Statistics and Its Interface Volume 12 (2019) 203–213

Robust change point detection for linear regression models

Aylin Alin*, Ufuk Beyaztas, and Michael A. Martin

Linear models incorporating change points are very common in many scientific fields including genetics, medicine, ecology, and finance. Outlying or unusual data points pose another challenge for fitting such models, as outlying data may impact change point detection and estimation. In this paper, we propose a robust approach to estimate the change point/s in a linear regression model in the presence of potential outlying point/s or with non-normal error structure. The statistic that we propose is a partial F statistic based on the weighted likelihood residuals. We examine its asymptotic properties and finite sample properties using both simulated data and in two real data sets.

AMS 2000 subject classifications: 62J05, 62G35, 62G09.
Keywords and phrases: Bootstrap, Hellinger distance, Simple linear regression, Robustness, Weighted likelihood.

1. INTRODUCTION

Linear regression is one of the most popular statistical models to describe a relationship between a response and a set of independent variables:

$$y_i = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \ldots + \beta_k x_{ik} + \epsilon_i \quad \text{for} \quad i = 1, 2, \ldots, n,$$

where y_i is the response variable, and x_{ij} is the ith observation on the jth independent variable for $j = 1, 2, \ldots, k$. The random quantities $\{\epsilon_i\}$ comprise the set of n independent and identically distributed Normal random variables with $E(\epsilon_i) = 0$ and $Var(\epsilon_i) = \sigma^2$ ($\epsilon_i \sim N(0, \sigma^2)$). While simple linear regression models describe many data sets well, in many cases the nature of the relationship between the response and the independent variables changes at some value or values of the explanatory variables so that the regression relationship is, effectively, "bent" at certain points, these points generally referred to as change points. Obviously, if a model involving such changes is justified, then using a simple linear regression model to study the data results in a poor fit and a lack of explanatory power [8]. Thus, testing for and estimating change points in regression settings are useful capabilities. Regression models incorporating change points have been employed in many areas of application such as cancer research [21, 32], animal science [7], genetics [36], finance [8], meteorology [33], and even in the tire industry [13], in each case reflecting experience or theory that the nature of the regression relationship fundamentally changes at certain threshold values of the covariates. Approaches to change point detection have been numerous in the literature, using a variety of techniques, for example "traditional" approaches based on residual sums of squares, Bayesian approaches, information theoretic approaches, and employing permutation tests, likelihood ratio tests, partial F tests or score tests [17, 8, 6, 9, 12, 19, 26, 30, 31, 34, 35, 27, 14, 15, 23, 20, 28]). [29] provides an in-depth discussion on a normal mean change point model that offers insight into the modelling situation underpinning this paper.

Quite apart from the broad question of how to detect and estimate change points in linear models, the effects of individual data points on change point detection and estimation need also to be considered. For example, there may be discordant data points that suggest a change point but which are, in fact, a result of errors or excessively "noisy" data. These points, often called outliers or influential points, are common in real data sets. Such points may reflect unusual values in the response variable or in the covariate space. Both types typically adversely affect the regression model, albeit in different ways. Outlying points may, in a sense, be false flags for change points – that is, they may be confused as evidence of a structural change in the relationship when they are, in fact, not of structural significance. Robust change point detection techniques have thus been proposed as a solution for outliers – for a detailed survey on robust change point detection see, for example, [18]. Test statistics based on divergence measures have also been used for robust change point detection. For example, [22] considered a method for robust change point estimation for a general parameter case based on the minimum density power divergence estimator proposed by [5]. A detailed study of extensions of some classical methods in change point analysis using divergence-based statistics is provided by [26]. In this paper, we propose a new robust approach to detect change points. Our approach will be shown to be robust not only to outliers but also to non-normal error structure. We introduce a robust partial F test, F_{wle}, based on weighted likelihood methodology to estimate the change points in the relationship between the response and the covariates. The statistic F_{wle} is a robust alternative to the classical F. Ours is the first study using a robust approach based on the

*Corresponding author.

weighted likelihood methodology in linear regression models with change points.

In our formulation, change may be in the mean structure or in the variance structure or in both. By way of introduction, we examine changes in the mean structure in the case of simple linear regression model ($k = 1$) with fixed variance structure and the proposed model is continuous at the change point. While our focus in this paper is on estimating one or two unknown change points, the procedure can be readily extended to situations with several change points.

The rest of the paper is organized as follows. In Section 2, we describe the change point problem for a simple linear regression model. In Section 3, we introduce the proposed test statistic and show the asymptotic equivalence of the proposed F_{wle} and the classical F statistic. We compare the finite sample performance of two statistics on simulated and real data sets in Section 4.

