods in literature

PRM (Partial Robust M-regression) :

Calculation of a PLS model with a defined **x** latent variables then weighting according to the Y residuals and leverage.

1. Introduction

b. Robust PLS methods in literature

PRM (Partial Robust M-regression) :

Calculation of a PLS model with a defined **x** latent variables then weighting according to the Y residuals and leverage.

Weighting X and Y matrices

1. Introduction

b. Robust PLS methods in literature

PRM (Partial Robust M-regression) :

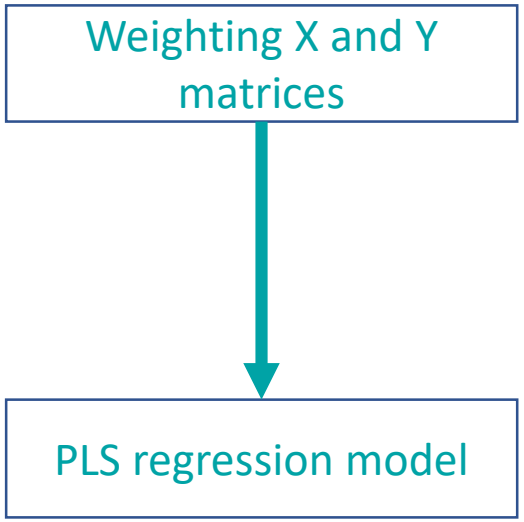
Calculation of a PLS model with a defined **x** latent variables then weighting according to the Y residuals and leverage.

1. Introduction

b. Robust PLS methods in literature

PRM (Partial Robust M-regression) :

Calculation of a PLS model with a defined **x** latent variables then weighting according to the Y residuals and leverage.

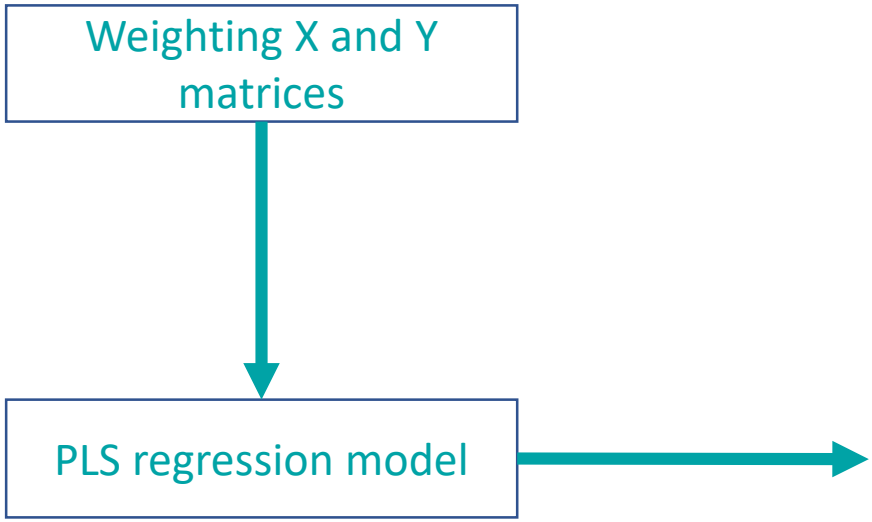


1. Introduction

b. Robust PLS methods in literature

PRM (Partial Robust M-regression) :

Calculation of a PLS model with a defined **x** latent variables then weighting according to the Y residuals and leverage.

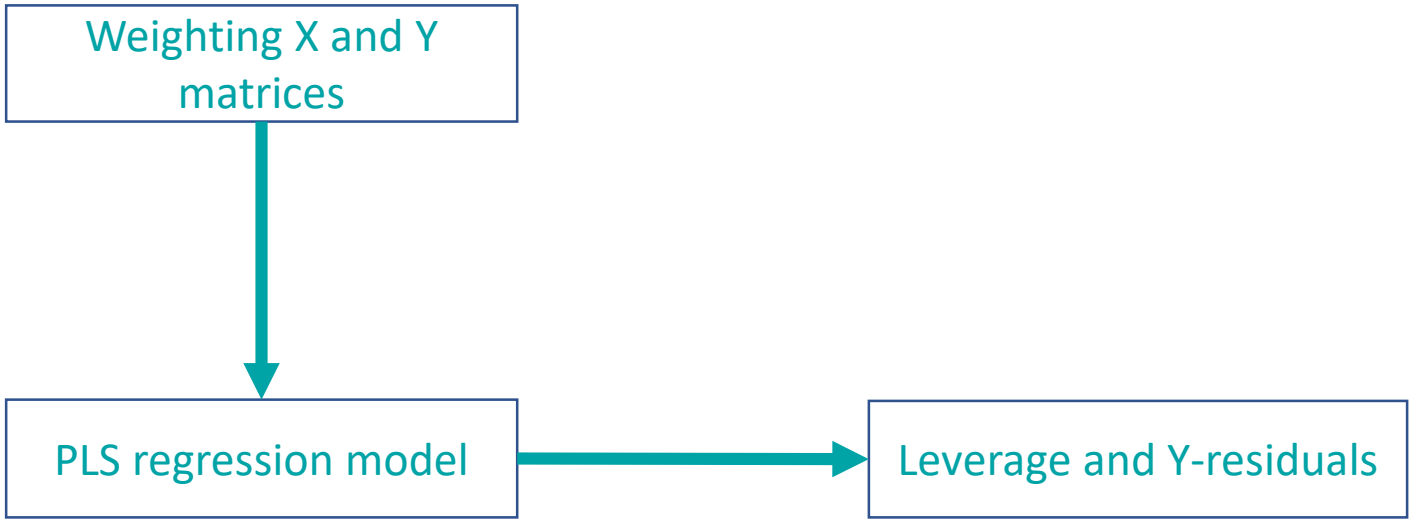


1. Introduction

b. Robust PLS methods in literature

PRM (Partial Robust M-regression) :

Calculation of a PLS model with a defined **x** latent variables then weighting according to the Y residuals and leverage.

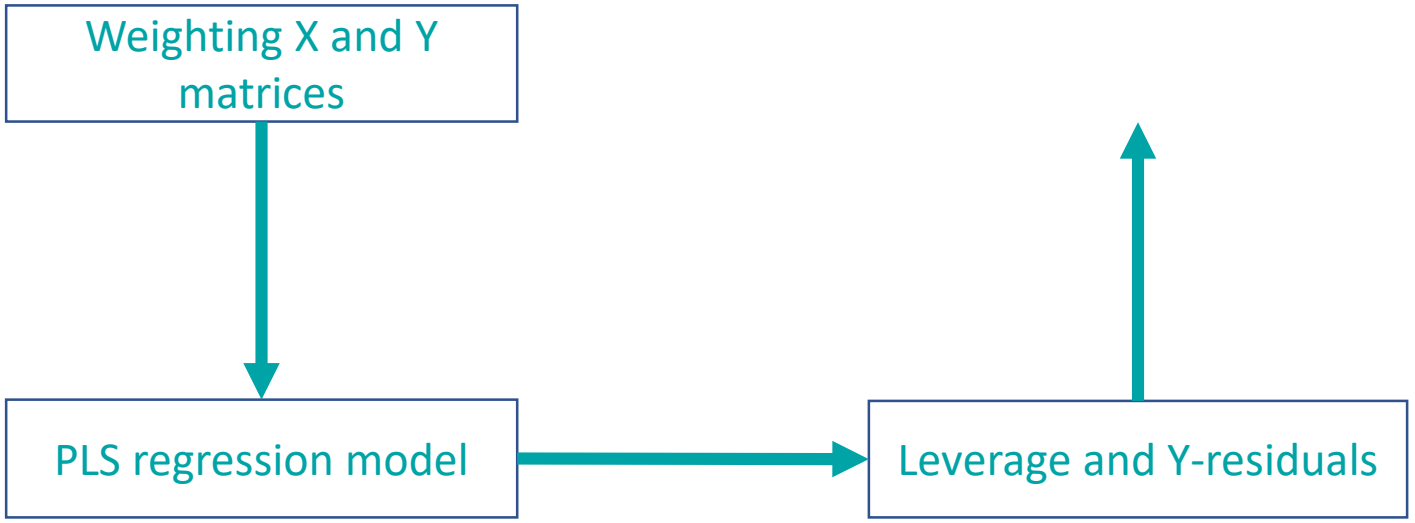


1. Introduction

b. Robust PLS methods in literature

PRM (Partial Robust M-regression) :

Calculation of a PLS model with a defined **x** latent variables then weighting according to the Y residuals and leverage.

