Change-point detection in an high-dimensional model with possibly asymmetric errors

Nicolas DULAC
Gabriela CIUPERCA
Cedric DEFFO-SIKOUNMO

ICJ - HD Technology

May 10, 2021

Table of Contents

The data

The dataset

The data

The dataset

- The data consists of monthly means of precipitation for all locations on the globe throughout the years starting in 1948.

The data

The dataset

- The data consists of monthly means of precipitation for all locations on the globe throughout the years starting in 1948.
- Locations in Eastern USA, Brazil, North East China, South Africa and India have been picked as target locations, and make up the dependent variable Y in the model.

The data

The dataset

- The data consists of monthly means of precipitation for all locations on the globe throughout the years starting in 1948.
- Locations in Eastern USA, Brazil, North East China, South Africa and India have been picked as target locations, and make up the dependent variable Y in the model.
- 40 locations near the target locations have been selected as the independent variables X in the model. The first 8 locations used as regressors are located near the first target location, the next 8 near the second target location and so on.

Math ICJ Institut Camille Jordan

HD Technology

The data

The dataset

- The data consists of monthly means of precipitation for all locations on the globe throughout the years starting in 1948.
- Locations in Eastern USA, Brazil, North East China, South Africa and India have been picked as target locations, and make up the dependent variable Y in the model.
- 40 locations near the target locations have been selected as the independent variables X in the model. The first 8 locations used as regressors are located near the first target location, the next 8 near the second target location and so on.
- For each of the target locations and the regressors we selected 400 values (from January 1948 up until April 1981). We concatenated the data to create a dataset with $n = 5 \times 400 = 2000$ observations and $p = 40$ regressors.

The dataset in picture

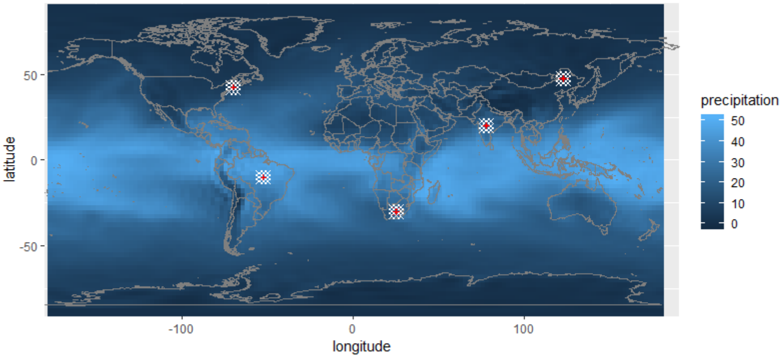


Figure: Data on the map.

Table of Contents

Institut Camille Jordan

HD Technology

Possible issues when modelling

Change points, feature selection, hypothesis...

Issues often faced in real life

Possible issues when modelling

Change points, feature selection, hypothesis...

Issues often faced in real life

1. In regression settings, change points may occur at random times and change the shape of the model.

Possible issues when modelling

Change points, feature selection, hypothesis...

Issues often faced in real life

1. In regression settings, change points may occur at random times and change the shape of the model.
2. Some assumptions needed for the classical least square method to be efficient might not be verified.

Possible issues when modelling

Change points, feature selection, hypothesis...

Issues often faced in real life

1. In regression settings, change points may occur at random times and change the shape of the model.
2. Some assumptions needed for the classical least square method to be efficient might not be verified.
3. When a large number of regressors is used, it is difficult to interpret the results.

Institut Camille Jordan

HD Technology

Proposed model

A model with K change points

$$Y_i = \mathrm{X}_i^\top \beta_1 \mathbb{1}_{1 \leqslant i \leqslant l_1} + \mathrm{X}_i^\top \beta_2 \mathbb{1}_{l_1 < i \leqslant l_2} + \cdots + \mathrm{X}_i^\top \beta_{K+1} \mathbb{1}_{l_K < i \leqslant n} + \varepsilon_i, \quad i = 1, \cdots, n. \quad (1)$$

Proposed model

A model with K change points

$$Y_i = \mathrm{X}_i^\top \beta_1 1\!\!1_{1 \leqslant i \leqslant l_1} + \mathrm{X}_i^\top \beta_2 1\!\!1_{l_1 < i \leqslant l_2} + \cdots + \mathrm{X}_i^\top \beta_{K+1} 1\!\!1_{l_K < i \leqslant n} + \varepsilon_i, \quad i = 1, \cdots, n. \quad (1)$$