2. THE CHANGE POINT MODEL

2.1 Simple linear regression with one change point

Consider two putative regression models relating the response $\boldsymbol{y}$ to a single covariate $\boldsymbol{x}$.

$$y_i = \beta_0 + \beta_1 x_i + \epsilon_i \quad i = 1, 2, \ldots, n, \tag{1}$$

and

$$y_i = \begin{cases} \beta_{10} + \beta_{11} x_i + \epsilon_i, & a \le x_i \le \gamma; \\ \beta_{20} + \beta_{21} x_i + \epsilon_i, & \gamma < x_i \le b, \end{cases} \tag{2}$$

under the continuity constraint

$$\beta_{10} + \beta_{11}\gamma = \beta_{20} + \beta_{21}\gamma. \tag{3}$$

Without loss of generality, the explanatory variable is ordered, with $x_i \le x_j$ for $i < j$, so that "change" is interpreted in the sense of a "bend" in the model when the covariate increases beyond the change point, γ, the condition (3) ensuring continuity of the model at the change point. The change-point γ is not constrained to be one of the data points, and can take on any value between two adjacent existing data points. The situations covered by our prescription fits both type 1 and type 2 joins defined in [17]. Combining (2) and (3), the model (2) can be rewritten as

$$y_i = \begin{cases} \beta_{10} + \beta_{11} x_i + \epsilon_i, & x_i \le \gamma; \\ \beta_{10} + \beta_{11}\gamma + \beta_{21}(x_i - \gamma) + \epsilon_i, & x_i > \gamma, \end{cases} \tag{4}$$

If γ is known, (3) reflects a linear constraint on the unknown parameters $\boldsymbol{\beta_1} = (\beta_{10}, \beta_{11})$ and $\boldsymbol{\beta_2} = (\beta_{20}, \beta_{21})$ but when γ is unknown it reflects a nonlinear constraint on the unknown parameters γ, $\boldsymbol{\beta_1} = (\beta_{10}, \beta_{11})$, and $\boldsymbol{\beta_2} = (\beta_{20}, \beta_{21})$ [17]. Under the continuity constraint, once $(\beta_{10}, \beta_{11}, \beta_{21})$ are estimated for model (4), β_{20} in model (2) can be calculated using (3).

We use a hypothesis testing approach to determine if there is any significant change point. We will test

$$\begin{aligned} H_0 &: \boldsymbol{\beta_1} = \boldsymbol{\beta_2} = \boldsymbol{\beta} \\ H_a &: \boldsymbol{\beta_1} \ne \boldsymbol{\beta_2}, \end{aligned} \tag{5}$$

where $\boldsymbol{\beta} = (\beta_0, \beta_1)$ are the coefficients in model (1). Under the continuity constraint in (3), (5) is equivalent to

$$\begin{aligned} H_0 &: \beta_{11} = \beta_{21} = \beta_1 \\ H_a &: \beta_{11} \ne \beta_{21}, \end{aligned}$$

since $\beta_{11} = \beta_{21}$ implies that $\beta_{10} = \beta_{20}$. Failing to reject H_0 means there is insufficient evidence from the data to conclude a change in the mean structure at $x = \gamma$. The classical partial F-statistic using ordinary least squares residuals can be used to compare model (1) with model (4):

$$F = \frac{(SSE_{(0)} - SSE_{(a)})/1}{SSE_{(a)}/(n-3)}, \tag{6}$$

where

$$SSE_{(a)} = \sum_{i=1}^{n} r_i(\tilde{\boldsymbol{\beta}}_{\mathbf{1}}; \tilde{\boldsymbol{\beta}}_{\mathbf{2}})^2,$$

$$SSE_{(0)} = \sum_{i=1}^{n} r_i(\tilde{\boldsymbol{\beta}}_{\mathbf{1}})^2,$$

are the residual sums of squares under (1) and (4), respectively, the vectors $\tilde{\boldsymbol{\beta}}_{\mathbf{1}} = \left(\tilde{\beta}_{10}, \tilde{\beta}_{11}\right)$ and $\tilde{\boldsymbol{\beta}}_{\mathbf{2}} = \left(\tilde{\beta}_{20}, \tilde{\beta}_{21}\right)$ are the ordinary least squares (OLS) estimates of the respective parameters, and the functions $r_i(\tilde{\boldsymbol{\beta}}) = y_i - (\tilde{\beta}_{10} + \tilde{\beta}_{11} x_i)$ and $r_i(\tilde{\boldsymbol{\beta}}_{\mathbf{1}}, \tilde{\boldsymbol{\beta}}_{\mathbf{2}}) = y_i - \{\tilde{\beta}_{10} + \tilde{\beta}_{11} x_i + I(\gamma < x_i \le b)\tilde{\beta}_{21}(x_i - \gamma)\}$ are the residuals, where $I(\cdot)$ is the indicator function. When γ is known, this quantity is asymptotically F distributed with 1 and $n-3$ degrees of freedom [19]. The larger the value of F, the greater the evidence is against H_0; that is, the more the evidence for a change point at γ. Use of the F statistic is justified through expressing model (1) as a suitably restricted, or nested, version of model (4). Following [16], this construction is made clear by expressing (4) in the form

