*Chilean Journal of Statistics*
Vol. 9, No. 1, April 2018, 19-32

# On the estimation of change points in the Beer-Lambert law problem

Michel Helcias Montoril*, and Clécio da Silva Ferreira

Department of Statistics, Federal University of Juiz de Fora, Juiz de Fora, MG, Brasil

(Received: May 15, 2017 · Accepted in final form: March 9, 2018)

**Abstract**

In this paper, we propose a simple method, based on the coefficient of determination, to estimate the change points in the Beer-Lambert law problems. We run Monte Carlo simulation studies in different scenarios in order to evaluate the performance of the method. Furthermore, we illustrate the approach with real datasets where we estimate the appropriate change points that yield the interval where the compound concentration and absorbance have linear relationship. In both cases, simulation and application, it is possible to see good performances of the method, that can be easily implemented.

**Keywords:** Beer-Lambert law · Change point · Coefficient of determination.

## 1. Introduction

This paper is motivated by an Analytical Chemistry problem, where the Beer-Lambert law dictates the relation between absorbance and a compound concentration, which can be described as

$$A = \beta C,$$

where $A$ corresponds to the absorbance, $C$ the concentration and $\beta = ab$, with $a$ being the absorptivity and $b$ the optical path. Flame atomic absorption spectrometry (FAAS) is an analytical technique widely used to measure this relation. In FAAS, the value of $b$ corresponds to the flame width, which is defined according to the spectrometer model (see, e.g., Skoog et al., 2013). Some applications of the Beer-Lambert law can be seen, for example, in Rodríguez-Marín et al. (2012), Dias et al. (2015) and Içelli et al. (2014).

It is well known that there is an interval where the Beer-Lambert law holds. Therefore, the goal is to find the minimum and maximum concentrations that provide the range where the analytical response (absorbance) is linearly related to the compound concentration.

A typical experiment can be observed in Figure 1. One can see that the relationship between zinc concentration and absorbance is almost linear for concentrations between around 0.1 and 1.7. Furthermore, it seems that the slope of a linear relationship changes around 0.3. Thus, there is an interval where the pair zinc concentration and absorbance holds due to the Beer-Lambert law.

*Corresponding author. Email: michelcias@gmail.com

ISSN: 0718-7912 (print)/ISSN: 0718-7920 (online)

http://www.soche.cl/chjs

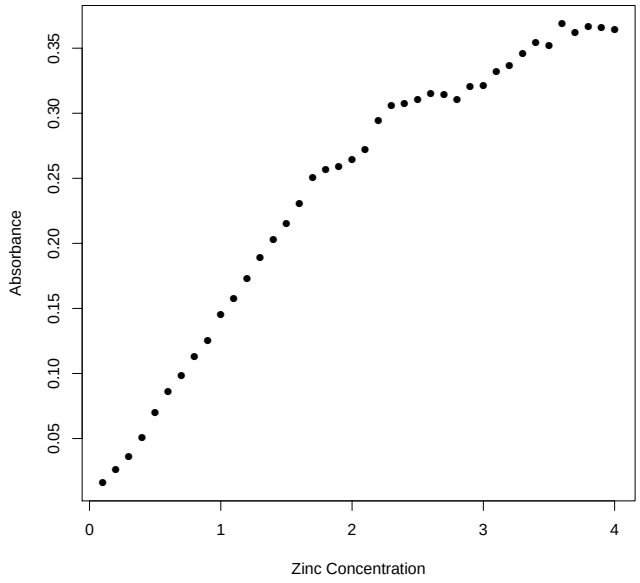


Figure 1. Scatter plot of absorbance versus zinc concentration.

In practice, one usually takes predetermined intervals depending on the chemical element, where the Beer-Lambert law must hold, or one defines the interval simply by observing the scatter plot between compound concentration and absorbance. However, the amplitude and location of the interval where the Beer-Lambert law holds may vary according to environmental conditions and even the device that is being used. Thus, predetermined intervals will likely be inappropriate. Although it may be possible to make a decision by graphical analysis, sometimes subjective choices may be a problem, and an automatic method based on the data would be of substantial interest, avoiding wasting material and time.

The choice of the interval where the Beer-Lambert law holds is clearly a typical change point situation. The change point problem has been widely studied for a long time. Pettitt (1979) introduced nonparametric techniques for the change point problem with binomial and continuous observations. Barry and Hartigan (1993) proposed a Bayesian model for the change point detection – when there is an unknown partition of a set into contiguous blocks where the observations are independent with normal distribution. Erdman and Emerson (2008) implemented this method in R Core Team (2017), allowed in the `bcp` package. Liu et al. (2013) proposed a change point detection algorithm based on nonparametric divergence estimation between time-series samples from two retrospective segments. In linear regression models, we can highlight Quandt (1958), Quandt (1960), Worsley (1983), Kim and Siegmund (1989), Chen (1998), Pastor and Guallar (1998), Julious (2001), Osorio and Galea (2006) and Chen et al. (2011) just to mention few. A comprehensive survey of this topic can be found in Chen and Gupta (2012).

On the other hand, our experience is that practitioners, despite their good performance, seldom accept sophisticated statistical techniques. This by itself can be interpreted as a challenge, which is in line with the classic saying that states "*Research is to realize the obvious*". This is used as a motivation in this paper. Furthermore, although there is a wide range of methods for change point detection, to the best of our acknowledge, there is not any approach in the literature that is capable of detecting the interval where the Beer-Lambert law holds.

In this paper, we propose a simple method to estimate the change point in partially linear models. Although we focus on the Beer-Lambert law, this method may also be used in other situations because of the interesting numerical results presented in the next sections.

This paper is organized as follows. In Section 2, we present the model, propose a simple method for efficiently estimating change points in partially linear models and discuss some features of our approach. In Section 3, we describe some simulation results. Typical applications, based on the Beer-Lambert law, illustrate the method in Section 4. Some conclusions and additional remarks are presented in Section 5.

## 2. Partially linear models with change point

Most of the studies to estimate change points in regression models consider that these points are limits for variations of the intercept or the slope (or both) of linear models. In this paper, we consider that part of the data (an interval) is linear, and outside the interval the relationship changes while not necessarily linear it presents a regression function that may assume any kind of functional behavior.

