

Détections de change-points dans un modèle en grande dimension avec des erreurs asymétriques ou avec des outliers

Gabriela Ciuperca

Département de Mathématiques,
Institut Camille Jordan,
Université Lyon 1

Journées MAS 2022,
Rouen, 29-31 Août 2022

Plan

I. Méthodes LASSO adaptatives de sélection automatique de variables

- 1 Modèle quantile avec LASSO adaptatif
- 2 Modèle expectile avec LASSO adaptatif

II. Modèles avec change-points

- 1 Détection rétroactive des changements
- 2 Détection en temps réel d'un changement

I. Méthodes LASSO adaptatives de sélection automatique de variables

Modèle linéaire:

$$Y_i = \mathbf{X}_i^\top \boldsymbol{\beta} + \varepsilon_i, \quad i = 1, \dots, n, \quad (1)$$

avec $\boldsymbol{\beta} \in \mathbb{R}^p$ paramètres, $\boldsymbol{\beta}^0$ la vraie valeur (inconnue) de $\boldsymbol{\beta}$.

- Cas classique: erreurs ε_i i.i.d., $\mathbb{E}[\varepsilon] = 0$ et $\text{Var}(\varepsilon) < \infty$.
 - Si $\dim(\boldsymbol{\beta})$ petite, alors, on estime $\boldsymbol{\beta}$ par moindres carrés (MC).
 - Si $\dim(\boldsymbol{\beta})$ très grande, alors, on estime $\boldsymbol{\beta}$ par MC avec une pénalité LASSO adaptative (Zou, 2006):

$$\sum_{i=1}^n (Y_i - \mathbf{X}_i^\top \boldsymbol{\beta})^2 + \lambda_n \widehat{\boldsymbol{\omega}}_n^\top |\boldsymbol{\beta}|,$$

avec $\lambda_n \rightarrow \infty$ paramètre de calibrage (tuning), $\widetilde{\boldsymbol{\beta}}^{LS}$ estimateur par moindres carrés, $\widehat{\boldsymbol{\omega}}_n^\top = (|\widetilde{\beta}_1^{LS}|^{-g}, \dots, |\widetilde{\beta}_p^{LS}|^{-g})$, $g > 0$.

- Si $\mathbb{E}[\varepsilon^2]$ n'existe pas, mais $\mathbb{E}[\text{sgn}(\varepsilon)] = 0$, on peut utiliser la méthode des moindres déviations pénalisée.
- Si les quatre premiers moments de ε existent mais la répartition de ε est asymétrique, alors la méthode des moindres carrés n'est pas appropriée. La méthode expectile peut être utilisée.
- Si les deux premiers moments de ε n'existent pas, on peut utiliser la méthode d'estimation quantile.

I.1. Modèle quantile avec LASSO adaptatif

- Pour un *niveau* $\tau \in (0, 1)$ fixé, soit la *quantile* b_τ d'ordre τ pour ε :

$$\tau = \mathbb{P}[\varepsilon < b_\tau].$$

On suppose b_τ inconnu, $\beta = (\beta_1, \dots, \beta_p) \in \mathbb{R}^p$, avec p très grand, $p < n$.

- Estimation, en minimisant le processus quantile:

$$(\tilde{b}_n^{(Q)}, \tilde{\beta}_n^{(Q)}) \equiv \arg \min_{(b, \phi)} \sum_{i=1}^n \rho_\tau(Y_i - b - \mathbf{X}_i^\top \beta), \quad (2)$$

la fonction $\rho_\tau(\cdot) : \mathbb{R} \rightarrow \mathbb{R}$ définie par $\rho_\tau(r) = r[\tau \mathbf{1}_{r>0} - (1 - \tau) \mathbf{1}_{r \leq 0}]$.

- Si (b_τ^0, β^0) sont les vraies valeurs (inconnues), alors les *estimateurs quantile*

$$(\tilde{b}_n^{(Q)}, \tilde{\beta}_n^{(Q)}) \xrightarrow[n \rightarrow \infty]{p.s.} (b_\tau^0, \beta^0),$$

$$\sqrt{n}(\tilde{\beta}_n^{(Q)} - \beta^0) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}\left(\mathbf{0}, \frac{\tau(1-\tau)}{f^2(b_\tau^0)} \Upsilon^{-1}\right).$$

- Pour $\tau = 1/2$, on obtient l'estimateur des moindres déviations.