- both the coefficients $\beta_1, \cdots, \beta_{K+1}$ and the locations $l_1, \cdots, l_K$ of the change-points are unknown.

$$Y_i = \mathrm{X}_i^\top \beta_1 \mathbb{1}_{1 \leqslant i \leqslant l_1} + \mathrm{X}_i^\top \beta_2 \mathbb{1}_{l_1 < i \leqslant l_2} + \cdots + \mathrm{X}_i^\top \beta_{K+1} \mathbb{1}_{l_K < i \leqslant n} + \varepsilon_i, \quad i = 1, \cdots, n. \quad (1)$$

- both the coefficients $\beta_1, \cdots, \beta_{K+1}$ and the locations $l_1, \cdots, l_K$ of the change-points are unknown.
- If K is unknown, its value will have to be estimated using a Schwarz-typed criterion.

Proposed model

A model with K change points

$$Y_i = \mathrm{X}_i^\top \beta_1 1\!\!1_{1 \leqslant i \leqslant l_1} + \mathrm{X}_i^\top \beta_2 1\!\!1_{l_1 < i \leqslant l_2} + \cdots + \mathrm{X}_i^\top \beta_{K+1} 1\!\!1_{l_K < i \leqslant n} + \varepsilon_i, \quad i = 1, \cdots, n. \quad (1)$$

- both the coefficients $\beta_1, \cdots, \beta_{K+1}$ and the locations $l_1, \cdots, l_K$ of the change-points are unknown.
- If K is unknown, its value will have to be estimated using a Schwarz-typed criterion.
- $(\varepsilon_i)_{1 \leqslant i \leqslant n}$ are i.i.d. such that $\mathbb{E}[\varepsilon_i^4] < \infty$.

Solving the issues
Estimating the parameters

Expectile LASSO adaptive process

In order to address the main issues mention on slide 6 we propose to estimate the parameters of the model (the K locations of the change points, as well as the $K+1$ sets of coeffecients in each phase) through an **expectile LASSO adaptive** objective function.

Institut Camille Jordan

HD Technology

Main results

Main results

- When the number of change points is known, then the estimators of the change points locations have an optimal convergence rate $O_{\mathbb{P}}(1)$

Institut Camille Jordan

HD Technology

Obtained results

adaptive LASSO expectile process

Main results

- When the number of change points is known, then the estimators of the change points locations have an optimal convergence rate $O_{\mathbb{P}}(1)$
- In each regime, the adaptive LASSO expectile estimators fulfill the sparsity property (all non significant coefficients are shrunk to 0), and the estimators of the nonzero coefficients are asymptotically Gaussian.

Main results

- When the number of change points is known, then the estimators of the change points locations have an optimal convergence rate $O_{\mathbb{P}}(1)$
- In each regime, the adaptive LASSO expectile estimators fulfill the sparsity property (all non significant coefficients are shrunk to 0), and the estimators of the nonzero coefficients are asymptotically Gaussian.
- We also propose a weakly consistent criterion for selecting K, the number of change points.

Table of Contents

Simulations study

Comparison with penalized least squares and quantile functions

Simulation settings

Simulation settings

- Model with 2 change points.

Simulation settings

- Model with 2 change points.
- $p = 10$ regressors, and samples of size $n = 200$ and $n = 500$.

Simulation settings

- Model with 2 change points.
- $p = 10$ regressors, and samples of size $n = 200$ and $n = 500$.
- 1000 Monte Carlo repetitions.

Institut Camille Jordan

HD Technology

Simulations study

Comparison with penalized least squares and quantile functions

Simulation settings

- Model with 2 change points.
- $p = 10$ regressors, and samples of size $n = 200$ and $n = 500$.
- 1000 Monte Carlo repetitions.
- 3 different types of errors (normal, mixture of normal and chi squared, and exponential).

Institut Camille Jordan

HD Technology

Simulation settings

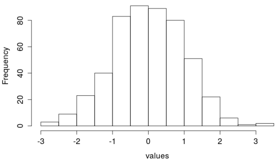
- Model with 2 change points.
- $p = 10$ regressors, and samples of size $n = 200$ and $n = 500$.
- 1000 Monte Carlo repetitions.
- 3 different types of errors (normal, mixture of normal and chi squared, and exponential).
- 3 different objective function used to estimate the parameters. Least squares (LS), expectile (EX), quantile (QU), all penalized with LASSO adaptive.