### 2.1 The partially linear model

In this paper, we assume that the relationship between the dependent and the independent variable is partially linear satisfying

$$Y_i = \begin{cases} g(x_i) + \varepsilon_i, & x_i < c_0 \\ \beta_0 + \beta_1 x_i + \epsilon_i, & c_0 \leq x_i \leq c_1 \\ f(x_i) + \varepsilon_i, & x_i > c_1, \end{cases} \tag{1}$$

where $Y_i$ denotes the response (dependent) variable in the experiment; $x_1 < x_2 < \cdots < x_n$ are values of the explanatory (independent) variable; $\boldsymbol{\beta} = (\beta_0, \beta_1)^\top$ is a $2 \times 1$ fixed parameter vector that defines the linear relationship between $Y$ and $x$, for $c_0 \leq x \leq c_1$; and $g$ and $f$ are arbitrary univariate (regression) functions that quantify the effect of the explanatory variable $x_i$ on the response $Y_i$, for $x_i < c_0$ and $x_i > c_1$, respectively. We also assume that the errors $\epsilon_i$ are independent and identically distributed (iid) random variables with Normal distribution, whose mean and standard deviation are zero and $\sigma$, respectively. For the error term $\varepsilon_i$, we suppose it has zero mean and standard deviation $\gamma$, not necessarily following a Normal distribution.

Observe that the error terms $\epsilon$ and $\varepsilon$ do not need to have the same distribution. Furthermore, since our method focus only on the linear term, the regression functions $g$ and $f$ may be different. This allows the model to be more flexible, with a wider range of different (possible) features. We approach different scenarios (including the heteroscedastic case) in the Section 3.

### 2.2 Estimation of the change point

This simple procedure is based on the coefficient of determination, which is well known in the literature, used in many different fields. Since we are considering the relationship between two variables (in a simple linear regression model), the coefficient of determination corresponds to the square of the coefficient of correlation, which is implemented in a wide range of computational packages.

The main idea behind the use of the coefficient of determination to detect change points relies on the consistency of this estimator. In other words, in a (simple) linear regression case, the estimator of the coefficient of determination converges almost surely to its population coefficient (see, e.g., Theorems 2.3.4 and 2.3.13 in Sen and Singer, 2000).

The arguments above mean that the more one adds pairs $(x, Y)$ to the dataset, the closer the estimator of the coefficient of determination tends to be to the population coefficient. Thus, as long as the (linear) relationship between two variables does not change, the estimates of the coefficient of determination tend to increase (although not monotonically) to the population value.

For the sake of clarity, we develop by separating the problem into three cases, from the linear situation to the general model (1).

### First case: $c_0 = x_1$ and $c_1 = x_n$

Consider initially that $c_0 = x_1$ and $c_1 = x_n$, i.e., the relationship between $(x, Y)$ is completely linear. Let us denote

$$R^2_{(1,k)} = \frac{\left(\sum_{i=1}^{k}(x_i - \bar{x})Y_i\right)^2}{\sum_{i=1}^{k}(x_i - \bar{x})^2 \sum_{i=1}^{k}(Y_i - \bar{Y})^2},$$

the estimator of the coefficient of determination based on the sample from the 1st to $k$-th pair. An illustration of how $R^2_{(1,k)}$ tends to converge as $k$ increases can be observed in a simulated example presented in Figure 2. Each gray point corresponds to the coefficient of determination $R^2_{1,k}$ estimated with respect to the data, where $k = 20, 21, \ldots, 100$, and the horizontal line corresponds to the true (population) coefficient of determination. In this example, we used a sample size $n = 100$.

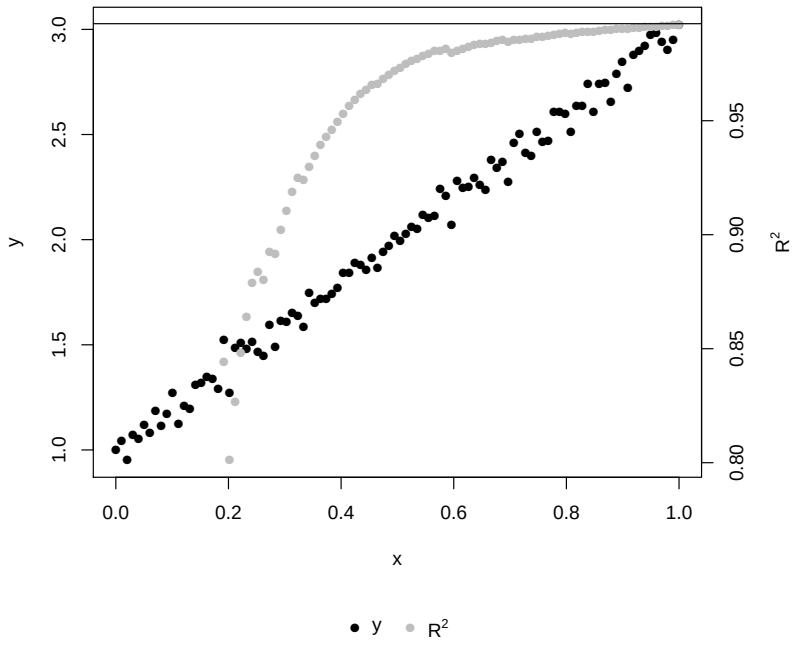


Figure 2. Typical simulated example that shows how the coefficient of determination increases as we increase the amount of data in the estimation.

