High dimensional change point detection for regression

Malte Londschien

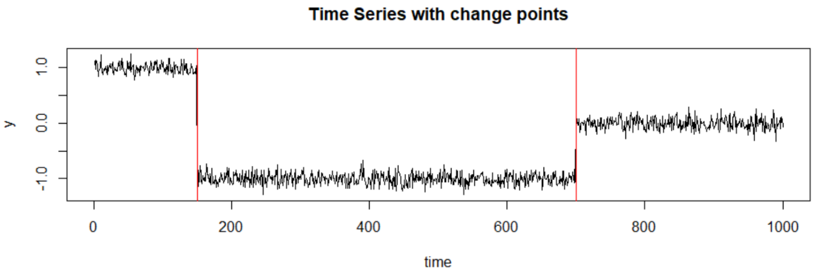
October 30, 2018

General setting, idea

Idea

Our situation is as follows:

- *Our observations* $\{X_i\}_{i=1}^n, \{Y_i\}_{i=1}^n$ *are ordered in time.*
- *We wish to fit some model* $Y_i = f(X_i, i) + \varepsilon_i$ *dependent on time.*
- *The map* $(i \mapsto f(\cdot, i))$ *is piecewise constant.*
- *We wish to estimate the points where the model changes.*



General setting

Given $n, p \in \mathbb{N}$ and random variables $\{X_i\}_{i=1}^n, \{Y_i\}_{i=1}^n$ taking values in $\mathbb{R}^p$ respectively $\mathbb{R}$, assume that there exist $\{\beta(i)\}_{i=1}^n \subseteq \mathbb{R}^p$ and $\{\varepsilon_i\}_{i=1}^n$ centered normal i.i.d. such that $\forall i$:

$$Y_i = X_i^T \beta(i) + \varepsilon_i.$$

Assume furthermore that $(i \mapsto \beta(i))$ is piecewise constant, i.e. that $\exists k \in \mathbb{N}, \alpha \in (\mathbb{N})^{k+2}$ with $1 = \alpha_0 < \alpha_1 < \ldots < \alpha_{k+1} = n + 1$ and $\{\beta^j\}_{j=0}^k \subseteq \mathbb{R}^p$ such that $\forall i$:

$$\beta(i) = \sum_{j=0}^{k} \beta^j 1_{[\alpha_j, \alpha_{j+1})}(i).$$

We wish to estimate α from data.

Introduction of some notation

Notation

For $u < v \in \mathbb{N}$, *let*

$$X_{[u,v)} := \begin{pmatrix} X_{u,1} & \dots & X_{u,p} \\ \vdots & & \vdots \\ X_{(v-1),1} & \dots & X_{(v-1),p} \end{pmatrix}, \quad Y_{[u,v)} := \begin{pmatrix} Y_u \\ \vdots \\ Y_{(v-1)} \end{pmatrix}.$$

Notation

For $1 = \alpha_0 < \alpha_1 < \ldots < \alpha_{k+1} = n + 1$ *let*

$$r_j(\alpha) := \alpha_{j+1} - \alpha_j, \; j = 0, \ldots, k$$

and

$$r(\alpha) := \min_{j=0,\ldots,k} r_j(\alpha).$$

Estimation of α knowing the total amount of change points

Definition

Let $L_n([u,v),\beta) := ||Y_{[u,v)} - X_{[u,v)}^T\beta||_2^2/n$. *Assume we know the total amount of change points* $k \in \mathbb{N}$. *Given a tuning parameter* $\delta \geqslant 0$ *and estimators* $\hat{\beta}([u,v))$ *for* $u < v$, *we can then estimate* α *as*

$$\hat{\alpha} := \underset{\substack{1=\alpha_0<\ldots<\alpha_{(k+1)}=n+1 \\ r(\alpha)\ \geqslant\ \delta n}}{\operatorname{argmin}} \sum_{j=0}^{k} L_n([\alpha_j,\alpha_{j+1}),\hat{\beta}([\alpha_j,\alpha_{j+1})))$$

Example

In a low dimensional setting, we can set

$$\hat{\beta}([u,v)) := \underset{\beta\in\mathbb{R}^p}{\operatorname{argmin}} L_n([u,v),\beta).$$

Estimation of α not knowing the total amount of c.p.