- *Avantage* de la méthode: estimateurs robustes.
- *Inconvénient*: comme d'autres méthodes d'estimation non pénalisées (MC, MV), les estimateurs quantiles n'ont pas les *propriétés oracle*:
 - *propriété de sparsité*: les composantes de β^0 égales à 0 sont estimées directement par 0, avec une probabilité tendant vers 1;
 - les paramètres non-nuls ont des estimateurs asymptotiquement normaux, avec la vitesse de convergence optimale $n^{-1/2}$.

↪ *Estimateur quantile LASSO adaptatif*:

$$(\hat{b}_n, \hat{\beta}_n) \equiv \arg \min_{(b, \beta)} \left(\sum_{i=1}^n \rho_\tau(Y_i - b - \mathbf{x}_i^\top \beta) + \lambda_n \hat{\omega}_n^\top |\beta| \right), \quad (3)$$

avec $\hat{\omega}_n^\top = (\hat{\omega}_{n,1}, \dots, \hat{\omega}_{n,p}) \equiv \frac{1}{|\hat{\beta}_{n,1}^{(0)}|^g} = \left(\frac{1}{|\hat{\beta}_{n,1}^{(0)}|^g}, \dots, \frac{1}{|\hat{\beta}_{n,p}^{(0)}|^g} \right)$, $g > 0$.

$(\lambda_n)_{n \in \mathbf{N}}$ une suite de régularisation $\rightarrow \infty$ pour $n \rightarrow \infty$.

Suppositions:

(A1) $n^{-1} \max_{1 \leq i \leq n} \mathbf{X}_i^\top \mathbf{X}_i \xrightarrow[n \rightarrow \infty]{} 0$ et $n^{-1} \sum_{i=1}^n \mathbf{X}_i \mathbf{X}_i^\top \xrightarrow[n \rightarrow \infty]{} \Upsilon$, avec Υ une matrice définie positive.

(A2) Pour tout $\mathbf{e} \in \text{int}(\mathcal{B})$, $u_0 \in \mathbb{R}$, $\mathbf{u} \in \mathbb{R}^p$ on a

$$\lim_{n \rightarrow \infty} \frac{1}{n} \sum_{i=1}^n \int_0^{u_0 + \mathbf{x}_i^\top \mathbf{u}} \sqrt{n} [F(\mathbf{e} + \frac{\mathbf{v}}{\sqrt{n}}) - F(\mathbf{e})] d\mathbf{v} = \frac{f(\mathbf{e})}{2} (u_0, \mathbf{u}^\top) \begin{bmatrix} 1 & \mathbf{0} \\ 0 & \Upsilon \end{bmatrix} (u_0, \mathbf{u}^\top)$$

Propriétés oracle

Soit les ensembles $\mathcal{A}^0 \equiv \{j \in \{1, \dots, p\}; \beta_j^0 \neq 0\}$,

$$\hat{\mathcal{A}}_n \equiv \{j \in \{1, \dots, p\}; \hat{\beta}_{n,j} \neq 0\}.$$

Théorème

Sous (A1), (A2), si $n^{-1/2} \lambda_n \rightarrow 0$ et $n^{(g-1)/2} \lambda_n \rightarrow \infty$, alors:

(i) $\sqrt{n}(\hat{\beta}_{\mathcal{A}^0} - \beta_{\mathcal{A}^0}^0) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}\left(\mathbf{0}_p, \frac{\tau(1-\tau)}{f^2(\mathbf{b}^0)} \Upsilon_{\mathcal{A}^0}^{-1}\right)$.

(ii) Si de plus $n^{g/2-1} \lambda_n \rightarrow \infty$, alors $\mathbb{P}[\hat{\mathcal{A}}_n = \mathcal{A}^0] \rightarrow 1$, pour $n \rightarrow \infty$.

Idée de preuve:

- Conditions d'optimalité de Karush-Kuhn-Tucker (KKT):

- pour tout $j \in \hat{\mathcal{A}}_n$, $\tau \sum_{i=1}^n X_{ij} - \sum_{i=1}^n X_{ij} \mathbf{1}_{Y_i < \mathbf{x}_i^\top \hat{\beta}_n} = \lambda_n \hat{\omega}_{n,j} \text{sgn}(\hat{\beta}_{n,j})$.

- pour tout $j \notin \hat{\mathcal{A}}_n$, on a $\left| \tau \sum_{i=1}^n X_{ij} - \sum_{i=1}^n X_{ij} \mathbf{1}_{Y_i < \mathbf{x}_i^\top \hat{\beta}_n} \right| \leq \lambda_n \hat{\omega}_{n,j}$.

- (i) Reparamétrisation $\mathbf{u}_n = \sqrt{n}(\hat{\beta}_n - \beta^0)$ et $u_{0,n} = \sqrt{n}(\hat{b}_n - b^0)$.