### Second case: $c_0 = x_1$ and $c_1 < x_n$

Now, consider the case where $c_0 = x_1$ and $c_1 < x_n$. Thus, the first part of the data has a linear relationship and it changes after some point $c_1$. Since the estimates of $R^2_{(1,k)}$ tends to increase as long as the data is linear, it makes sense to think that, if the coefficient of determination starts to decrease as $k$ increases, then the relationship is no longer linear. Therefore, a change point estimator for $c_1$ corresponds to the value $x_k$ with index $k$ that maximizes $R^2_{(1,k)}$, i.e.,

$$\hat{c}_1 = x_{k'}, \quad \text{where} \quad k' = \arg\max_k R^2_{(1,k)}.$$

### Third case: $c_0 > x_1$ and $c_1 < x_n$

This is the most general case. The data is assumed to have all the three components (nonlinear-linear-nonlinear relationships) presented in (1). Thus, we must adapt $R^2_{(1,k)}$ to a more general scenario, where we consider the estimator $R^2_{(j,k)}$, which is based only on the data from the $j$-th to the $k$-th pair. For each different value of $j$, we set $c_0 = x_j$ and estimate $c_1$ as in the second case above. Thus, for each $j$ we will have an adequate pair denoted by $(\hat{c}_1^{(j)}, R^2_{(j,k^{(j)})})$, which corresponds to the estimated value of $c_1$ and the maximum coefficient of determination estimate obtained, based on the sample from the $j$-th pair to the $k^{(j)}$-th pair. Observe that we use the superscript $(j)$ for the $k$ value selected, because it may vary for different values of $j$. Since the greater the $R^2$ the better the fitting is, the appropriate value of $j$ will be the one that maximizes $R^2_{(j,k^{(j)})}$, which can provide the estimator of $c_0$. Therefore, it is easy to see that the estimator of the pair $(c_0, c_1)$ corresponds to

$$(\hat{c}_0, \hat{c}_1) = (x_{j'}, x_{k'}), \quad \text{where} \quad (j', k') = \arg\max_{j,k} R^2_{(j,k)}. \tag{2}$$

Besides the third case above can be easily implemented, it will work properly as long as the data have a linear relationship for some interval. In other words, it is important that we have a minimum interval where the dependence between dependent and independent variables is linear, which can be even the full range of the independent variable. This means that our approach is able to detect not only the scenario of model (1), but also its variations. For example, the situation of the second case above (see the second dataset application, in Section 4).

## 2.3 On the candidate indexes

With respect to the indexes $(j, k)$ to be used during the estimation, we highlight the importance of being careful on the choice of the grid that must be used to estimate $(c_0, c_1)$ in (2). Observe that the estimate $R^2_{j,k}$ is based on a sample with size $k - j + 1$. Thus, depending on the choice of $k$ and $j$, the sample size may be too small and, then, $R^2_{j,k}$ may not be considered a reliable estimate. Although it is not reported here, we have observed that our method works well with grids of the form $I_{i_0,i_1} = \{1, \ldots, i_0\} \times \{i_1, \ldots, n\}$, where $i_0$ and $i_1$ are such that $i_1 - i_0 \geq 5$.

An advantage of considering grids as above is that it allows our method to work with variations of (2) (e.g., the second case above). Therefore, if for some reason the researcher has knowledge about the value of any change point, this restriction can be imposed in the grid of candidates. For example, the second case above will be used if we consider the grid of candidates $I_{1,i_1} = \{1\} \times \{i_1, \ldots, n\}$.

### 2.4 Decision rule to test full linearity

Our proposal does not need that the nonlinear terms in model (1) exist (as discussed above). This means that there is no requirement about the sample size of the nonlinear terms, which can be seen as an advantage, if compared to other change point methods proposed in the literature (see, e.g., Chen, 1998; Osorio and Galea, 2006; Chen et al., 2011). Thus, we can think on testing the hypothesis that the data have only linear relationship, against an alternative hypothesis which states that there is also an interval where the regression function is nonlinear.

The hypothesis to be tested could be written as

$$H_0 : c_0 = x_1 \text{ and } c_1 = x_n$$
$$H_1 : c_0 > x_1 \text{ and/or } c_1 < x_n.$$

The null hypothesis $H_0$ states that the relationship between both variables is totally linear, while the alternative $H_1$ says that there is also a nonlinear relationship between the data, which can be before and/or after the linear term.

The null hypothesis can be rejected based on the decision rule $R^2_{(1,n)} < \max\{R^2_{(j,k)} : (j,k) \in I_{i_0,i_1}\}$. This procedure is clearly similar to the Schwarz Information Criterion (SIC), that has been used in several papers related to change point selection in partially linear models (PLM) (see, e.g., Chen, 1998; Chen and Gupta, 2001; Osorio and Galea, 2006).

It is important to mention that the model (1) and PLM considered in the afore mentioned papers are different. In the PLM used in these papers, it is assumed that there are deterministic relationships in subintervals separated by the change points. In this work, since we focus specially on cases similar to the Beer-Lambert law, out main interest is on the interval where the relationship between variables is linear, and we do not need to worry about the functional behavior of the regression function outside the linear interval. This can be seen as a powerful feature in our approach, because the regression function does not need to have a previously specified behavior that is not linear.

## 3. Simulation study

In order to evaluate the performance of the approach in Section 2.2, we run Monte Carlo simulations using different sample sizes and different situations. We approach both, the second and third cases in Section 2.2.

Based initially on the second case, where $c_0 = x_1$, we consider four different partially linear models, which are enumerated and presented in Table 1. It is important to mention that, in model 4, $\Phi$ corresponds to the cumulative distribution function of a standard Normal distribution. For the sake of simplicity, we used as covariate a sequence of equally spaced points between zero and one, with a change point in the middle, i.e., $c_1 = 0.5$. Moreover, in order to make sure that the two parts of the model (the linear and nonlinear terms) have the same amount of data, we chose odd numbered sample sizes. Therefore, we took 1,000 replicas for each model and for each sample size $n = 51, 101, 251, 501$.

Considering a sample with size $n = 51$, we illustrate typical examples with their regression functions in Figure 3. With respect to the indexes $(j,k)$ to be used in $R^2_{(j,k)}$, we fixed $j = 1$ to ensure that $c_0 = x_1 = 0$, and we tested each value of $k$ such that $x_k \geq 0.4$.