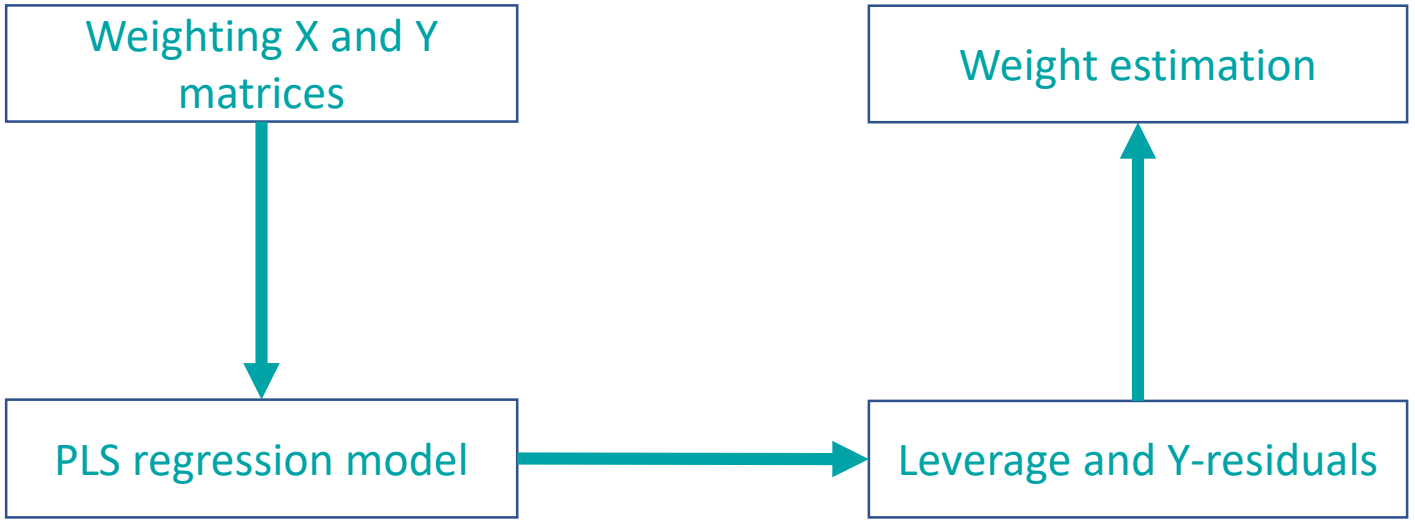


1. Introduction

b. Robust PLS methods in literature

PRM (Partial Robust M-regression) :

Calculation of a PLS model with a defined **x** latent variables then weighting according to the Y residuals and leverage.

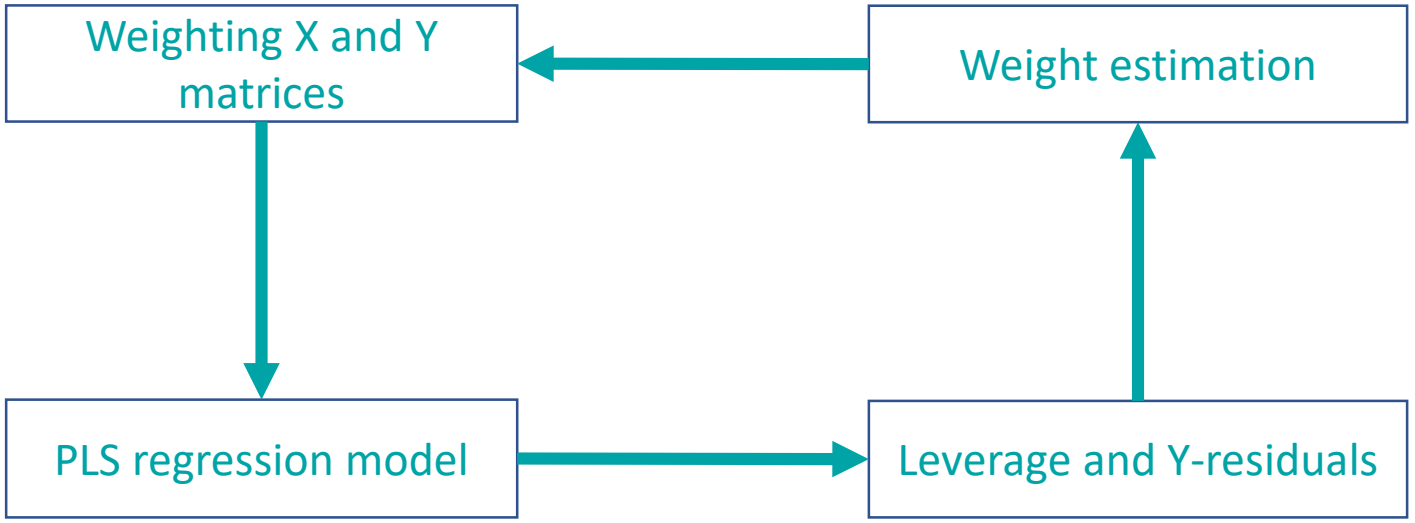


1. Introduction

b. Robust PLS methods in literature

PRM (Partial Robust M-regression) :

Calculation of a PLS model with a defined **x** latent variables then weighting according to the Y residuals and leverage.

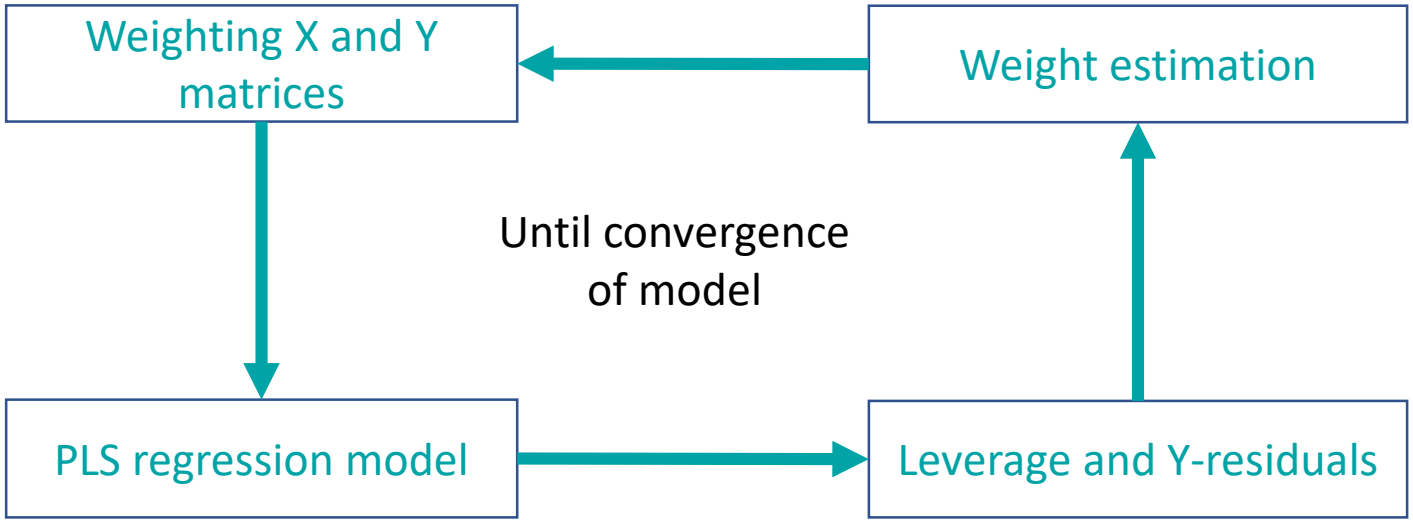


1. Introduction

b. Robust PLS methods in literature

PRM (Partial Robust M-regression) :

Calculation of a PLS model with a defined **x** latent variables then weighting according to the Y residuals and leverage.