$(u_{0,n}, \mathbf{u}_n)$ le minimiseur du processus:

$\sum_{i=1}^n \left[\rho_\tau \left(\varepsilon_i - b^0 - n^{-1/2}(u_0 + \mathbf{x}_i^\top \mathbf{u}) \right) - \rho_\tau(\varepsilon_i - b^0) \right] + \lambda_n \hat{\omega}_n^\top [|\beta^0 + n^{-1/2} \mathbf{u}| - |\beta^0|]$, qui converge en loi vers

$$z_0 u_0 + \mathbf{z}^\top \mathbf{u} + \frac{1}{2} f(b_\tau^0) u_0^2 + \frac{1}{2} \mathbf{u}^\top \Upsilon \mathbf{u} + \sum_{j=1}^p W(\beta_j, u_j)$$

avec W soit 0 (quand $\beta_j^0 \neq 0$ ou $\beta_j^0 = 0, u_j = 0$) soit ∞ quand $\beta_j^0 = 0, u_j \neq 0$), et z_0, \mathbf{z} processus gaussiens centrés.

- (ii) Par (i) on a $\lim_{n \rightarrow \infty} \mathbb{P}[\hat{\mathcal{A}}_n \supseteq \mathcal{A}^0] = 1$.

Pour l'inclusion inverse, en utilisant les conditions KKT, on montre que $\mathbb{P}[j \in \hat{\mathcal{A}}_n, j \notin \mathcal{A}^0] \rightarrow 0$, quand $n \rightarrow \infty$.

(G. Ciuperca, *Statistics*, 2016)

I.2. Modèle expectile avec LASSO adaptatif

Soit $\tau \in]0, 1[$ le niveau expectile (fixé) et soit la fonction $\rho_\tau : \mathbb{R} \rightarrow \mathbb{R}_+$

$$\rho_\tau(r) = |r - \mathbf{1}_{r < 0}|r^2, \quad \forall r \in \mathbb{R}.$$

Définissons également les fonctions:

$$d_\tau(r) = \rho'_\tau(r - t)|_{t=0} = 2\tau r \mathbf{1}_{r \geq 0} + 2(1 - \tau)r \mathbf{1}_{r < 0},$$

$$h_\tau(r) = \rho''_\tau(r - t)|_{t=0} = 2\tau \mathbf{1}_{r \geq 0} + 2(1 - \tau) \mathbf{1}_{r < 0}.$$

Concernant les erreurs $(\varepsilon_i)_{1 \leq i \leq n}$ et le design $(\mathbf{X}_i)_{1 \leq i \leq n}$ les suppositions suivantes sont considérées (voir Liao et al. (2019) et Ciuperca(2021)):

- **(B1)** $(\varepsilon_i)_{1 \leq i \leq n}$ sont i.i.d. avec $\mathbb{E}[\varepsilon^4] < \infty$ et $\mathbb{E}[d_\tau(\varepsilon)] = 0$.

- **(B2)** $\Upsilon = \lim_{n \rightarrow \infty} n^{-1} \sum_{i=1}^n \mathbf{X}_i \mathbf{X}_i^\top$ et Υ est définie positive.

Remarques. 1) On utilise la méthode expectile quand la loi de ε est asymétrique ou elle a des valeurs extrêmes.

2) Pour $\tau = 1/2$ on obtient la méthode des moindres carrés.

Définition *L'estimateur expectile* de niveau τ du paramètre β est défini par:

$$\tilde{\beta}_n^{(E)}(\tau) = \arg \min_{\beta \in \mathbb{R}^p} \sum_{i=1}^n \rho_\tau(Y_i - \mathbf{X}_i^\top \beta)$$

Remarque. Comme l'estimateur quantile, l'estimateur expectile n'a pas de forme explicite et dépend de l'ordre τ tel que la supposition $\mathbb{E}[d_\tau(\varepsilon)] = 0$ de (B1) soit satisfaite.

Les éléments du vecteur sont $\tilde{\beta}_n^{(E)} = (\tilde{\beta}_{n,1}^{(E)}, \dots, \tilde{\beta}_{n,p}^{(E)})$. L'estimateur expectile LASSO adaptatif $\hat{\beta}_n^{aeLASSO}$ de β est défini par:

$$\hat{\beta}_n^{aeLASSO} = \arg \min_{\beta \in \mathbb{R}^p} \left[\sum_{i=1}^n \rho_\tau(Y_i - \mathbf{x}_i^\top \beta) + \lambda_n \sum_{j=1}^p \hat{\omega}_{n,j}^{(E)} |\beta_j| \right], \quad (4)$$

avec $\hat{\omega}_{n,j}^{(E)} = |\tilde{\beta}_{n,j}^{(E)}|^{-g}$ et $g > 0$ un paramètre connu.