Definition

Let $L_n([u,v),\beta) := ||Y_{[u,v)} - X_{[u,v)}^T\beta||_2^2/n$. *Without knowledge of* k, *given tuning parameters* $\gamma, \delta \geqslant 0$ *and estimators* $\hat{\beta}([u,v))$ *for* $u < v$, *we estimate* α *via*

$$\hat{\alpha} := \underset{k\in\mathbb{N}}{\operatorname{argmin}} \underset{\substack{1=\alpha_0<\ldots<\alpha_{k+1}=n+1 \\ r(\alpha)\geqslant\delta n}}{\operatorname{argmin}} \sum_{j=0}^{k} L_n([\alpha_j,\alpha_{j+1}),\hat{\beta}([\alpha_j,\alpha_{j+1}))+\gamma k.$$

Remark

For unbiased estimators $\hat{\beta}([u,v))$ *and* $\delta = \gamma = 0$, *a minimum is always given by* $\hat{\alpha} = (1, 2, \ldots, n+1)$.

How to solve this optimisation problem?

- Dynamic programming (Computationally very expensive)
- Binary Segmentation: Define the gains function

$$\Delta L_n(t) := L_n([1,n], \hat{\beta}([1,n])) - \Big(L_n([1,t), \hat{\beta}([1,t))) + L_n([t,n], \hat{\beta}[t,n]) \Big)$$

and iteratively split intervals at maxima.

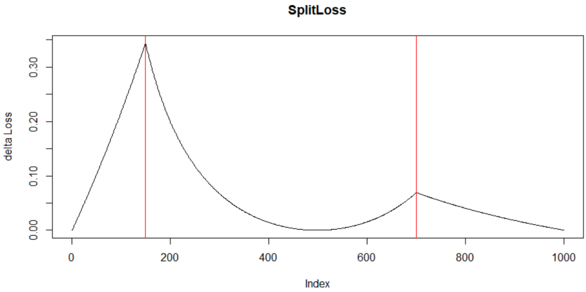


Figure: Example of difference in loss in population case.

Changepoint detection in high dimensions

Again, we wish to estimate

$$\hat{\alpha} := \operatorname*{argmin}_{k \in \mathbb{N}} \operatorname*{argmin}_{\substack{1=\alpha_0<\ldots<\alpha_{k+1}=n+1 \\ r(\alpha)\geqslant\delta n}} \sum_{j=0}^{k} L_n([\alpha_j, \alpha_{j+1}), \hat{\beta}([\alpha_j, \alpha_{j+1}))) + \gamma k.$$

In high dimensions, we use estimators

$$\hat{\beta}([u, v)) := \operatorname*{argmin}_{\beta \in \mathbb{R}^p} L_n([u, v), \beta) + \lambda_{[u,v)} \|\beta\|_1.$$

How to choose the $\lambda_{[u,v)}$? Remember that we have consistency for $\lambda \asymp \sqrt{log(p)/n}$. We thus choose $\lambda_{[u,v)} = \sqrt{v-u}\,\lambda_0$ for some λ_0.

Applications in Finance

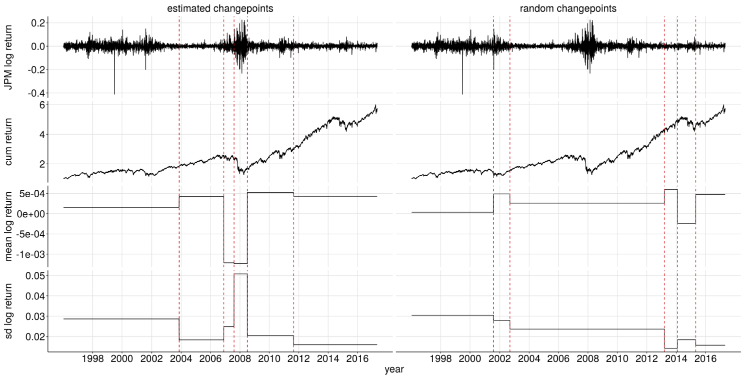


Figure: Estimated changepoints (left) versus randomly sampled changepoints (right) for the log returns of a subset of the S&P 500 stocks. From 'Optimistic Binary Segmentation' by Lorenz Haubner.