Multiple change points detection in linear regression by Filtered Derivative and False Discovery Rate method

ELMI MOHAMED*
University of Djibouti, Djibouti, Republic of Djibouti  mohamed.elmi.abdillahi@univ.edu.dj

Abstract

Linear models are widely used in statistics to describe between two variables : X the explanatory variable and Y the response variable. In this work, we consider the simple linear regression model with change on the parameters slope and intercept. We use the Filtered Derivative and False Discovery Rate method (FDqV) for estimating the coefficients of linear regression. We indicate that it exists a previous work made for estimating the parameter slope by using the Filtered Derivative with p-value (FDpV). We specify in this work, that the estimating of the slope done by FDpV is wrong and we give the correct estimation.

Keywords: Linear regression, Filtered Derivative and False discovery rate, Time Series, Gaussian Derivative.

1. Introduction

We study in this paper the problem of change point analysis in the case of simple linear regression. For an updated overview, the reader can see the book (Basseville (1993)) or the article (Bai & Perron (1998)). The goal is to detect or to estimate the instant of abrupt changes and the parameter (slope and intercept) corresponding of linear regression. Others methods for the problem of change points exist such that the Penalized Least Square Error (PLS), the Hierarchical Binary Splitting (HBS), but the both-methods are expensive in times of calculation. The PLS- method has the time and memory complexity of order \(O(N^2)\). Recently, an other method called the filtered derivative with p-value (FDpV) (Bertrand, & Fhima, & Gullin (2011)) is introduced for the detection of change point analysis, this method for the detection of slope in linear regression is wrong, because the filtered derivative considered in (Bertrand, & Fhima, & Gullin (2011)) is a hat-function which presents an each instant of change point a maximum and this is not the case. We provide in this work the exactly order to detect of slope parameter in linear regression. In our method, the time and memory complexity are the order \(O(N)\). By consequent, the (FDqV)-method is more advantageous not only the criteria (time and memory complexity) but also the FDqV-method realize the best results for the detection of abrupt changes and so the estimation of corresponding parameters ( for example the parameter mean) if we base ourselves on the criteria ( Mean Integrate Square Error (MISE), Number of False Alarms (NFA) and Number of No-Detection (NND)). The dominance of FDqV than FDpV is well treated in (Elmi (2014)) or (Elmi (2014)). The rest of this paper is structured as following : The section 1 describes the description of change points in simple linear regression, the section 2 recall back the FDqV-method for the mean. The section 3 apply the FDqV-method in the coefficients of linear regression.

2. Description of the problem

In this section, we describe the problem of change analysis and we study the change in the coefficients of linear regression. In fact, there are two types change points detection in linear regression : The model with a discontinuous change points and the model with a continuous change points.

Model discontinuous for change points in linear regression.

The model with a discontinuous change points is defined as :

Let \((X_i, Y_i), i = 1, 2, \ldots, n\), the observations where each \(Y_i\) describes the response of the explanatory variable \(X_i\). A simple linear regression with no change is defined as \(Y_t = aX_t + b + \epsilon_t\) for \(t = 1, 2, \ldots, n\), and \(a\) and \(b\) are the slope and the intercept of linear regression. Here, we suppose that the parameters \(a\) and \(b\) change. Then , we have:

\[X = (X_1, X_2, \ldots, X_n)\] is a family of independent random variables indexed by the time \(t=1, 2, \ldots, n\). It exists a segmentation \(\tau = (\tau_1, \tau_2, \ldots, \tau_K)\) with \(\tau_k \in [1, 2, \ldots, n]\) and \(0 < \tau_1 < \tau_2 < \ldots < \tau_K < n\). \(K\) denotes the number of changes.
By convention, we set $\tau_0 = 1$ and $\tau_{K+1} = n$, thus $K$ can be equal to zero (this means no change) or any integer smaller than $n$. Thus

$$Y_t = \begin{cases} a_o \times X_t + b_o + \epsilon_t, \text{ for } t = 1, \ldots, \tau_1 \\ a_1 \times X_t + b_1 + \epsilon_t, \text{ for } t = \tau_1 + 1, \ldots, \tau_2 \\ \vdots \\ a_K \times X_t + b_K + \epsilon_t, \text{ for } t = \tau_K + 1, \ldots, n \end{cases}$$

where $\epsilon_t \sim N(0, \sigma^2)$ denote the Gaussian law with mean 0 and standard deviation 1. For simplicity of the presentation, we have assumed that the variance $\sigma^2$ remains constant.