Les éléments du vecteur $\hat{\beta}_n^{aeLASSO}$ sont $(\hat{\beta}_{n,1}^{aeLASSO}, \dots, \hat{\beta}_{n,p}^{aeLASSO})$.

Considérons l'ensemble des indices:

$$\hat{\mathcal{A}}_n^{aeLASSO} = \{j \in \{1, \dots, p\}, \hat{\beta}_{n,j}^{aeLASSO} \neq 0\}.$$

Theorem

Sous les suppositions (B1), (B2) si $(\lambda_n)_{n \in \mathbf{N}}$ est tel que $\lambda_n \rightarrow \infty$, $n^{-1/2} \lambda_n \xrightarrow[n \rightarrow \infty]{} 0$ alors $\|\hat{\beta}_n^{aeLASSO} - \beta^0\|_2 = O_{\mathbb{P}}(n^{-1/2})$.

Theorem

(Théorème 3 de Liao et al.(2019)) Si les supposition (B1), (B2) sont vraies et le paramètre de tuning satisfait $(\lambda_n)_{n \in \mathbf{N}}$ est tel que $\lambda_n \rightarrow \infty$, $n^{-1/2} \lambda_n \xrightarrow[n \rightarrow \infty]{} 0$ et $n^{g-1)/2} \lambda_n \xrightarrow[n \rightarrow \infty]{} \infty$. Alors:

(i) $\mathbb{P}[\hat{\mathcal{A}}_n^{aeLASSO} = \mathcal{A}] \rightarrow 1$, pour $n \rightarrow \infty$.

(ii) $n^{1/2}(\hat{\beta}_n^{aeLASSO} - \beta^0)_{\hat{\mathcal{A}}_n^{aeLASSO}} \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}(\mathbf{0}_p, \frac{\text{var}[d_{\tau}(\varepsilon)]}{\mathbb{E}^2[h_{\tau}(\varepsilon)]} \Upsilon_{\mathcal{A}}^{-1})$.

(G. Ciuperca, CSDA, 2021). (L. Liao et al., AISM, 2019)

- 1 Méthodes LASSO adaptatives de sélection automatique de variables
 - Modèle quantile avec LASSO adaptatif
 - Modèle expectile avec LASSO adaptatif
- 2 Modèles avec change-points (rupture)
 - Détection rétroactive des changements
 - Détection d'un changement en temps réel

Deux types de problèmes de rupture:

- *rétroactif (à posteriori)*: on traite les données une fois que l'événement s'est terminé.
- *on-line (en temps réel)*: détection d'un changement dans le modèle au fur et à mesure que les observations sont réalisées.

Veiller que la propriété de sparsité (sélection automatique des variables) reste valable (elle n'est pas influencée par l'estimation des change-points).

2.1. Détection rétroactive des changements

Considérons un modèle à $(K + 1)$ phases:

$$Y_i = \mathbf{X}_i^\top \beta_1 \mathbf{1}_{1 \leq i < l_1} + \mathbf{X}_i^\top \beta_2 \mathbf{1}_{l_1 \leq i < l_2} + \dots + \mathbf{X}_i^\top \beta_{K+1} \mathbf{1}_{l_K \leq i \leq n} + \varepsilon_i, \quad i = 1, \dots, n.$$

Pour K fixé, les paramètres inconnus du modèle sont:

- par la méthode quantile: les quantiles d'ordre τ des erreurs pour chaque segment;
- les coefficients β_k , $k = 1, \dots, K + 1$, pour chaque segment;
- les change-points l_1, \dots, l_K .

On estime ces paramètres en deux étapes (avec la fonction de perte ρ_τ soit quantile, soit expectile):

$$\textcircled{1} \quad S(l_1, \dots, l_K) \equiv \inf_{(\beta_r, b_r, \dots)} \sum_{r=1}^{K+1} \left[\sum_{i=l_{r-1}+1}^{l_r} \rho_\tau(Y_i - b_r - \mathbf{X}_i^\top \beta_r) + \lambda_{(l_{r-1}; l_r)} \hat{\omega}_{(l_{r-1}; l_r)}^\top | \beta_r | \right].$$

et alors $(\hat{l}_1, \dots, \hat{l}_K) \equiv \arg \min_{(l_1, \dots, l_K)} S(l_1, \dots, l_K)$.