2. Theory

a. RoBoost-PLS

2. Theory

a. RoBoost-PLS

Weighted PLSR model with 1 LV

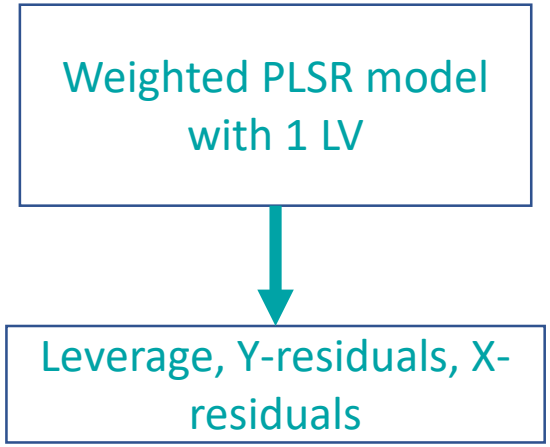
2. Theory

a. RoBoost-PLS

Weighted PLSR model with 1 LV

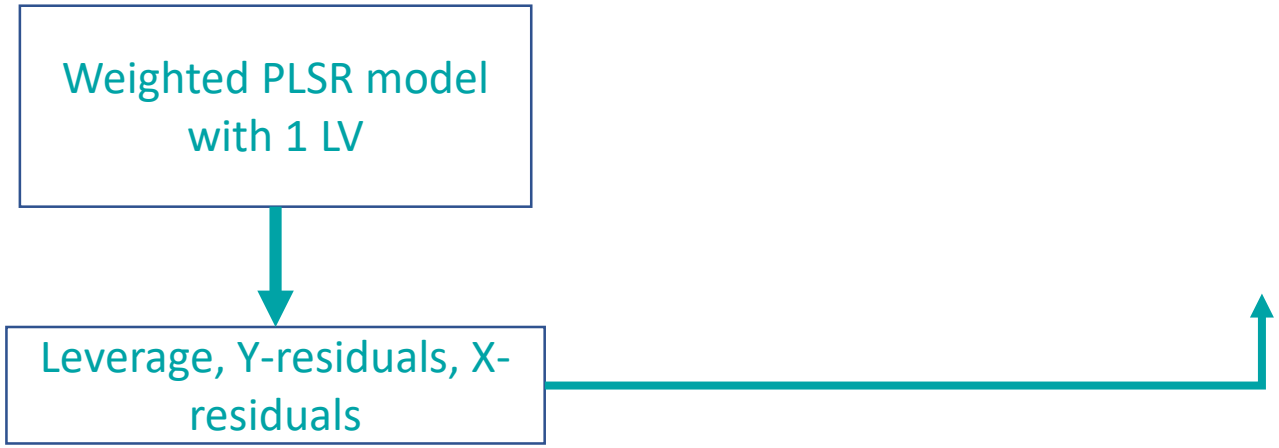
2. Theory

a. RoBoost-PLS



2. Theory

a. RoBoost-PLS

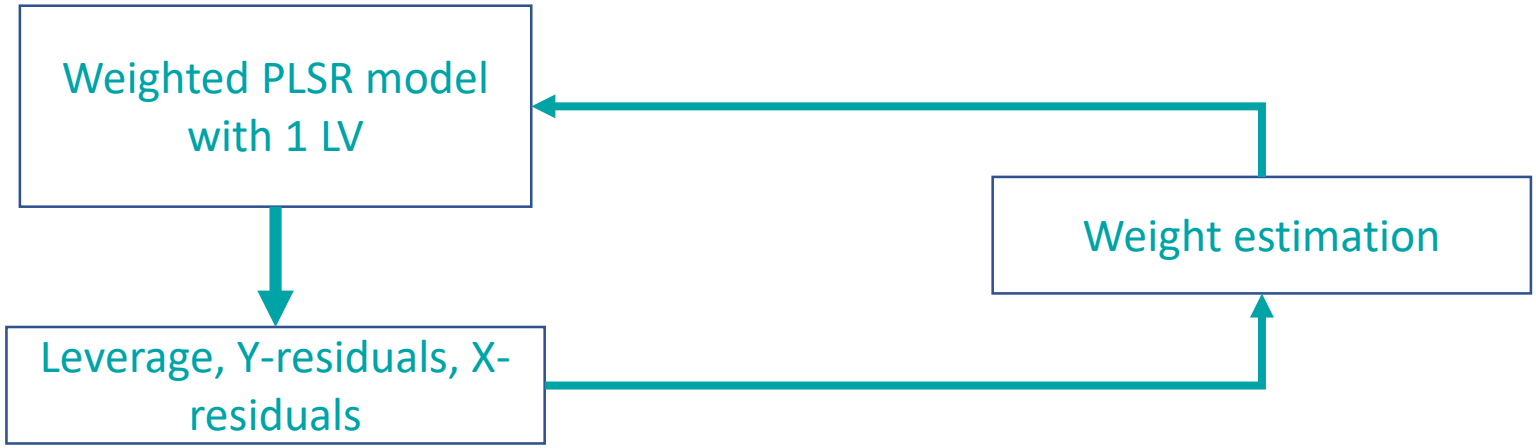


2. Theory

a. RoBoost-PLS

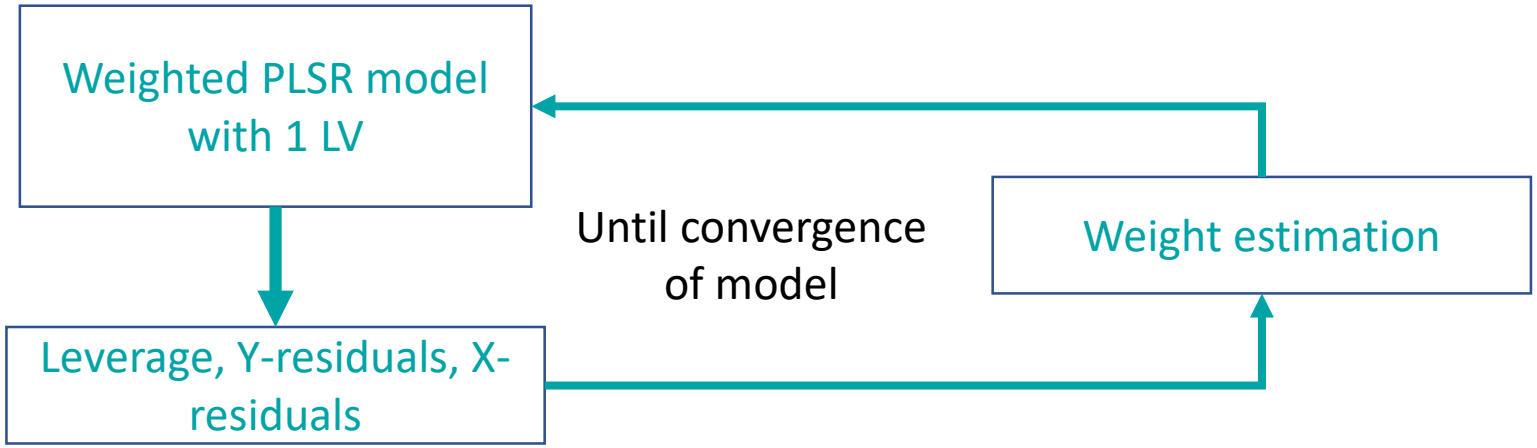
2. Theory

a. RoBoost-PLS



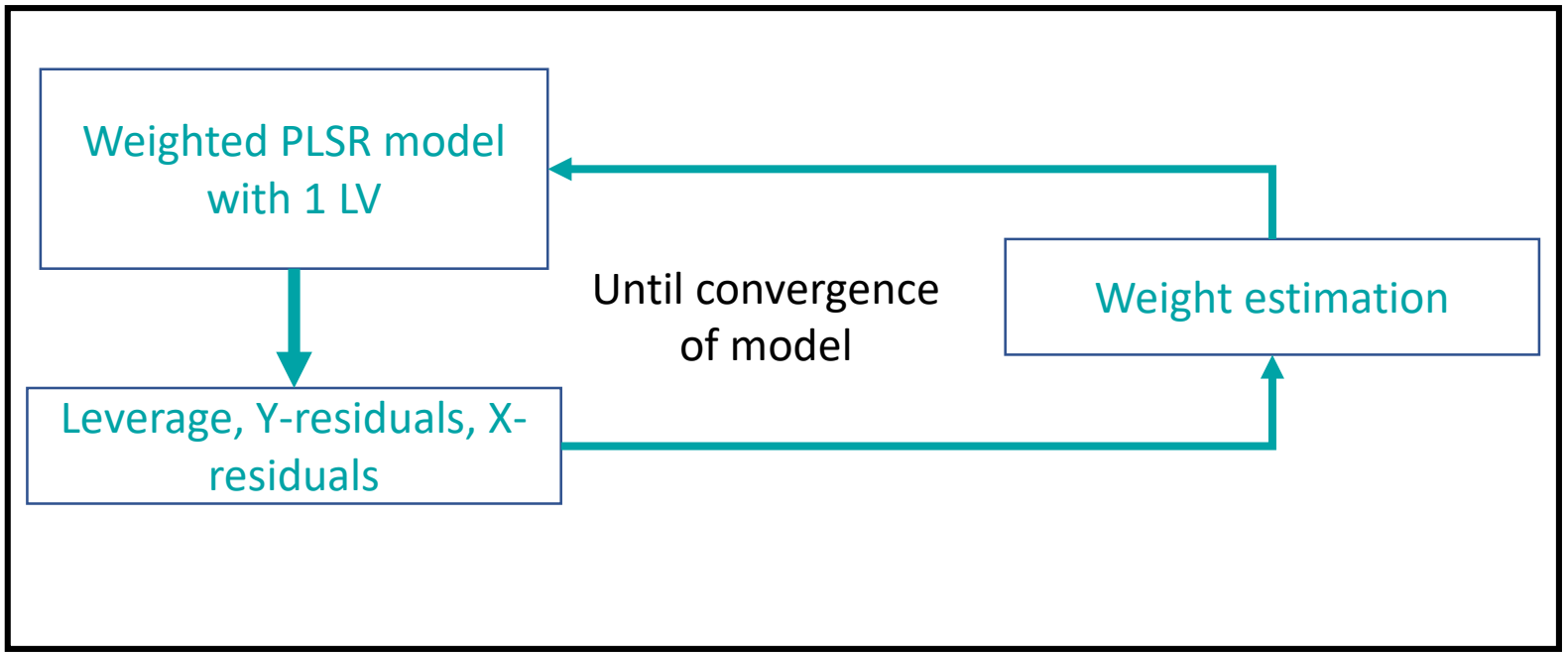
2. Theory