Simulation settings

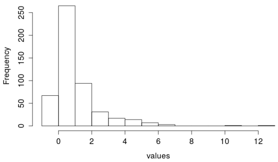
- Model with 2 change points.
- $p = 10$ regressors, and samples of size $n = 200$ and $n = 500$.
- 1000 Monte Carlo repetitions.
- 3 different types of errors (normal, mixture of normal and chi squared, and exponential).
- 3 different objective function used to estimate the parameters. Least squares (LS), expectile (EX), quantile (QU), all penalized with LASSO adaptive.
- Computation of several indicators (bias of coefficients, true positive rate & false positive rate of coefficients, precision of location of change points, execution time).

Errors

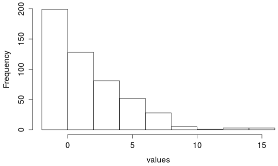
Homoscedastic and heteroscedastic



(a) Errors of type $\varepsilon \sim \mathcal{N}(0,1)$



(b) Errors of type $\varepsilon \sim 0.2 \cdot \mathcal{N}(0,1) + \chi^2_1$



(c) Errors of type $\varepsilon \sim 3 \cdot \mathcal{E}(1) - 1.5$

Figure: Histograms for 500 realizations of each type of error

Simulation study results

Normal & mixed errors

n	loss	$\left\|\widehat{\beta}-\beta^{0}\right\|_2$	$\% \sum_{r=1}^{2}(\mathcal{A}_r^{0} \subseteq \widehat{\mathcal{A}}_r)$	$\% \sum_{r=1}^{2}(\widehat{\mathcal{A}}_r \cap \mathcal{A}_r^{0\,c} \neq \emptyset)$	$\lvert\widehat{l}/n - l^{0}/n\rvert$	time (s)
200	LS	0.704	100%	0.133%	0.00243	0.00287
	EX	0.704	100%	0.127%	0.00233	0.00299
	QU	0.694	100%	1.20%	0.00249	0.00402
500	LS	0.450	100%	6.67e-5%	2.2e-5	0.00303
	EX	0.450	100%	6.67e-5%	2.4e-5	0.00308
	QU	0.422	100%	0.273%	4.2e-5	0.00549

Table: Results when $\varepsilon \sim \mathcal{N}(0,1)$.

n	loss	$\left\|\widehat{\beta}-\beta^{0}\right\|_2$	$\% \sum_{r=1}^{2}(\mathcal{A}_r^{0} \subseteq \widehat{\mathcal{A}}_r)$	$\% \sum_{r=1}^{2}(\widehat{\mathcal{A}}_r \cap \mathcal{A}_r^{0\,c} \neq \emptyset)$	$\lvert\widehat{l}/n - l^{0}/n\rvert$	time (s)
200	LS	0.938	99.6%	2.69%	0.000635	0.00289
	EX	0.955	99.7%	0.000133%	0.00480	0.00330
	QU	0.380	100%	6.67e-5%	0.00455	0.00406
500	LS	0.572	100%	0.233%	3.2e-5	0.00304
	EX	0.558	100%	0%	1.4e-5	0.00371
	QU	0.181	100%	0%	4e-6	0.00562

Table: Results when $\varepsilon \sim 0.2 \cdot \mathcal{N}(0,1) + \chi_1^2$.

Simulation study results

Exponential errors

n	loss	$\lVert\widehat{\beta}-\beta^0\rVert_2$	$\% \sum_{r=1}^{2}(\mathcal{A}_r^0 \subseteq \widehat{\mathcal{A}}_r)$	$\% \sum_{r=1}^{2}(\widehat{\mathcal{A}}_r \cap \mathcal{A}_r^{0^c} \neq \emptyset)$	$\lvert\widehat{l}/n - l^0/n\rvert$	time (s)
	LS	1.72	97.2%	25.0%	0.00403	0.00291
200	EX	1.30	98.7%	9.26%	0.00374	0.00303
	QU	1.40	98.2%	8.29%	0.00456	0.00400
	LS	1.02	99.8%	13.7%	0.00313	0.00306
500	EX	0.722	100%	2.06%	0.00338	0.00334
	QU	0.812	99.9%	3.15%	0.00334	0.00549

Table: Results when $\varepsilon \sim 3 \cdot \mathcal{E}(1) - 1.5$.