$$\textcircled{2} \quad ((\hat{b}_{(\hat{l}_0; \hat{l}_1)}, \hat{\beta}_{(\hat{l}_0; \hat{l}_1)}), \dots, (\hat{b}_{(\hat{l}_K; \hat{l}_{K+1})}, \hat{\beta}_{(\hat{l}_K; \hat{l}_{K+1})})) = \arg \min_{(\beta_r, b_r, \dots)} \sum_{r=1}^{K+1} \left[\sum_{i=\hat{l}_{r-1}+1}^{\hat{l}_r} \rho_\tau(Y_i - b_r - \mathbf{X}_i^\top \beta_r) + \lambda_{(\hat{l}_{r-1}; \hat{l}_r)} \hat{\omega}_{(\hat{l}_{r-1}; \hat{l}_r)}^\top | \beta_r | \right]$$

Considérons l'ensemble d'indices des coefficients non nuls pour la phase r , avec $r \in \{1, \dots, K+1\}$:

$$\mathcal{A}_r^0 \equiv \{j \in \{1, \dots, p\}; \beta_{r,j}^0 \neq 0\}.$$

et

$$\widehat{\mathcal{A}}_r^0 \equiv \{j \in \{1, \dots, p\}; \widehat{\beta}_{(l_{r-1}^0, l_r^0), j} \neq 0\},$$

$$\widehat{\mathcal{A}}_r \equiv \{j \in \{1, \dots, p\}; \widehat{\beta}_{(\widehat{l}_{r-1}, \widehat{l}_r), j} \neq 0\}.$$

Propriétés de ces estimateurs

1 $\widehat{l}_r - l_r^0 = O_{\mathbb{P}}(1)$, pour $\forall r = 1, \dots, K$.

2 Propriétés oracle entre 2 change-points estimés:

(i) $(\widehat{l}_r - \widehat{l}_{r-1})^{1/2} \left(\widehat{\beta}_{(\widehat{l}_{r-1}, \widehat{l}_r)} - \beta_r^0 \right)_{\mathcal{A}_r^0} =$

$(l_r^0 - l_{r-1}^0)^{1/2} \left(\widehat{\beta}_{(\widehat{l}_{r-1}, \widehat{l}_r)} - \beta_r^0 \right)_{\mathcal{A}_r^0} (1 + o_{\mathbb{P}}(1))$

$$\xrightarrow[n \rightarrow \infty]{\mathcal{L}} \mathcal{N}(\mathbf{0}_{|\mathcal{A}_r^0|}, \sigma_{d_\tau}^2 \mu_{h_\tau}^{-2} \Omega_{r, \mathcal{A}_r^0}^{-1}).$$

(ii) $\lim_{n \rightarrow \infty} \mathbb{P} \left[\widehat{\mathcal{A}}_r^0 = \widehat{\mathcal{A}}_r = \mathcal{A}_r^0 \right] = 1.$

K inconnu

Estimation de K par un critère (généralisation du critère de Schwarz) consistant:

$$\hat{K}_n \equiv \arg \min_K \left(n \log \frac{S(\hat{l}_1, \dots, \hat{l}_K)}{n} + G(K, p) B_n \right)$$

avec $B_n \rightarrow \infty$ et $G(K, p)$ croissante en K .

Theorem

Si la suite $(B_n)_{n \in \mathbf{N}}$ est telle que $n^{-u} B_n \rightarrow 0$, $n^{-1/2} B_n \rightarrow \infty$, avec $u \in [3/4, 1]$ alors $\hat{K}_n \xrightarrow[n \rightarrow \infty]{\mathbb{P}} K^0$.

(G. Ciuperca, *Statistics*, 2016). (G. Ciuperca, N. Dulac, *JSTP*, 2022)

Simulations

Méthode quantile $K^0 = 2$, $n = 200$, $l_1^0 = 30$, $l_2^0 = 100$, $\beta_1^0 = \beta_2^0 \neq \beta_3^0$;
 $\varepsilon_1, \varepsilon_3 \sim \mathcal{E}_1$, $\varepsilon_2 \sim \mathcal{E}_2$, $\tau = 0.55$.