We compute the average (and standard deviation) of the change point estimates based on the replicas. The results are presented in Table 2, where it can be observed that the method tends to provide estimates very close to the real change point, including for data

Table 1. Models used for detecting the change point in the Monte Carlo simulations.

| Model | Linear term | $\sigma$ | $f(x)$ | $\gamma$ |
|---|---|---|---|---|
| 1 | $x$ | 0.015 | $x$ | 0.085 |
| 2 | $1+x$ | 0.030 | $-1.5+6x$ | 0.090 |
| 3 | $2x$ | 0.030 | $1+\cos(2\pi(x-0.75))$ | 0.060 |
| 4 | $-1.49+3.99x$ | 0.015 | $(1-1/(16x)\sin(16\pi(x-0.5)))\Phi\left(\frac{x-0.5}{0.1}\right)$ | 0.030 |

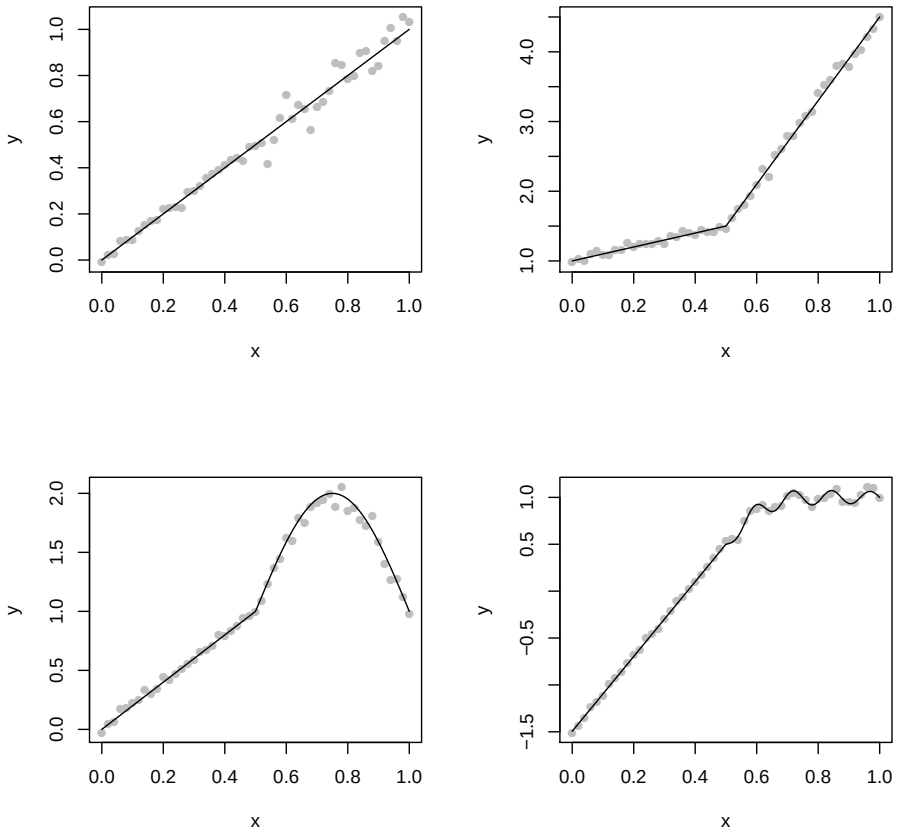


Figure 3. Examples of each model used in the Monte Carlo simulations with sample size $n = 51$. In the top, datasets from models 1 (letf) and 2 (right); in the bottom, datasets from models 3 (left) and 4 (right).

sets with the smallest sample size $n = 51$. It is also important to mention the ability of the method to successfully estimate the change point even in the heteroscedastic case (model 1), which highlights the efficiency of the approach. Furthermore, in Figure 4, it is also possible to observe good indicatives that the estimates are converging on the real change point $c_1 = 0.5$ as $n$ increases.

Table 2. Average (standard deviation) of the change point estimates of $c_1$ for the models presented in Table 1.

| Model | $n$ | | | |
|---|---|---|---|---|
| | 51 | 101 | 251 | 501 |
| 1 | 0.4835 (0.0315) | 0.4912 (0.0149) | 0.4970 (0.0049) | 0.4985 (0.0023) |
| 2 | 0.4883 (0.0240) | 0.4961 (0.0158) | 0.5003 (0.0079) | 0.5007 (0.0048) |
| 3 | 0.4933 (0.0253) | 0.5016 (0.0170) | 0.5050 (0.0092) | 0.5048 (0.0068) |
| 4 | 0.4917 (0.0268) | 0.4993 (0.0142) | 0.5021 (0.0067) | 0.5028 (0.0047) |