Institut Camille Jordan

HD Technology

Summary of the simulation study

key takeaways

Summary of the simulation study

key takeaways

1. The adaptive LASSO expectile process performs as well as the adaptive LASSO least squares process when homoscedasticity is verified.

Summary of the simulation study

key takeaways

1. The adaptive LASSO expectile process performs as well as the adaptive LASSO least squares process when homoscedasticity is verified.
2. The penalized expectile process features similar performances as the penalized quantile process when homoscedasticity is not verified.

Summary of the simulation study

key takeaways

1. The adaptive LASSO expectile process performs as well as the adaptive LASSO least squares process when homoscedasticity is verified.
2. The penalized expectile process features similar performances as the penalized quantile process when homoscedasticity is not verified.
3. The adaptive LASSO expectile process is faster than the adaptive LASSO quantile process, and does not encounter numerical problems.

Application on real data

Weather data

- First we apply our criterion to determine the number of change points. We search for $K \in 0, 1, 2, 3, 4, 5, 6$. The smallest criterion value is obtained for $K = 4$.

Application on real data

Weather data

- First we apply our criterion to determine the number of change points. We search for $K \in 0, 1, 2, 3, 4, 5, 6$. The smallest criterion value is obtained for $K = 4$.
- Then, we estimate the locations of the 4 change points and the coefficients simultaneously. The four change points locations are estimated as $\hat{l}_1 = 401,\ \hat{l}_2 = 801,\ \hat{l}_3 = 1201,\ \hat{l}_4 = 1601.$

Math Institut Camille Jordan

HD Technology

Application on real data

Weather data

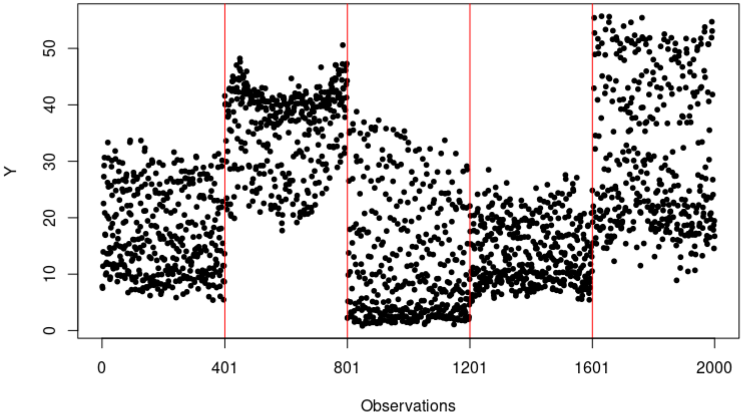
- First we apply our criterion to determine the number of change points. We search for $K \in 0, 1, 2, 3, 4, 5, 6$. The smallest criterion value is obtained for $K = 4$.
- Then, we estimate the locations of the 4 change points and the coefficients simultaneously. The four change points locations are estimated as $\hat{l}_1 = 401$, $\hat{l}_2 = 801$, $\hat{l}_3 = 1201$, $\hat{l}_4 = 1601$.
- In each regime, only a few (between 2 and 4) coefficients are estimated as significant. Those significant coefficients correspond to locations near the target location associated with the regime (e.g. for the first segment, Eastern USA, the two nonzero coefficients are two of the first eight coefficients).

- First we apply our criterion to determine the number of change points. We search for $K \in 0, 1, 2, 3, 4, 5, 6$. The smallest criterion value is obtained for $K = 4$.
- Then, we estimate the locations of the 4 change points and the coefficients simultaneously. The four change points locations are estimated as $\hat{l}_1 = 401,\ \hat{l}_2 = 801,\ \hat{l}_3 = 1201,\ \hat{l}_4 = 1601$.
- In each regime, only a few (between 2 and 4) coefficients are estimated as significant. Those significant coefficients correspond to locations near the target location associated with the regime (e.g. for the first segment, Eastern USA, the two nonzero coefficients are two of the first eight coefficients).

$\Rightarrow$ In conclusion, our method was able to detect the change points accurately, and to select some appropriate significant coefficients.

The obtained regimes

Figure: The different regimes.



The end

Thank you for your time, feel free to ask questions!