- $\text{med}(\hat{l}_1 = 31)$, $\text{med}(\hat{l}_2 = 101)$ par Quantile LASSO adaptative;
- $\text{med}(\hat{l}_1 = 25)$, $\text{med}(\hat{l}_2 = 100)$ par LS LASSO adaptative;

Estimation du nombre de changements

$K^0 = 1$, pour 100 répliques Monte Carlo

β_1^0, β_2^0	distribution erreurs	QUANT+aLASSO				LS+aLASSO			
		$\hat{K}_n = 0$	$= 1$	$= 2$	$= 3$	$= 0$	$= 1$	$= 2$	$= 3$
$\beta_1^0 = \beta_2^0$	$\varepsilon_1 \sim \mathcal{E}_1, \varepsilon_2 \sim \mathcal{E}_3$	36	62	1	1	88	10	1	1
	$\varepsilon_1 \sim \mathcal{E}_1, \varepsilon_2 \sim \mathcal{N}(0, 1)$	0	99	1	0	54	66	0	0
$\beta_1^0 \neq \beta_2^0$	$\varepsilon_1 \sim \mathcal{E}_1, \varepsilon_2 \sim \mathcal{E}_3$	0	100	0	0	0	100	0	0
	$\varepsilon_1 \sim \mathcal{E}_1, \varepsilon_2 \sim \mathcal{N}(0, 1)$	0	100	0	0	0	100	0	0

- Si le changement dans le modèle est dû au changement des coefficients, alors les 2 critères choisissent correctement le nombre de changements.
- Si le changement dans le modèle est dû au quantile de ε , alors le critère LS+Lasso adaptatif très mauvais.

Méthode expectile

ε	loss	$\ \hat{\beta} - \beta^0\ _2$	$\% \sum_{r=1}^2 (\mathcal{A}_r^0 \subseteq \hat{\mathcal{A}}_r)$	$\% \sum_{r=1}^2 (\hat{\mathcal{A}}_r \cap \mathcal{A}_r^{0c} \neq \emptyset)$	time (s)
$\varepsilon \sim 0.2 \cdot \mathcal{N}(0, 1) + \chi_1^2$	LS	0.968	100%	2.55%	0.002
	EX	1.920	97.2%	0%	0.003
	QU	0.212	100%	0%	0.003
$\varepsilon \sim \mathcal{E}(1) - 1.5$	LS	1.690	99.0%	21.5%	0.002
	EX	0.792	100%	2.49%	0.002
	QU	1.000	99.9%	5.44%	0.003

Table: Précision des estimations et temps d'exécution pour un modèle avec $K^0 = 1$ change-points, $n = 200$.

ε	Méthode estimation	\hat{K}_n							
		6	7	8	9	10	11	12	13
$\mathcal{N}(0, 1)$	LS	0	0	0	0	0	0	100	0
	QU	0	0	0	100	0	0	0	0
	EX	0	0	0	0	0	0	100	0
$0.2 \cdot \mathcal{N}(0, 1) + \chi_1^2$	LS	0	0	0	0	0	0	100	0
	QU	0	100	0	0	0	0	0	0
	EX	0	0	0	0	1	0	99	0
$3 \cdot \mathcal{E}(1) - 1.5$	LS	0	0	0	0	0	0	100	0
	QU	0	13	21	66	0	0	0	0
	EX	0	0	0	0	0	0	100	0

Table: \hat{K}_n pour 100 Monte Carlo, $n = 1000$, $K^0 = 10$.

2.2. Détection d'un changement en temps réel

Supposons que jusqu'à l'observation m on a un seul modèle:

$$Y_i = \mathbf{X}_i^\top \beta^0 + \varepsilon_i, \quad i = 1, \dots, m$$

A chaque instant, après l'obs m on teste si le modèle change:

$$H_0 : \beta = \beta^0, \quad i = m+1, m+2, \dots$$

$$H_1 : \exists k^0 \geq 1, \text{ tel que, } \beta = \beta^0 \text{ pour } i = m+1, \dots, m+k^0 \text{ et} \\ \beta = \beta_2^0 \neq \beta^0 \text{ pour } i = m+k^0+1, \dots.$$

Estimateur quantile/expectile LASSO adaptatif

$$\hat{\beta}_m \equiv \arg \min_{\beta} \left[\sum_{i=1}^m \rho_{\tau}(Y_i - \mathbf{X}_i^\top \beta) + \lambda_m \hat{\omega}_m^\top |\beta| \right]$$

et les résidus correspondants: $\hat{\varepsilon}_i = Y_i - \mathbf{X}_i^\top \hat{\beta}_m$.