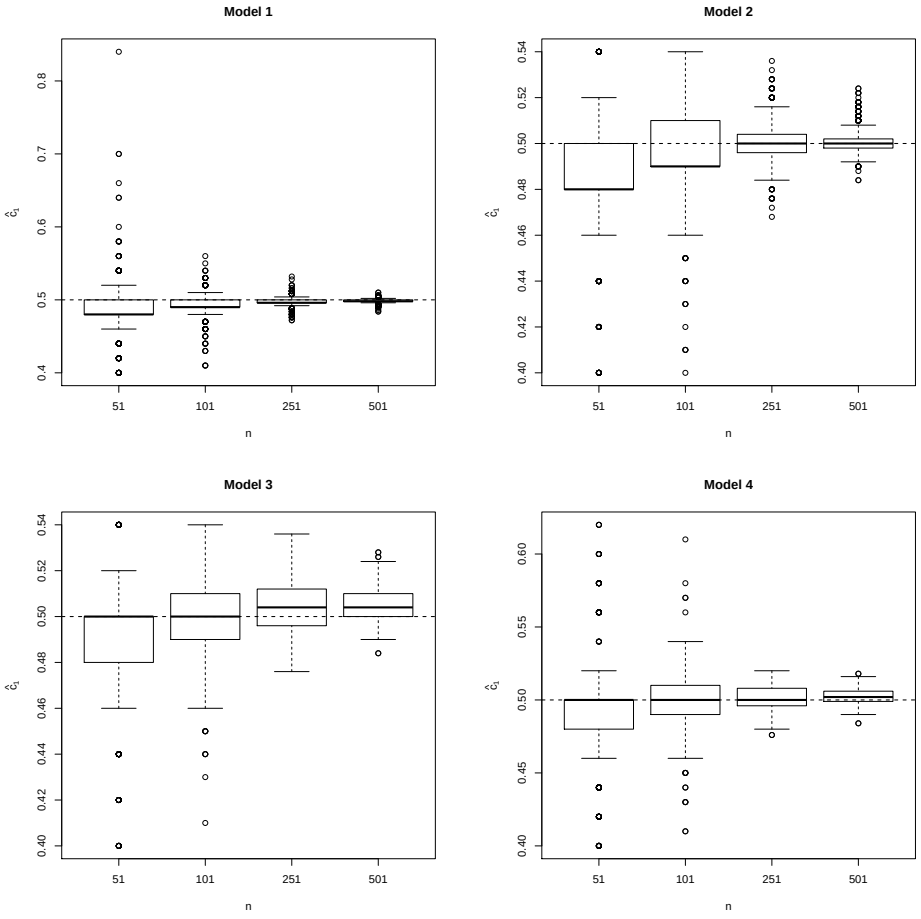


Figure 4. Boxplots of the change point estimates of $c_1$ for the models presented in Table 1. The dashed lines correspond to the real value of $c_1$.

In a more general framework, we extended the models presented in Table 1 to the case where the linear relationship is in an interval inside of the data range, i.e., $c_0 > 0$ and $c_1 < 1$, where $c_0 = 0.16$ and $c_1 = 0.84$. This corresponds to the third case of Section 2.2. Models 5-8 presented in Table 3 are similar to models 1-4, respectively, and can be seen as generalizations. In this case, we have also generated 1,000 replicas for each model and each sample size $n = 51, 101, 251, 501$. The indexes $(j, k)$ used to calculate $R^2_{(j,k)}$ were the values of $j$ such that $x_j \leq 0.3$ and the values of $k$ such that $x_k \geq 0.7$.

Table 3. Models used for detecting the change point in the Monte Carlo simulations. The standard deviations $\sigma$ and $\gamma$ for models 5-8 are the same used for models 1-4, respectively.

| Model | $g(x)$ | Linear term | $f(x)$ |
|---|---|---|---|
| 5 | $x$ | $x$ | $x$ |
| 6 | $0.2 + 6x$ | $1 + x$ | $-3.2 + 6x$ |
| 7 | $1.03 + \cos(32(x - 2.16))/3$ | $1 + x$ | $1.97 - \cos(32(x + 1.16))/3$ |
| 8 | $2.21 + (1.52 + 5x)\sin(80x)/5$ | $2 + 2x$ | $4.03 + (5.36 - 2x)\sin(80x)/10$ |

Illustrations of typical examples for each model, with sample size $n = 51$ are presented in Figure 5. Averages (and standard deviations) for the estimated values of $c_0$ and $c_1$ are calculated based on the 1,000 replicas of each model in Table 3. The results are presented in Table 4, where we can observe, as in the first simulation results, good performances in estimating the change points, in this case, $c_0$ and $c_1$. Boxplots are presented in Figures 6

and 7, where it is possible to see indicatives of convergence of the estimates of $c_0$ and $c_1$, respectively.

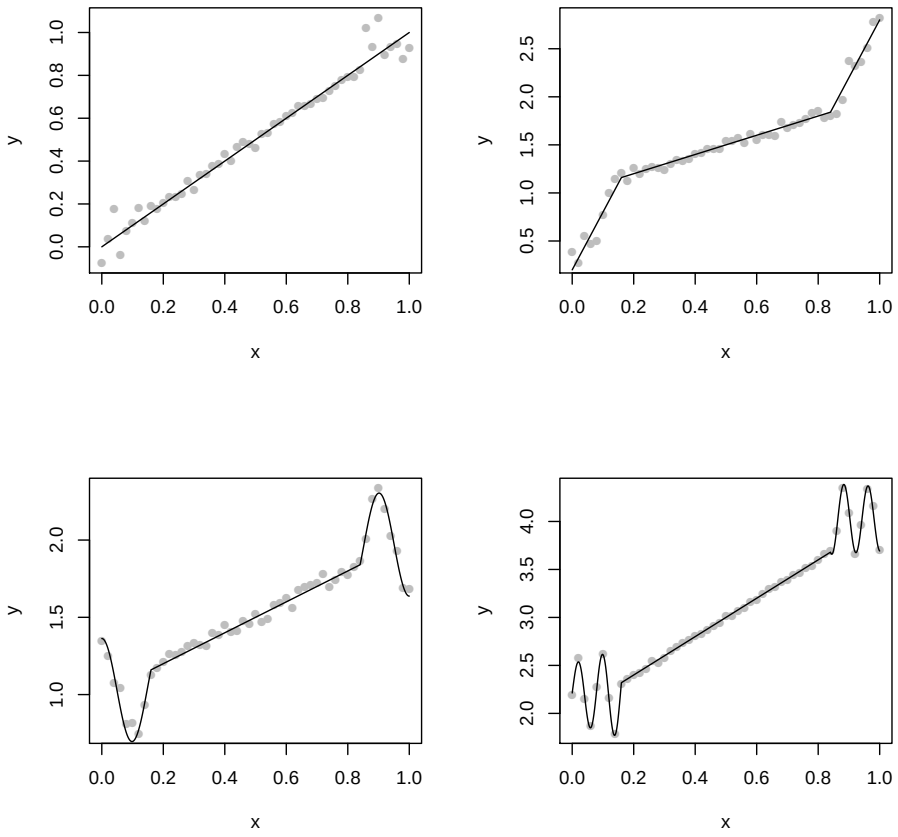


Figure 5. Examples of each model used in the Monte Carlo simulations with sample size $n = 51$. In the top, datasets from models 5 (letf) and 6 (right); in the bottom, datasets from models 7 (left) and 8 (right).