a. RoBoost-PLS



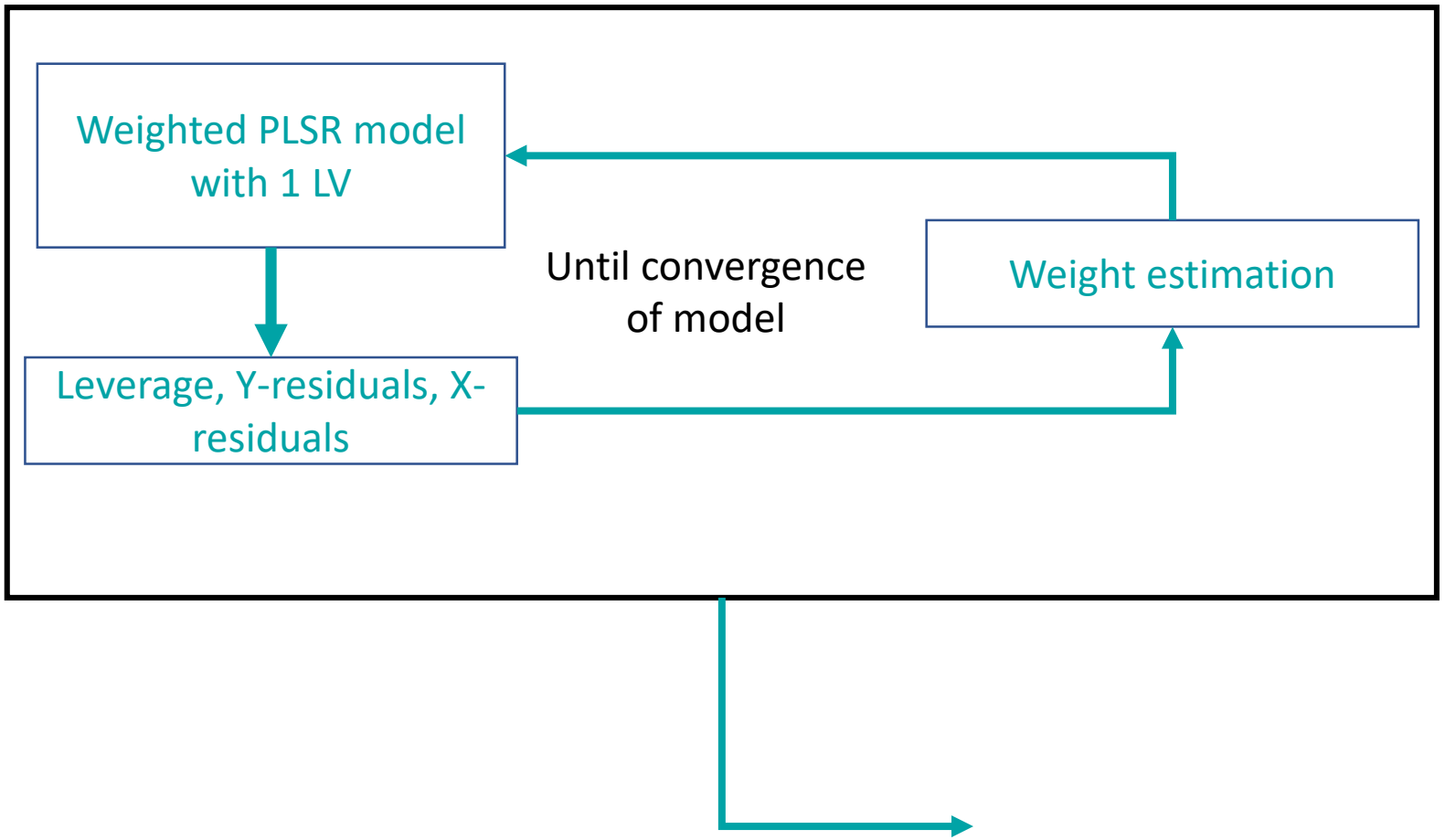
2. Theory

a. RoBoost-PLS



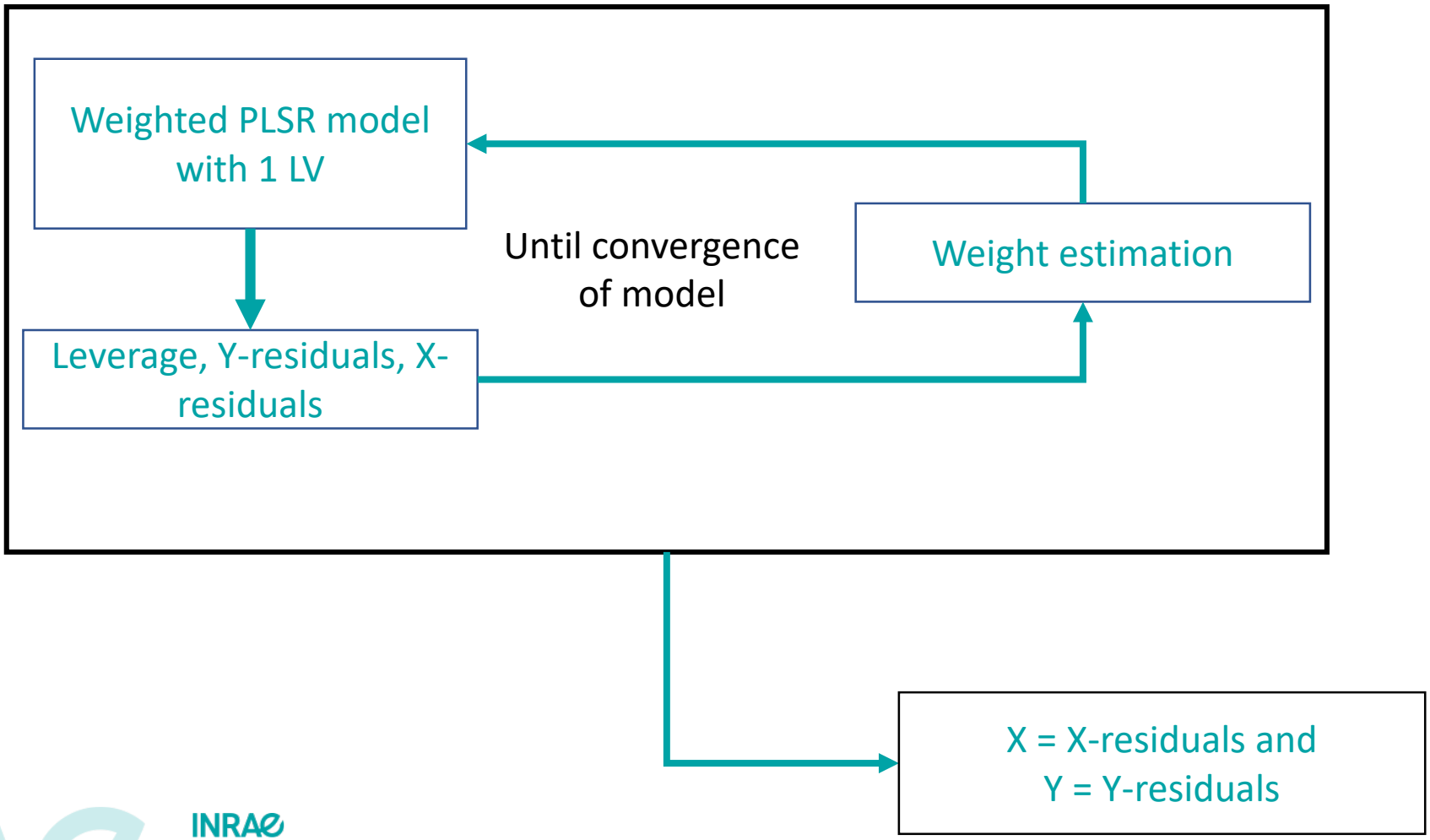
2. Theory

a. RoBoost-PLS



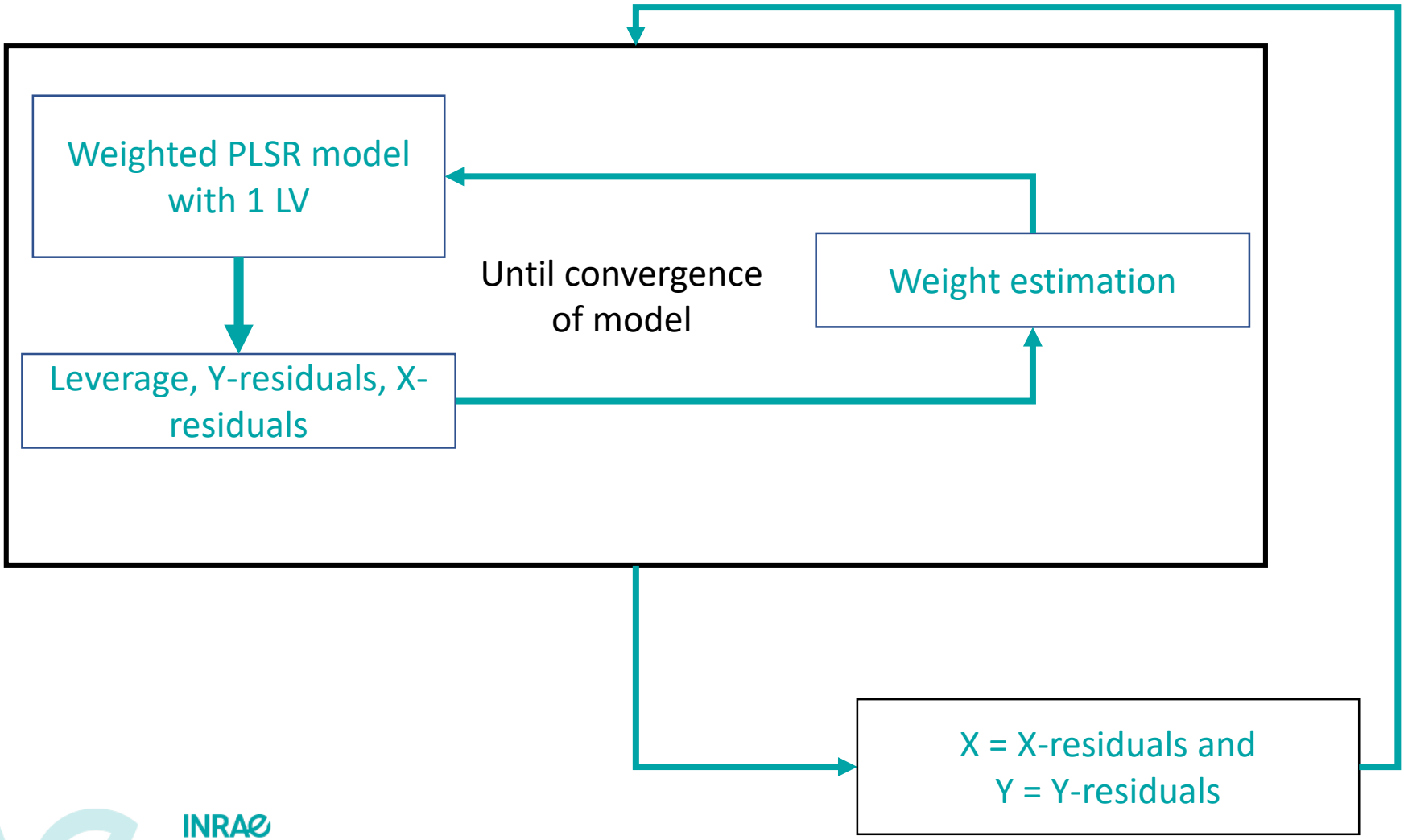
2. Theory

a. RoBoost-PLS



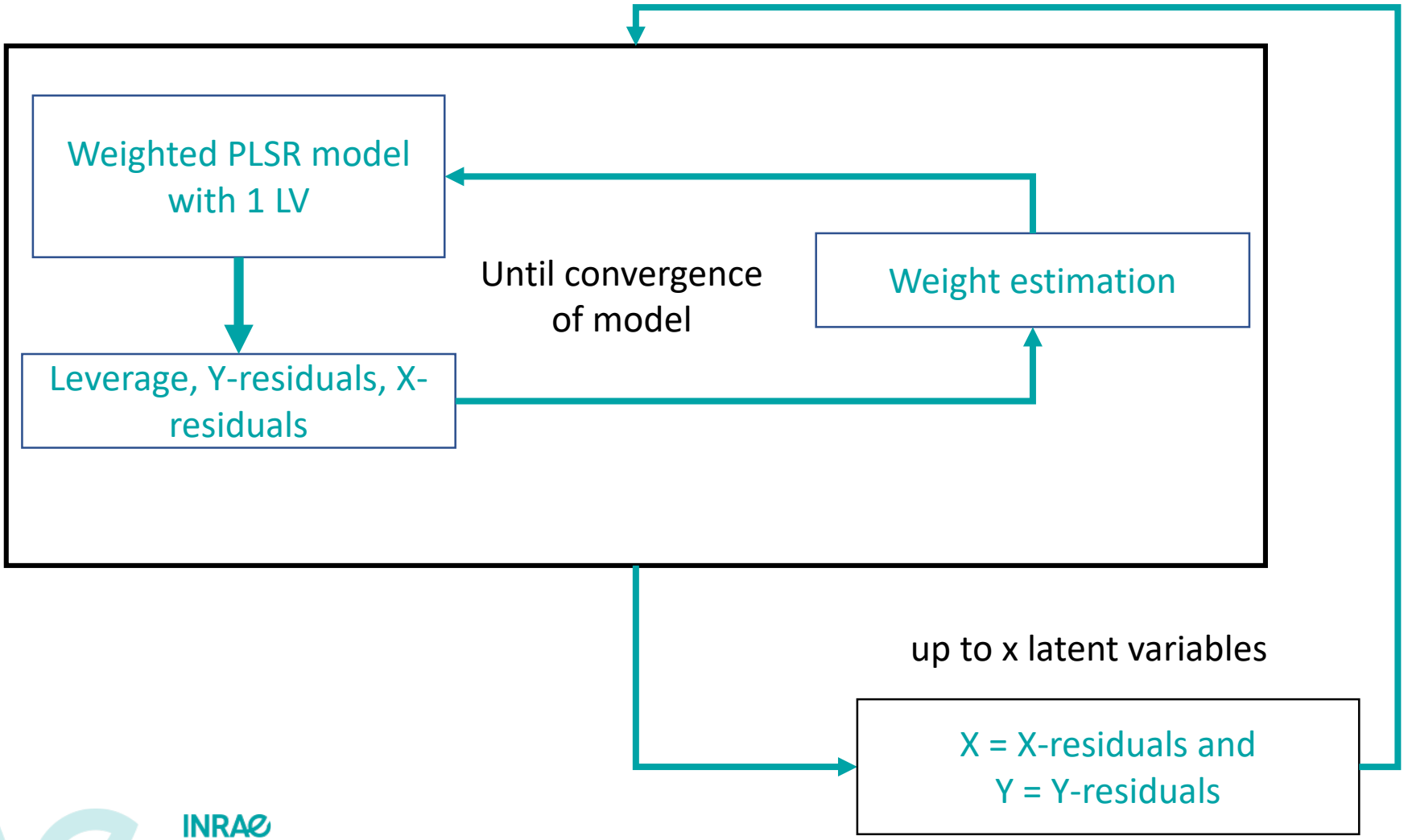
2. Theory

a. RoBoost-PLS



2. Theory

a. RoBoost-PLS



2. Theory

b. Pros and Cons

Pros	Cons
- Facilitates the weighting of leverage point - Apprehends X-residuals	- B-coef not observable - Scores : non-orthogonal

2. Theory

b. Pros and Cons

Pros	Cons
- Facilitates the weighting of leverage point - Apprehends X-residuals	- B-coef not observable - Scores : non-orthogonal

The distance to the center of the scores is easy to define for one dimension

3. Materials and methods

a. Methods

4 different methods :
PLSR with outliers in the training set
PLSR without outliers in the training set
PRM with outliers in the training set
RoBoost-PLSR with outliers in the training set