**Model continuous for change points in linear regression.**

The model continuous for change points in linear regression is defined as the discontinuous case but an each change point $\tau_k$, for $t=1, \ldots, \tau_k$ there are a continuity constraint. For more details, we have:

$$Y_t = \begin{cases} a_o \times X_t + b_o + \epsilon_t, \text{ for } t = 1, \ldots, \tau_1 \\ a_1 \times X_t + b_1 + \epsilon_t, \text{ for } t = \tau_1 + 1, \ldots, \tau_2 \\ \vdots \\ a_K \times X_t + b_K + \epsilon_t, \text{ for } t = \tau_K + 1, \ldots, n \end{cases}$$

With

$$\begin{cases} a_o \times X_t + b_o + \epsilon_t = a_1 \times X_t + b_1 + \epsilon_t, \text{ for } (\tau_0, Y_{\tau_1}). \\ a_1 \times X_t + b_1 + \epsilon_t = a_2 \times X_t + b_2 + \epsilon_t, \text{ for } (\tau_1, Y_{\tau_2}). \\ \vdots \\ a_K \times X_t + b_K + \epsilon_t = a_n \times X_t + b_n + \epsilon_t, \text{ for } (\tau_K, Y_{\tau_n}). \end{cases}$$

For an illustration, we draw the following figures.

![Figure 1: First drawing: the linear regression with model discontinuous change points. Second drawing: the linear regression with model continuous change points](image)

3. **Recall the Filtered Derivative and False Discovery Rate Method (FDqV)**
The Filtered Derivative and False Discovery Rate (FDqV) (Elmi(2014)) is a method derived the Filtered Derivative. For reading of this section, see the article (Elmi(2014)).

4. Detection the parameters of linear regression by FDqV-method

Our goal is to detect the abrupt changes and the estimation the both parameters slope and intercept an each box.

Simulation : change on the slope

For $n = 5,000$, we have simulated a simple linear regression of the random variables $(X_t, Y_t), t = 1, 2, \ldots, n$ with four change points $\tau = (1000, 1500, 3000, 4500)$ with the corresponding slopes $a_t = (0.2, 0.8, -0.5, 0.3, -0.1)$ (here, the intercept remains constant).

The FDqV-method

We want to estimate the instant $(\tau_1, \tau_2, \ldots, \tau_K)$ and $a_k$ an each box $[\tau_k + 1; \tau_{k+1}]$ for $1 < k < K$.

Step 1: The Filtered Derivative (FD) for the slope

The FD for the slope is defined as:

$$FD(A, k) = \tilde{a}(k, A) - \tilde{a}(k - A, A)$$

where

$$\tilde{a}(k, A) = [A \times \sum_{t=k+1}^{k+A} X_t Y_t - \sum_{t=k+1}^{k+A} X_t \sum_{t=k+1}^{k+A} Y_t] [A \times \sum_{t=k+1}^{k+A} X_t^2 - (\sum_{t=k+1}^{k+A} X_t)^2]^{-1} \quad (2)$$

is the estimator of the slope on the box $[k+1,k+A]$ obtained by the least square method.

We give below a lemma which allow us to well define the FD function.

Lemma.

Let $Y_j = a_k \times X_j + b_k + e_j$ for $\tau_k + 1 - 1 \leq k \leq \tau_k, 1 \leq j < n$ and $e_j$ is the error of Gaussian law of mean zero and standard deviation $\sigma$. Then

$$\tilde{a}_k \sim N(a_k, \sigma^2_{\tilde{a}_k})$$

$$\frac{\sigma^2_{\tilde{a}_k}}{\sigma^2} = \left[ \sum_{i=\tau_k+1}^{\tau_k} (X_j - \bar{X}_k)^2 \right]^{-1}$$

Proof.

The proof is clear by using the hypothesis that $e_j$ is a Gaussian law.

By definition, the Filtered Derivative function (FD) is mathematically defined in terms of difference between the estimators of the slope computed in two sliding windows respectively at the right and at the left of the time $k$, both of size $A$. We can re-define the FD using the lemma as the difference between two Gaussian functions respectively at the right and the left of the time $k$, both of size $A$. Then we have exactly the definition of the Derivative Gaussian.

How to detect the potential change points.

In the figure 2, we clearly see that the Gaussian Derivative (GD) change sign on each box $[\tau_k - A, \tau_k + A]$ for $k = 1, \ldots, K$, so for the localization of abrupt changes, we use the following algorithm.

Algorithm.

We choose a threshold $C_1$ and $K_{max}$ corresponding the number of maximum of change points.

Step 1. We calculate the maximum of GD function and the argument of maximum and we set for $k = 1$,

$$\tau_1 = \frac{\arg \max_{\epsilon \in [A, n-A]} GD(\epsilon) + \arg \min_{\epsilon \in [A, n-A]} GD(\epsilon)}{2}$$
and on \([\tau_1 - A; \tau_1 + A]\) we put \(GD(k) = 0\).