Table 4. Average (standard deviation) of the change point estimates $c_0$ and $c_1$ for the models presented in Table 3.

| Model | Points | $n$ | | | |
|---|---|---|---|---|---|
| | | 51 | 101 | 251 | 501 |
| 5 | $c_0$ | 0.1562 (0.0265) | 0.1581 (0.0117) | 0.1592 (0.0042) | 0.1596 (0.0021) |
| | $c_1$ | 0.8426 (0.0284) | 0.8418 (0.0134) | 0.8407 (0.0024) | 0.8405 (0.0024) |
| 6 | $c_0$ | 0.1602 (0.0230) | 0.1562 (0.0122) | 0.1563 (0.0065) | 0.1576 (0.0042) |
| | $c_1$ | 0.8394 (0.0247) | 0.8434 (0.0144) | 0.8432 (0.0067) | 0.8424 (0.0043) |
| 7 | $c_0$ | 0.1667 (0.0209) | 0.1598 (0.0098) | 0.1567 (0.0049) | 0.1566 (0.0038) |
| | $c_1$ | 0.8334 (0.0215) | 0.8398 (0.0121) | 0.8432 (0.0057) | 0.8433 (0.0036) |
| 8 | $c_0$ | 0.1662 (0.0185) | 0.1624 (0.0075) | 0.1608 (0.0026) | 0.1603 (0.0014) |
| | $c_1$ | 0.8337 (0.0197) | 0.8439 (0.0112) | 0.8434 (0.0060) | 0.8432 (0.0044) |

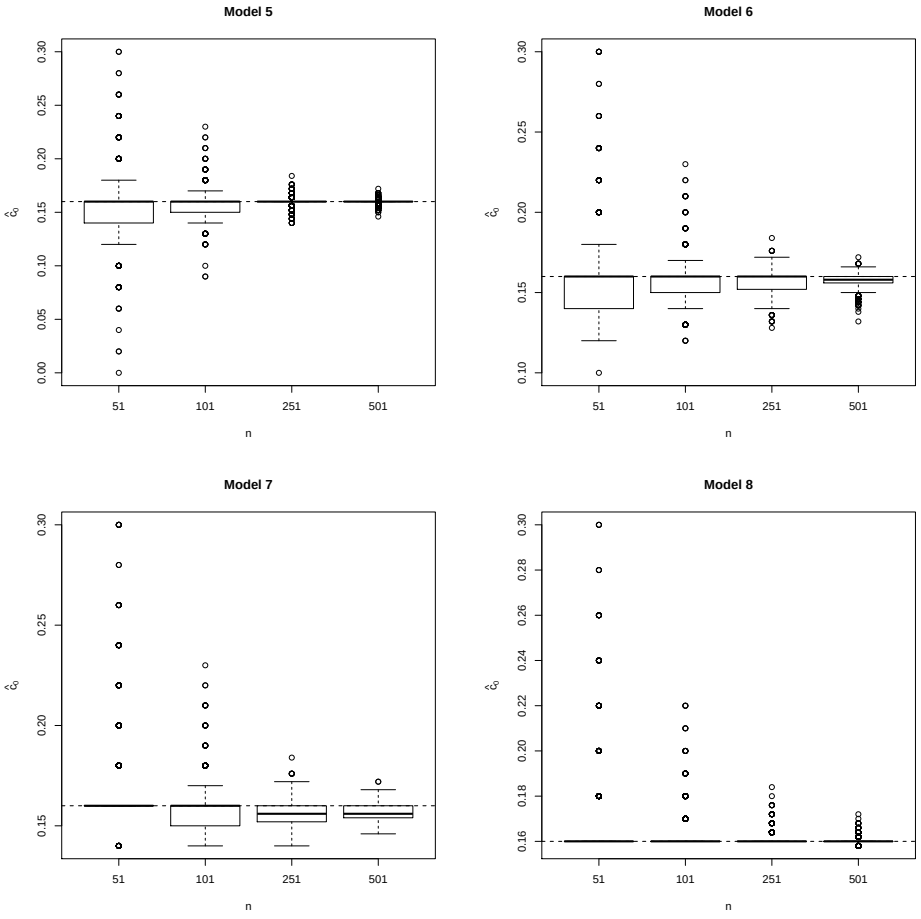


Figure 6. Boxplots of the change point estimates of $c_0$ for the models presented in Table 3.

## 4. Application

Here we apply the method to two real datasets in the field of Analytical Chemistry, where the absorbance and the compound concentrations are analyzed. In the examples, we consider the zinc and iron compounds. In this context, the aim is to find the interval where the Beer-Lambert law holds, i.e., where the relationship between absorbance and concentration is linear.

The datasets were collected in the Analytical Chemistry Laboratory of the Department of Chemistry, at Federal University of Juiz de Fora, Brazil.

### Absorbance vs. zinc concentration

The study was performed by using a stock solution of concentration with 1000 mg $L^{-1}$ of zinc (Vetec, São Paulo, Brazil). By dilution, an analytic curve containing $n = 40$ levels of concentration was prepared for the sequence $Zn = 0.1(0.1)4.0$ mg $L^{-1}$. The solutions were analyzed using FAAS (Thermo Scientific model, Solaar M5 Series). Measurements of the absorbance were performed in triplicates, at the wavelength of 213.9 nm, with air and acetylene flame. The resulting dataset is the same presented before in Figure 1.

In the application of the proposed method to the zinc dataset, we considered the indexes in the range $1 \leq j \leq 5$ for estimating $c_0$ and $10 \leq k \leq 40$ for estimating $c_1$. The method provided the change point estimates $\hat{c}_0 = 0.5$ and $\hat{c}_1 = 1.6$, which corresponds to the