3. Materials and methods

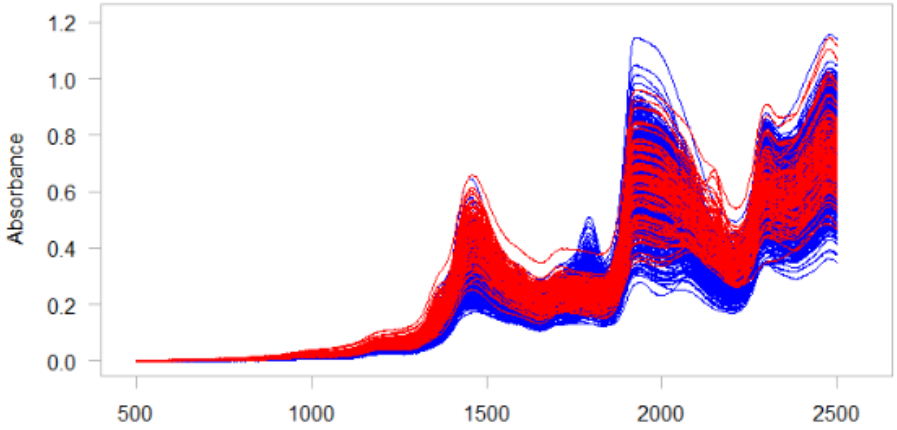
a. Methods

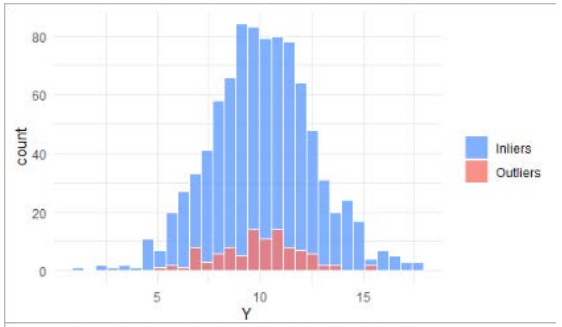
4 different methods :
PLSR with outliers in the training set
PLSR without outliers in the training set → Reference
PRM with outliers in the training set
RoBoost-PLSR with outliers in the training set