$$y_i = \beta_{10} + \beta_{11} x_i + z_i(\eta_0 + \eta_1 x_i) + \epsilon_i, \quad i = 1, 2, \ldots, n,$$

where $\boldsymbol{\beta_2} = \boldsymbol{\beta_1} + \boldsymbol{\eta}$ with $\boldsymbol{\eta} = (\eta_0, \eta_1)'$ and $z_i = I(x_i > \gamma)$.

2.2 The model with two change points

As an extension to the model discussed in the previous section, we consider models with two change points at γ_1 and γ_2:

$$y_i = \begin{cases} \beta_{10} + \beta_{11} x_i + \epsilon_i, & x_i \le \gamma_1, \\ \beta_{20} + \beta_{21} x_i + \epsilon_i, & \gamma_1 < x_i \le \gamma_2, \\ \beta_{30} + \beta_{31} x_i + \epsilon_i, & x_i > \gamma_2, \end{cases} \tag{7}$$

again assuming without loss of generality that the explanatory variable is ordered as $x_i \le x_j$ for $i < j$, and with continuity constraints:

$$\beta_{10} + \beta_{11}\gamma_1 = \beta_{20} + \beta_{21}\gamma_1, \qquad (8)$$
$$\beta_{20} + \beta_{21}\gamma_2 = \beta_{30} + \beta_{31}\gamma_2.$$

Using the continuity constraints in (8), the model in (7) can be rewritten as

$$y_i = \begin{cases} \beta_{10} + \beta_{11}x_i + \epsilon_i, & x_i \le \gamma_1 \\ \beta_{10} + \beta_{11}\gamma_1 + \beta_{21}(x_i - \gamma_1) + \epsilon_i, & \gamma_1 < x_i \le \gamma_2 \\ \beta_{10} + \beta_{11}\gamma_1 + \beta_{21}(\gamma_2 - \gamma_1) + & \\ \beta_{31}(x_i - \gamma_2) + \epsilon_i & x_i > \gamma_2, \end{cases}$$

Evidence for the existence of two change points, γ_1 and γ_2 can be gathered through a hypothesis test of

$$H_0 : \beta_{11} = \beta_{21} = \beta_{31} = \beta_1$$
$$H_a : \beta_{11} \ne \beta_{21} \quad \text{and} \quad \beta_{21} \ne \beta_{31}.$$

The classical F statistic similar to that in (6), modifying the degrees of freedom to be 2 and $n-4$, can again be used to test these hypotheses, with the null distribution of the statistic F being F with 2 and $n-4$ degrees of freedoms when the location of the change point is known.

3. THE ROBUST WEIGHTED PARTIAL F TEST

3.1 The weighted likelihood methodology

Weighted likelihood methodology based on the minimum disparity estimation was proposed by [25] as a means of improving the efficiency and robustness in estimation, and discussed in the regression context by [1], [2], and [4]. Let us consider the simple linear regression model (1) with y having density function $f = f(y; \boldsymbol{x}, \boldsymbol{\beta}, \sigma)$. Assume a density $m(\epsilon; \sigma)$, $\sigma \in \mathbb{R}^+$ for the theoretical residual ϵ having zero mean for each given σ. Denote by $r_i(\boldsymbol{\beta}) = y_i - (\beta_0 + \beta_1 x_i)$ the residuals for a specific value of the parameter vector $\boldsymbol{\beta}$. Let $f^*(r; \boldsymbol{\beta}) = \int k(r, t, h) d\hat{F}_n(t; \boldsymbol{\beta})$ represent a kernel density estimator with bandwidth h based on the empirical distribution $\hat{F}_n(\cdot; \boldsymbol{\beta})$ of the residuals $r_i(\boldsymbol{\beta})$. Let $m^*(r; \sigma) = \int k(r, t, h) dM(t; \sigma)$ be the kernel smoothed version of the residual model density where $M(\cdot; \sigma)$ is the distribution function corresponding to the density $m(\cdot; \sigma)$. The Pearson residuals are $\delta(r; \sigma, \hat{F}_n(\boldsymbol{\beta})) = \frac{f^*(r;\boldsymbol{\beta})}{m^*(r;\sigma)} - 1$, which capture the agreement between the empirical distribution of the residuals and their assumed probability model. Following [25], we use $h = \sqrt{k\sigma^2}$, where k is a constant independent of the scale of the data, so that very small weight is assigned to an outlying observation. Unlike the usual definition of an outlier as a point lying far away from the bulk of the data, in the present context an outlier is defined as a point unlikely to occur under the assumed probabilistic model. The weighted likelihood of the parameter $\boldsymbol{\beta}$ and the scale parameter σ are the solutions to the estimating equations