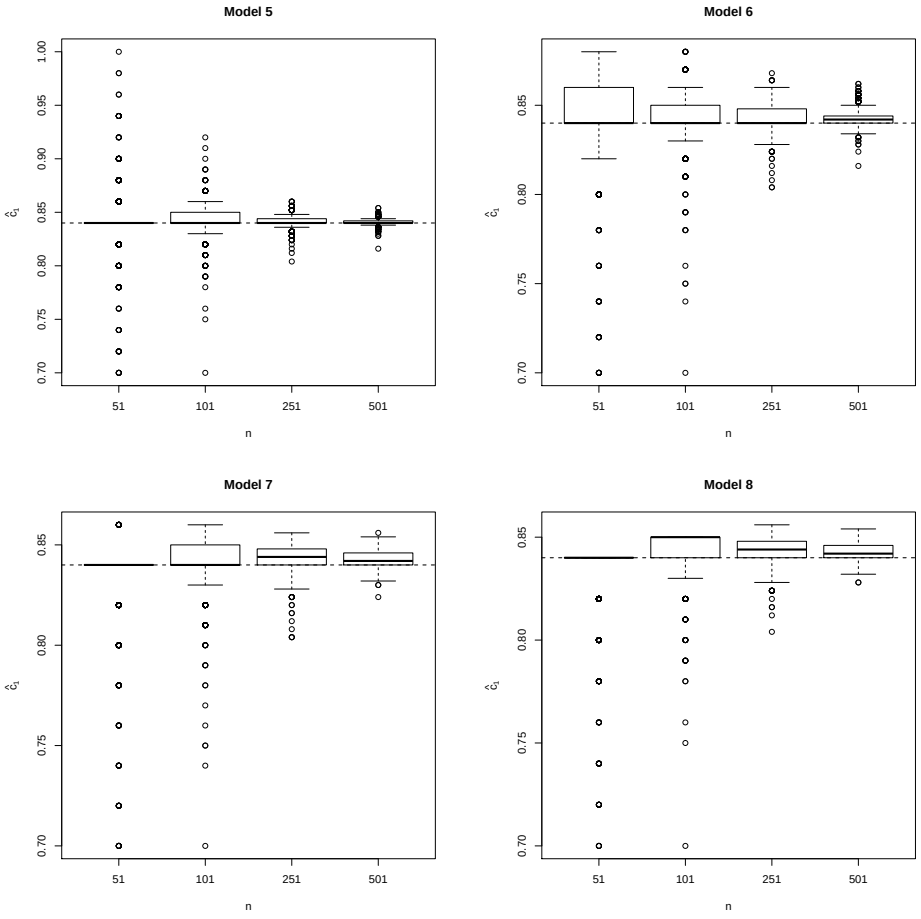


Figure 7. Boxplots of the change point estimates of $c_1$ for the models presented in Table 3.

boundaries where the Beer-Lambert law holds. Furthermore, the coefficient of determination obtained from the data with zinc concentrations in the interval $[0.5, 1.7]$ is 0.9993. In other words, it means that, in the aforementioned interval, 99.93% of the variation of the absorbance can be linearly explained by the zinc concentration. The example is illustrated in Figure 8, where the black points correspond to the data that, according to our approach, is mostly linear. Their boundaries are highlighted by dashed lines, which are the change point estimates.

According to the equipment manual, the linear working range of the analytical curve comes from 0.2 mg $L^{-1}$ to 2 mg $L^{-1}$. This is a standard interval, and it does not take into account random conditions such as, for example, the environment where the equipment is installed. In this case, absorbance values related to zinc concentrations close to 2 mg $L^{-1}$ must lead to unreliable results.

## Absorbance vs. iron concentration

In this study, it was considered a stock solution of concentration with 1000 mg $L^{-1}$ of iron (Vetec, São Paulo, Brazil). $n = 40$ levels of iron concentration was collected, satisfying $Fe = 1(1)40$ mg $L^{-1}$. The solutions were analyzed using FAAS (Thermo Scientific model, Solaar M5 Series) and the absorbance measurements were performed in triplicates, at the wavelength of 248.3 nm, with air and acetylene flame. The dataset is presented in Figure 9.

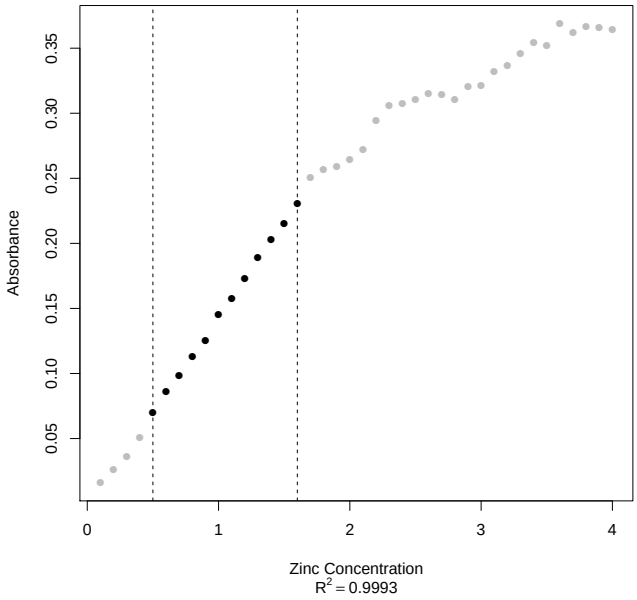


Figure 8. Scatter plot of the absorbance versus the zinc concentration. The dashed lines correspond to the estimated change points $\hat{c}_0$ and $\hat{c}_1$. The black points correspond to the data whose relationship is mostly linear (its coefficient of determination is right below).

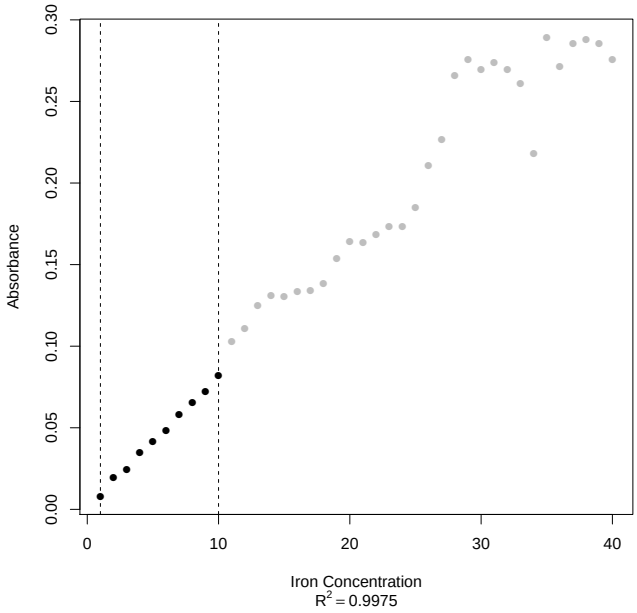


Figure 9. Scatter plot of the absorbance versus the iron concentration. The dashed lines correspond to the estimated change points $\hat{c}_0$ and $\hat{c}_1$. The black points correspond to the data whose relationship is mostly linear (its coefficient of determination is right below).