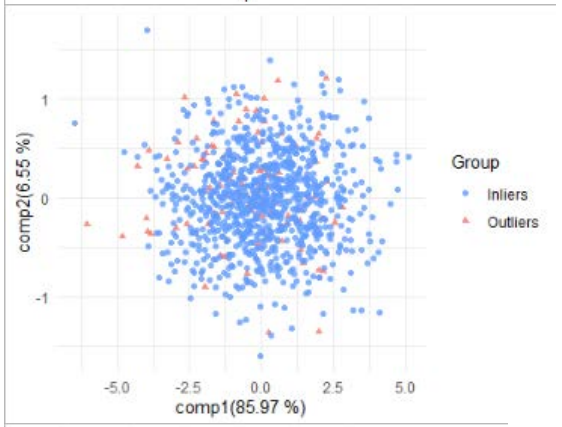
3. Materials and methods

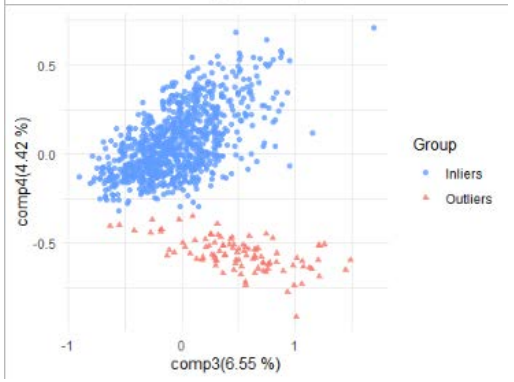
b. Data set

Simulated data set





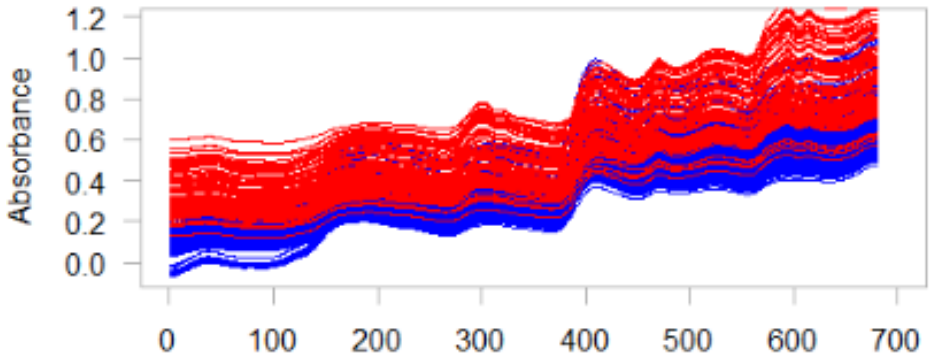


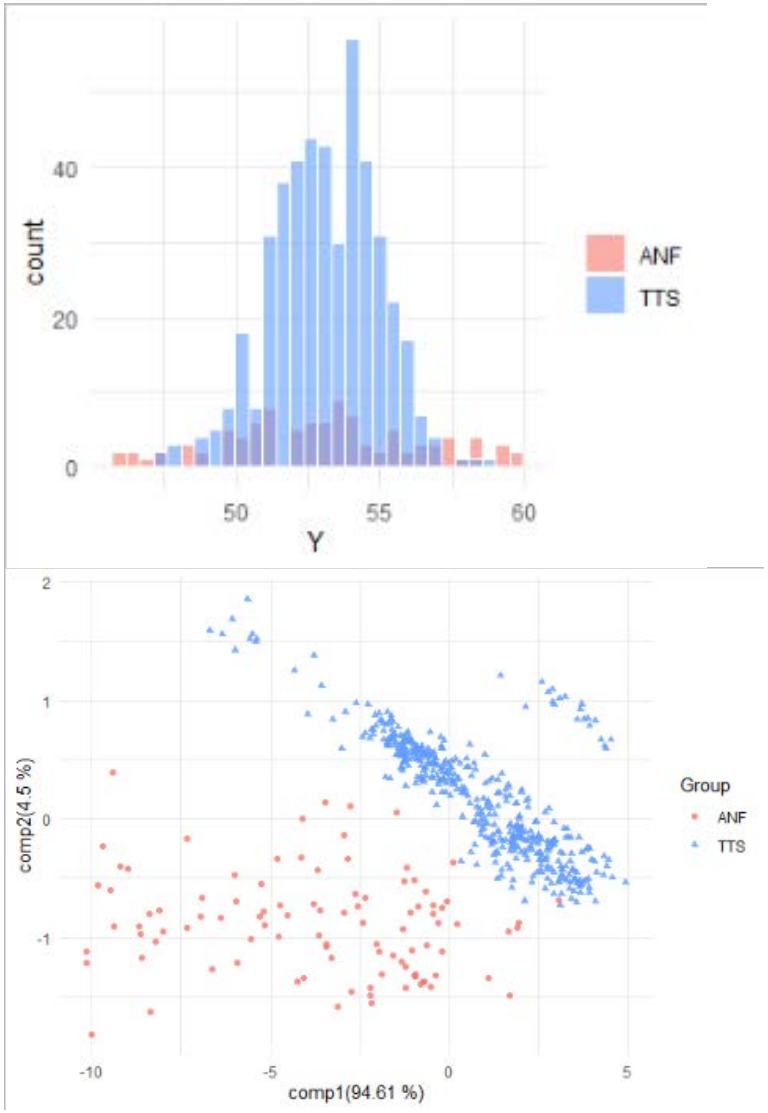


3. Materials and methods

b. Data set

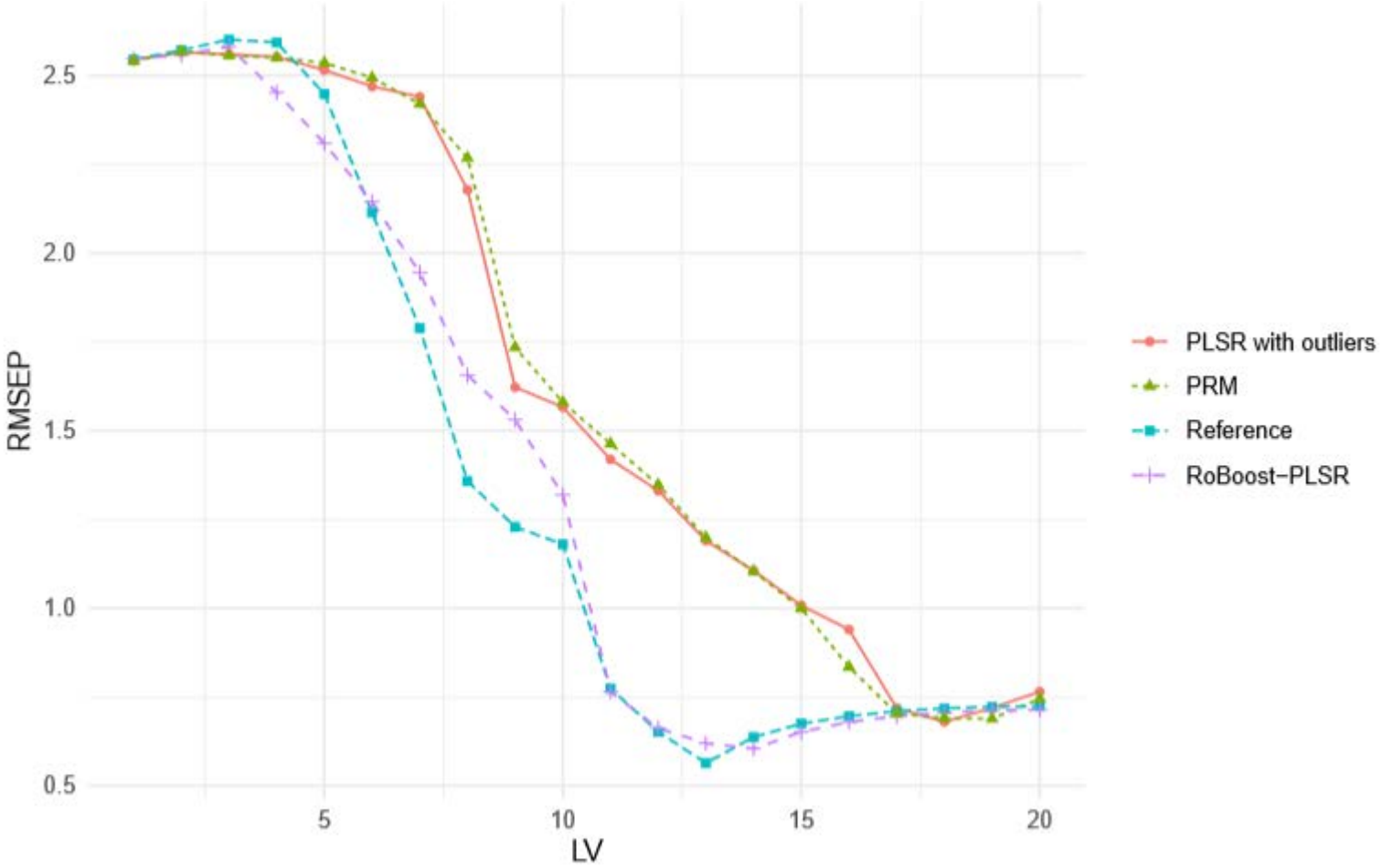
Real data set





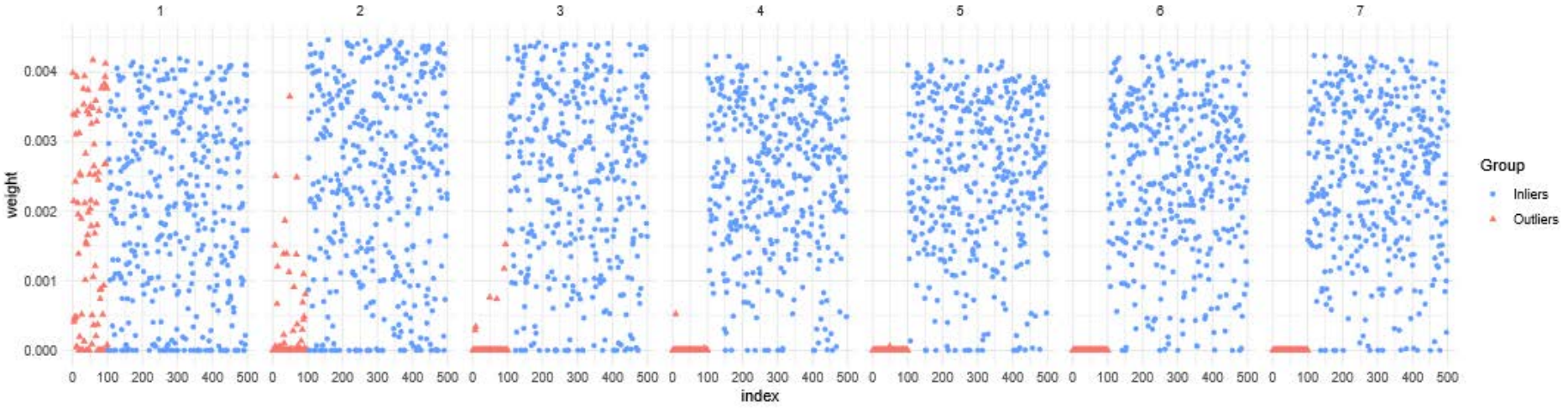
4. Results and discussions

a. Simulated data set



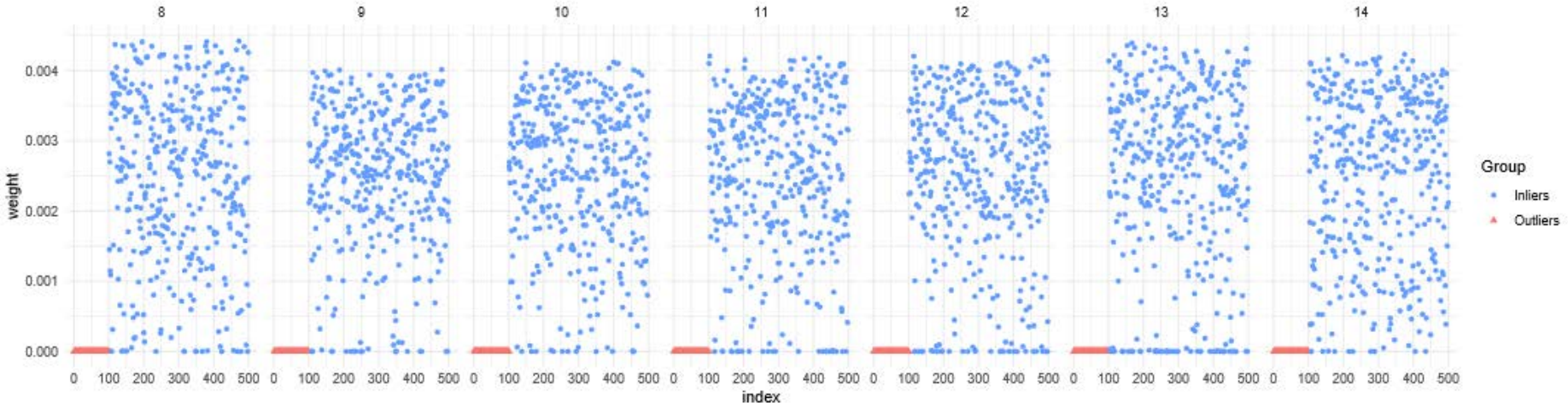
4. Results and discussions

a. Simulated data set



4. Results and discussions

a. Simulated data set

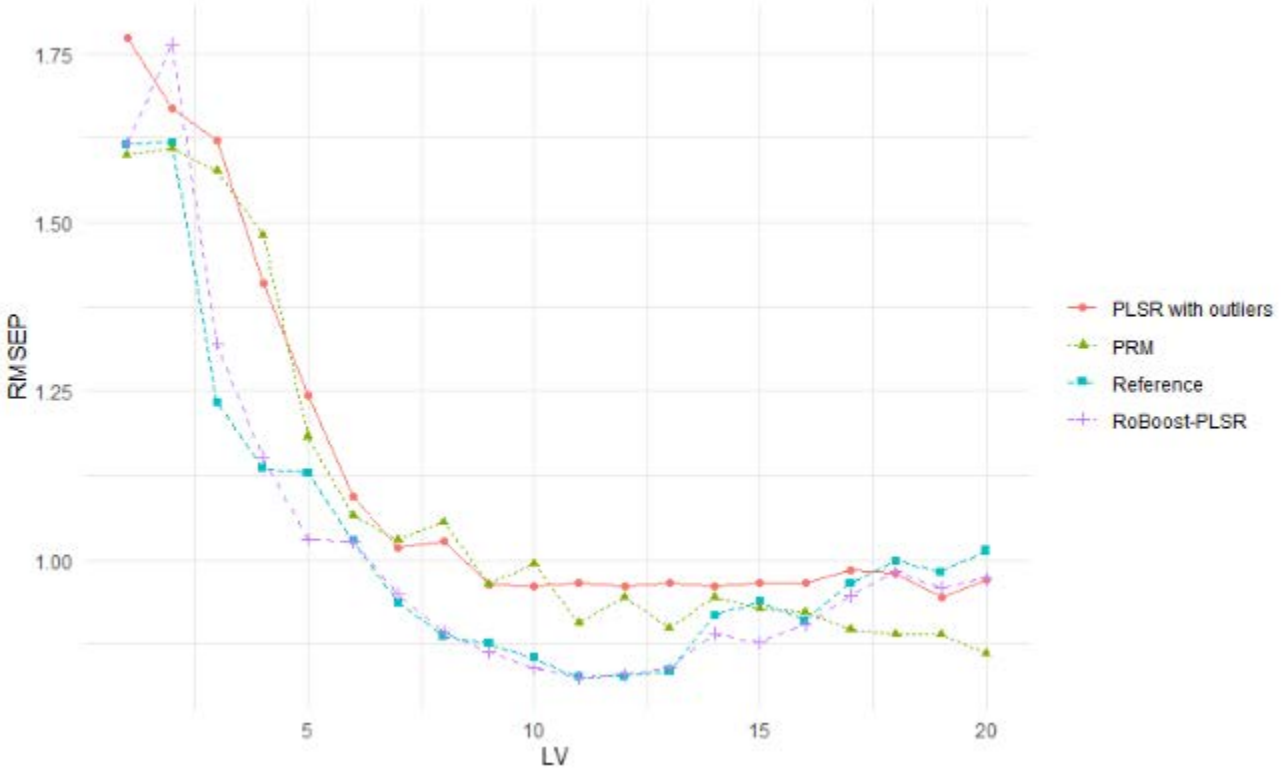


4. Results and discussions

b. Real data set