Step 2. While \((k < K_{max})\) and \((C_{max} > C_1)\) do

\[
\hat{\tau}_k = \frac{\arg \max_{k \in [A;\tau_1 - A]} GD(k) + \arg \min_{k \in [\tau_1 + A; A]} GD(k)}{2}
\]

and on \([\tau_k - A; \tau_k + A]\), we set \(GD(k) = 0\)

Step 3. We sort out in order increasing the instant of potential change points.
Thus, we keep the instant of potential change points \(\tau_1 < \tau_2 < \ldots < \tau_K\) with \(K < \hat{K}\).
Below, we draw the Gaussian Derivative with noise and without noise.

![Graphs showing linear regression and filtered derivatives](image)

Figure 2: First drawing: The observed signal, the third drawing: The Filtered derivative function for the slope without noise, the fourth drawing: The Filtered Derivative with noise.

**Remark**

We specify that the manner done in (Bertrand, Fhima, & Gullin (2011)) for detection of change points in linear regression for the parameter slope is wrong. In fact, they use in step 1, a hat-function for the localization of change points. We realize in this work that the filtered derivative function is a GD and not a hat-function, see figure 2.

**Step 2: The False Discovery Rate**

At the end of step 1, we have the instant of potential change points. Among these points, there are the right points and also the false detections. The false discovery rate allow us to separate these two kind points and keep the false detections at level close to zero. For this, we proceed in the same way as for the parameter mean describes in the section 2. Thus, we obtain the estimated instants \((\tau_1^* < \tau_2^* < \ldots < \tau_{K^*}^*)\) with \(K \leq K^*\). Also, each time detected, we can estimate the corresponding slopes \(a_k\), for \(k = 1; 2; \ldots; K^*\).

**The FDqV-method for the intercept**

In subsection, we want to apply the FDqV-method, the detection of abrupt changes and the estimation of the intercept on each box. The FD for the intercept is:

\[
FD(A, k) = \hat{b}(k, A) - \hat{b}(k - A, A)
\]
Where
\[ \hat{b}(k, A) = \frac{1}{A} \sum_{j=k+1}^{k+A} X_j - a \times \frac{1}{A} \sum_{k+1}^{k+A} X_j \]

**Simulation: Change on the intercept**

For \( n = 5,000 \), we have simulated a simple linear regression of the random variables \((X_t, Y_t), \ t = 1, 2, \ldots, n\) with four change points \( \tau = (1000, 1500, 3000, 4500) \) on the intercept \( b_t = (100, 500, 2000, -1000, 1500) \) (here, the slope remains constant). Below, we have the corresponding figure of this simulation and the filtered derivative.

![Simulation figure](image-url)

**Figure 3:** First drawing : The observed signal, second drawing: The Filtered Derivative function for the intercept.

We see through this above figure that the filtered derivative is a hat-function, so in the first time we select the first instant \( \hat{\tau}_1 \) of abrupt changes as the maximum value of the filtered derivative such that \( |FD| > C_1 \), where \( C_1 \) is the threshold chosen. We put around of the point \( \hat{\tau}_1 \), \( FD[\hat{\tau}_1 - A, \hat{\tau}_1 + A] = 0 \), where \( A \) is the window size. We begin again the same procedure for a new function FD, Thus we keep the second estimated point \( \hat{\tau}_2 \). We apply the above procedure, while \((k < K_{max})\), where the \( K_{max} \) is the maximum number change points. Finally, we have the estimated instants \((\hat{\tau}_1; \hat{\tau}_2; \ldots; \hat{\tau}_{K^*})\). For more details see (Elmi(2014)). At the step 2, For eliminating the false alarms or having the number of false alarms at level close to zero, we did \( \hat{K} \) hypothesis where the null hypothesis \( H_{(0,k)} : \hat{b}_k = \hat{b}_{k+1} \) versus the alternative hypothesis \( H_{(1,k)} : \hat{b}_k \neq \hat{b}_{k+1} \) with \( k = 1; 2; \ldots; \hat{K} \). When, we did all hypothesis tests, we have the p-value \( \hat{p}_1; \hat{p}_2; \ldots; \hat{p}_{\hat{K}} \). In the sequel, we apply the Benjamini and Hochberg’s procedure (Benjamini & Hochberg(2014)), so we only keep the p-values \( (p_1; p_2; \ldots; p_{K^*}) \), with \( K \leq K^* \). At the end of this procedure, we have the estimated instants and the corresponding intercept \( b_k \) for \( k = (1; 2; \ldots; K^*) \). For more details see (Elmi(2014)) or (Elmi(2014)).

**5. Conclusions**

In this work, we have given the estimating of the coefficients of linear regression. We corrected the mistake done in (Bertrand,& Fhima,& Gullin (2011)) concerning the estimating of the slope. It is logical to do the comparison of existing methods in the literature such the Penalized Square Error or the Hierarchic Binary Split. We already affirm that the FDqV-method is advantageous on times and memory complexity see (Elmi(2014)) or (Elmi(2014)). Finally I will do a real application of this method in the case of linear regression. We have already done a real application of this method concerning the wind turbines for the parameter mean (Elmi(2014)).

**References**