Pour trouver une statistique de test, on considère la somme cumulative (CUSUM) et une fonction de normalisation:

$$g(m, k, \gamma) = m^{1/2} \left(1 + \frac{k}{m} \right) \left(\frac{k}{k+m} \right)^{\gamma}, \quad \gamma \in [0, 1/2).$$

Méthode quantile LASSO adaptative

(G. Ciuperca, *Metrika*, 2018)

Considérons la matrice $\mathbf{J}_{m, \mathcal{A}^0} \equiv m^{-1} \tau (1 - \tau) \sum_{i=1}^m \mathbf{X}_{i, \mathcal{A}^0} \mathbf{X}_{i, \mathcal{A}^0}^t$, la fonction $\phi_\tau(u) = \tau - \mathbf{1}_{u \leq 0}$ et la statistique:

$$\Theta(m, k, \gamma) \equiv \frac{\|\mathbf{J}_{m, \mathcal{A}^0}^{-1/2} \sum_{i=m+1}^{m+k} \mathbf{X}_{i, \mathcal{A}^0} \phi_\tau(\hat{\varepsilon}_i)\|_\infty}{g(m, k, \gamma)}, \quad (5)$$

Pour le nombre d'observations T_m après les observations historiques, deux types de procédure sont possibles:

- *open-end procedure*: $T_m = \infty$ or $(T_m < \infty, \lim_{m \rightarrow \infty} T_m/m = \infty)$;
- *closed-end procedure*: $\lim_{m \rightarrow \infty} T_m/m = T \in (0, \infty)$.

On considère une suite (a_m) telle que $a_m \rightarrow \infty$ et $a_m/m \rightarrow 0$ quand $m \rightarrow \infty$.

Théorème

Sous l'hypothèse H_0 , si la suite $a_m \rightarrow \infty$ telle que $a_m/m \rightarrow 0$ pour $m \rightarrow \infty$, alors, pour tout $\gamma \in [0, 1/2)$:

(i) pour la procédure closed-end on a:

$$\sup_{a_m \leq k \leq T_m} \Theta(m, k, \gamma) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \sup_{0 < t < \frac{T}{1+\gamma}} \frac{\|\mathbf{W}(t)\|_\infty}{t^\gamma},$$

(ii) pour la procédure open-end on a:

$$\sup_{a_m \leq k \leq T_m} \Theta(m, k, \gamma) \xrightarrow[n \rightarrow \infty]{\mathcal{L}} \sup_{0 < t < 1} \frac{\|\mathbf{W}(t)\|_\infty}{t^\gamma},$$

avec $\mathbf{W}(t)$ un processus de Wiener de dimension $|\mathcal{A}^0|$.

- *Idée preuve:* $\mathbb{S}(m, k) = \mathbf{J}_{m, \mathcal{A}^0}^{-1/2} \sum_{i=m+1}^{m+k} \mathbf{X}_{i, \mathcal{A}^0} \phi_\tau(\varepsilon_i) - km^{-1} \mathbf{J}_{m, \mathcal{A}^0}^{-1/2} \sum_{i=1}^m \mathbf{X}_{i, \mathcal{A}^0} \phi_\tau(\varepsilon_i) + \mathcal{O}_{\mathbb{P}}(m^{-1/4} k^{1/2} \sqrt{\log k}) + o_{\mathbb{P}}(km^{-1/2})$.

Théorème

Sous l'hypothèse H_1 , pour la open-end et closed-end procedures:

$$\sup_{1 \leq k \leq T_m} \Theta(m, k, \gamma) \xrightarrow[m \rightarrow \infty]{\mathbb{P}} \infty.$$

Idée preuve: Il existe un k_m tel que $\Theta(m, k_m, \gamma) \xrightarrow[m \rightarrow \infty]{\mathbb{P}} \infty$.

Corollaire. *L'hypothèse H_0 sera rejetée, pour un niveau $\alpha \in (0, 1)$ fixé, dans le point, appelé "temps d'arrêt" (stopping time):*

$$\tilde{k}_m \equiv \begin{cases} \inf\{k \geq 1; \Theta(m, k, \gamma) > c_\alpha(\gamma)\}, \\ \infty, & \text{autrement.} \end{cases}$$

avec $c_\alpha(\gamma)$ le fractile d'ordre $1 - \alpha$ de la distribution de

$$\sup_{0 < t < L(T)} \frac{\|W(t)\|_\infty}{t^\gamma}.$$

Simulations

$m = 300$, $\gamma \in \{0, 0.35, 0.49\}$, $\beta^0 = (0, 0, 1, 0, 2, 0, -1, 0, 0, 0)$.

Après le change-point $k^0 = 100$, le paramètre

$\beta^1 = \beta^0 + (0, 0, 0, 0, \delta, 0, \delta, 0, 0, 0)$.