4. Results and discussions

b. Real data set

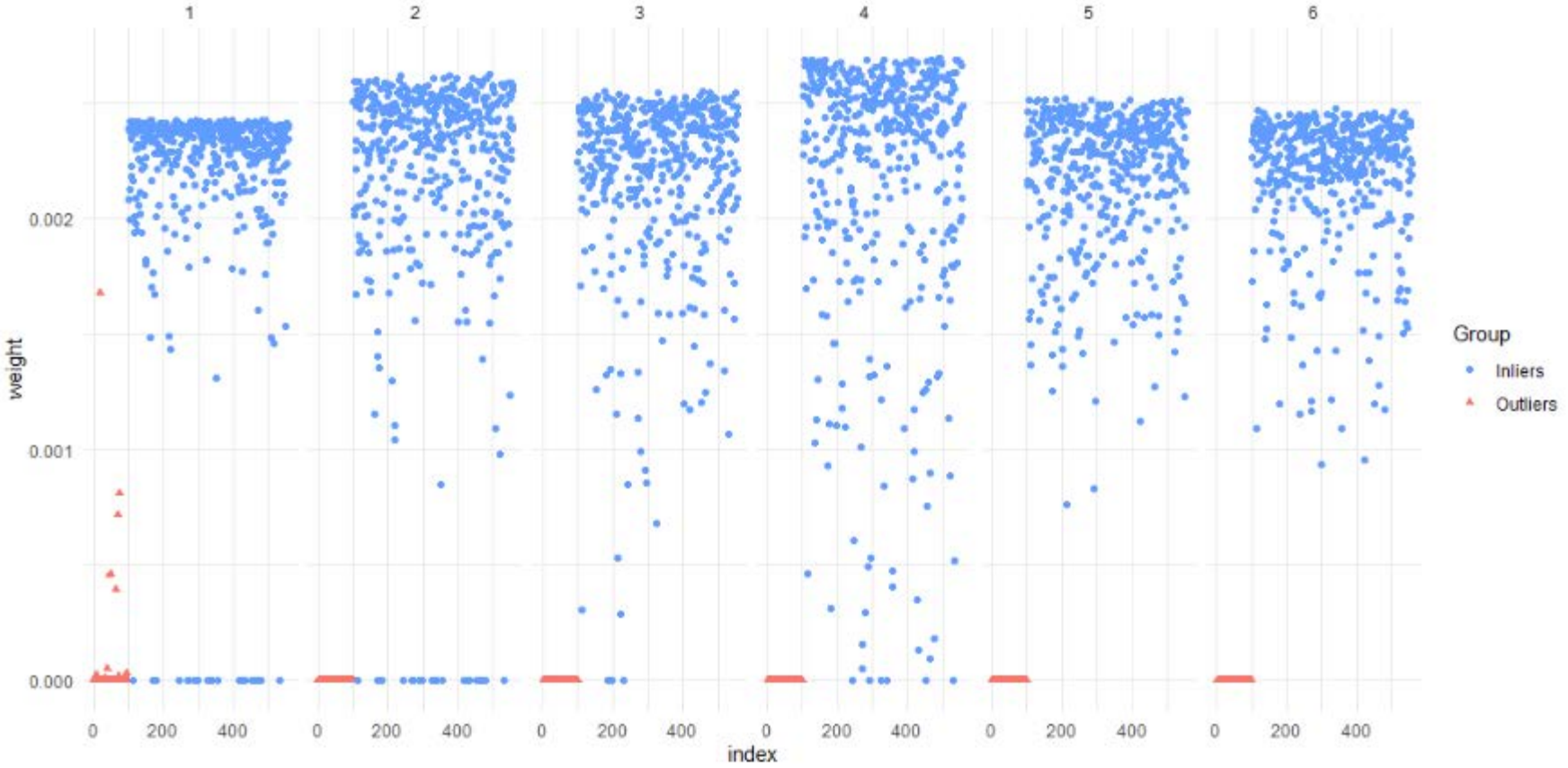


4. Results and discussions

b. Real data set

4. Results and discussions

b. Real data set

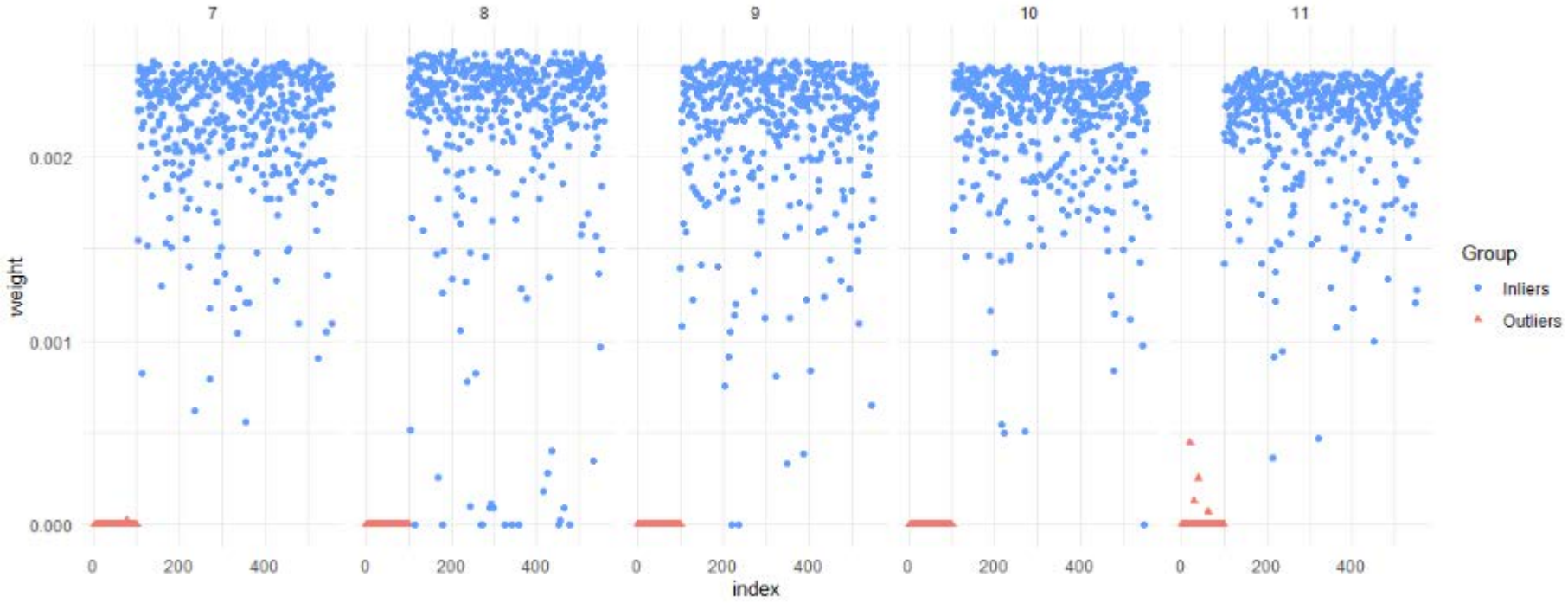


4. Results and discussions

b. Real data set

4. Results and discussions

b. Real data set



5. Conclusion

The results highlighted the good predictive capacity of the RoBoost-PLSR method, however some points need to be developed :

- Weight functions
- Weight optimisation
- Weight combinations
- Development of metrics for weighting
- The presence of outliers in the test
- Interpretation of the meta-model

Email : maxime.metz@inrae.fr