As in the zinc study, we considered the indexes in the range $1 \leq j \leq 5$ for estimating $c_0$ and $10 \leq k \leq 40$ for estimating $c_1$. The change point estimates obtained are $\hat{c}_0 = 0.1$ and $\hat{c}_1 = 10$, which correspond to the boundaries where the Beer-Lambert law holds. The coefficient of determination based on the data with iron concentration in the interval $[0.1, 10]$ is 0.9975, which means that 99.75% of the variation of the absorbance can be

linearly explained by the iron concentration. The result is illustrated in Figure 9.

The results obtained are, in this case, in agreement with the linear working range provided by the equipment manual, which states that the Beer-Lambert law holds up to 10 mg $L^{-1}$.

## 5. Concluding and additional remarks

In this paper, we present a simple method to detect the change points in partially linear models, based on the well-known coefficient of determination. The approach is illustrated by real dataset applications, and its performance is evaluated by Monte Carlo simulation studies, where we consider different scenarios and sample sizes. The method can efficiently detect the change of relationships.

Our proposal can also be interpreted as a way to find an interval where the data is mostly linear. Technically, we can say that the method provides the best interval where the variation of the dependent variable is linearly explained by the independent variable.

The figures and estimations were performed by using the statistical package R Core Team (2017). The codes to estimate the change points are provided as supplementary material.

We have approached the case where the relationship between dependent and independent variables becomes linear after a change point, say $c_0$, and then, after another change point, say $c_1$, the linearity ends. The method was developed in stages, in order to be easily understood even for practitioners from other fields.

The datasets used for illustration correspond to studies of absorbance and concentration of chemical elements (here we evaluated two cases: the zinc and iron compounds). The idea was to analyze the interval where the Beer-Lambert law holds (in other words, the interval where the relationship between concentration and absorbance is linear). The Beer-Lambert law is well known in Analytical Chemistry and, briefly speaking, it states that there is an interval where the relationship between the absorbance and a specific chemical compound (e.g., the zinc element) is linear.

We believe that our method, if compared to other proposals in the literature, has such a powerful advantage, that the dataset to be analyzed does not need to have a predetermined relationship outside the interval with linear behavior. Furthermore, the proposed approach can detect cases that are variations of model (1). Cases like nonlinear-linear or linear-nonlinear are some examples. Another example is the case where the regression function is fully linear, and a decision rule to test this hypothesis was presented in Section 2.4.

## References

Barry, D., and Hartigan, J.A. (1993). A bayesian analysis for change point problems. Journal of the American Statistical Association 88, 309-319.

Chen, C.W.S., Chan, J.S.K., Gerlach, R., and Hsieh, W.Y.L. (2011). A comparison of estimators for regression models with change points. Statistics and Computing 21, 395-414.

Chen, J. (1998). Testing for a change point in linear regression models. Communications in Statistics - Theory and Methods 27, 2481-2493.

Chen, J., and Gupta, A.K. (2001). On change point detection and estimation. Communications in Statistics - Simulation and Computation 30, 665-697.

Chen, J., and Gupta, A.K. (2012). Parametric Statistical Change Point Analysis. Springer, New York.

Dias, R., Garcia, N.L., Ludwig, G., and Saraiva, M.A. (2015). Aggregated functional data model for near-infrared spectroscopy calibration and prediction. Journal of Applied Statistics 42, 127-143.

Erdman, C., and Emerson, J.W. (2008). A fast bayesian change point analysis for the segmentation of microarray data. Bioinformatics 24, 2143-2148.

Içelli, O., Yalçin, Z., Karakaya, V., and Ilgaz, I.P. (2014). The spectral applications of Beer-Lambert law for some biological and dosimetric materials. AIP Conference Proocedings 1611, 199.

Julious, S.A. (2001). Inference and estimation in a changepoint regression problem. Journal of the Royal Statistical Society, Series D 50, 51-61.

Kim, H.J., and Siegmund, D. (1989). The likelihood ratio test for a change-point in simple linear regression. Biometrika 76, 409-423.

Liu, S., Yamada, M., Collier, N., and Sugiyama, M. (2013). Change-point detection in time series data by relative density-ratio estimation. Neural Networks 43, 72-83.

Osorio, F., and Galea, M. (2006). Detection of a change-point in Student-$t$ linear regression model. Statistical Papers 47, 31-48.

Pastor, R., and Guallar, E. (1998). Use of two-segmented logistic regression to estimate change-points in epidemiologic studies. American Journal of Epidemiology 148, 631-642.

Pettitt, A.N. (1979). A non-parametric approach to the change-point problem. Applied Statistics 28, 126-135.

Quandt, R.E. (1958). The estimation of the parameters of a linear regression system obeying two separate regimes. Journal of the American Statistical Association 53, 873-880.

Quandt, R.E. (1960). Tests of the hypothesis that a linear regression system obeys two separte regimes. Journal of the American Statistical Association 55, 324-330.

R Core Team (2017). R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. URL: www.R-project-org

Rodríguez-Marín, F., Anera, R.G., Alarcón, A., Hita, E., and Jiménez, J.R. (2012). A correction factor for ablation algorithms assuming deviations of Lambert-Beer's law with a gaussian-profile beam. Applied Physics Letters 100, 173703.

Sen, P.K., and Singer, J.M. (2000). Large Sample Methods in Statistics: An Introduction with Applications. Chapman & Hall, London.

Skoog, D.A., West, D.M., Holler, F.J., and Crouch, S.R. (2013). Fundamentals of Analytical Chemistry, 9th Edition. Brooks Cole, Boston.

Worsley, K.J. (1983). Testing for a two-phase multiple regression. Technometrics 25, 35-42.