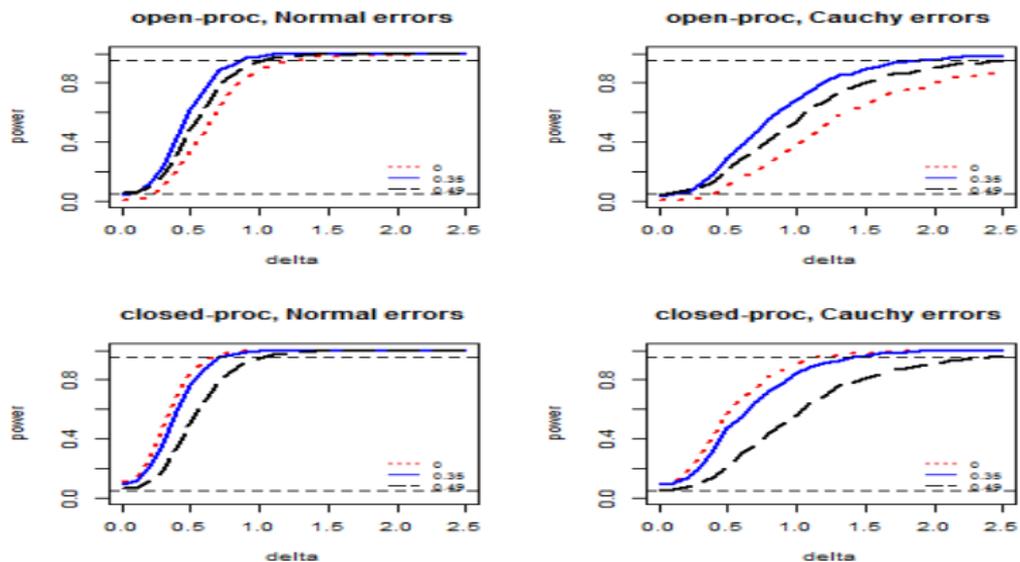


Figure: Etude de la puissance empiriques fonction de δ .

Méthode expectile LASSO adaptative

(G. Ciuperca, *Statistical Papers*, 2022)

Dans ce cas, on prend $\mathbf{J}_{m, \widehat{\mathcal{A}}_m} \equiv \text{Var}[\rho'_\tau(\varepsilon)] \frac{1}{m} \sum_{i=1}^m \mathbf{X}_{i, \widehat{\mathcal{A}}_m} \mathbf{X}_{i, \widehat{\mathcal{A}}_m}^\top$

Theorem

(i) Si l'hypothèse H_0 est vraie, alors:

$$\sup_{1 \leq k \leq T_m} \frac{\|\mathbf{J}_{m, \widehat{\mathcal{A}}_m}^{-1/2} \sum_{i=m+1}^{m+k} \rho'_\tau(\widehat{\varepsilon}_i) \mathbf{X}_{i, \widehat{\mathcal{A}}_m}\|_\infty}{g(m, k, \gamma)}$$

$$\xrightarrow[m \rightarrow \infty]{\mathcal{L}} \sup_{0 < t < L(T)} \frac{\|\mathbf{W}(t)\|_\infty}{t^\gamma},$$

avec $\{\mathbf{W}(t); t \in (0, \infty)\}$ un processus de Wiener de dimension $|\mathcal{A}^0|$, $L(T) = 1$ pour la procédure open-end et $L(T) = T/(1 + T)$ pour la procédure closed-end.

(i) Si l'hypothèse H_1 est vraie, alors la statistique de test converge vers ∞ .

Application: NO2

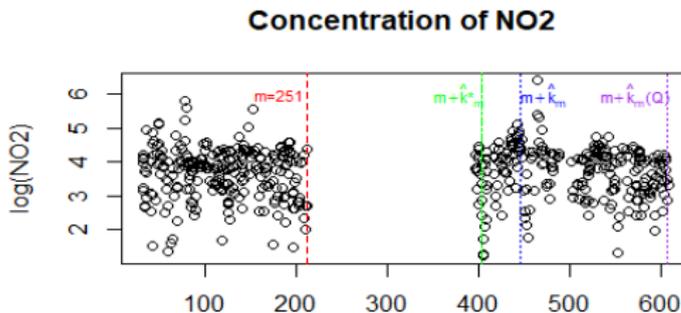
Détection d'un changement dans la pollution de NO2 par rapport à six variables quantitatives (nombre de voitures par heures, température à 2m, vitesse vent, diff temp 25m et 2m, direction vent, heure de max)

Données historiques: 1er novembre 2001 - 30 avril 2002. Données prévues: 1er novembre 2002- août 2003.

Variables rélevantes: nb voitures, diff température 25m et 2m

Table: Détection changement

$Jour[m]$	$m + \hat{k}_m$ Lasso expectile	$m + \hat{k}_m(Q)$ Lasso quantile
30 avril 2002	20 décembre 2002	30 juillet 2003



MERCI de votre attention !