$$\sum_{i=1}^{n} \omega(r_i(\boldsymbol{\beta})) u_{\boldsymbol{\beta}}(y_i) = 0,$$
$$\sum_{i=1}^{n} \omega(r_i(\boldsymbol{\beta})) u_{\sigma}(y_i) = 0,$$

where the functions $u_{\boldsymbol{\beta}}(y_i) = \frac{\partial}{\partial \boldsymbol{\beta}} \ln f(y_i|\boldsymbol{x}; \boldsymbol{\beta}; \sigma)$ and $u_\sigma(y_i) = \frac{\partial}{\partial \sigma} \ln f(y_i|\boldsymbol{x}; \boldsymbol{\beta}; \sigma)$ are the usual score functions and the weight function $\omega(r_i(\boldsymbol{\beta}))$ is

$$\omega(r_i(\boldsymbol{\beta})) = \min\left\{1, \frac{[A(\delta(r_i(\boldsymbol{\beta}))) + 1]^+}{\delta(r_i(\boldsymbol{\beta})) + 1}\right\},$$

where $[\cdot]^+$ is the positive part function and weights are constrained to be in $[0, 1]$. The function $A(\cdot)$ as introduced by [24] is a residual adjustment function. The choice $A(\delta) = \delta$ corresponds to usual maximum likelihood estimates (MLE) for the parameters, while the choice

$$A(\delta) = 2(\delta + 1)^{1/2} - 1 \qquad (9)$$

produces weighted likelihood estimates (WLE) based on Hellinger distance weights, the latter producing more stable estimates than the usual maximum likelihood approach when certain assumptions do not hold.

3.2 Change point location known case

Since the classical F statistic is calculated using the ordinary least squares (equivalent to maximum likelihood in the case of normal errors) residuals, it is sensitive to outlying points. Therefore, we propose using an F-statistics based on the weighted likelihood residuals to test the null hypothesis of the no change model in (1) against change point alternatives. We propose using the Hellinger distance weights given in (9). The use of these weights promoted robustness against both outliers having unusual y values and against points having high leverage. The proposed F statistic for testing the no change model against the one change model is

$$F_{wle} = \frac{(SSE_{wle(0)} - SSE_{wle(a)})/1}{SSE_{wle(a)}/\left(\sum_{i=1}^{n} \hat{\omega}(r_i) - 3\right)}. \qquad (10)$$

Here, $SSE_{wle(0)}$ and $SSE_{wle(a)}$ are the weighted likelihood versions of the error sum of squares under the null and alternative hypotheses

$$SSE_{wle(a)} = \sum_{i=1}^{n} \hat{\omega}(r_i) r_i(\hat{\boldsymbol{\beta}}_{\mathbf{1}}; \hat{\boldsymbol{\beta}}_{\mathbf{2}})^2,$$
$$SSE_{wle(0)} = \sum_{i=1}^{n} \hat{\omega}(r_i) r_i(\hat{\boldsymbol{\beta}}_{\mathbf{1}})^2,$$

where $\hat{\boldsymbol{\beta}}_{\mathbf{1}} = (\hat{\beta}_{10}, \hat{\beta}_{11})$, $\hat{\boldsymbol{\beta}}_{\mathbf{2}} = (\hat{\beta}_{20}, \hat{\beta}_{21})$ are the weighted likelihood estimates (WLE) of the parameters estimated under the assumption that $m(r;\sigma)$ is $N(0,\sigma^2)$, $r_i(\hat{\boldsymbol{\beta}}_1) = y - (\hat{\beta}_{10} + \hat{\beta}_{11}x)$ and $r_i(\hat{\boldsymbol{\beta}}_1, \hat{\boldsymbol{\beta}}_2) = y_i - (\hat{\beta}_{10} + \hat{\beta}_{11}\gamma + I(\gamma < x_i \leq b)\hat{\beta}_{21}(x_i - \gamma))$ are the residual functions, and the weights $\hat{\omega}(r_i) = \omega(r_i(\hat{\boldsymbol{\beta}}_1;\hat{\boldsymbol{\beta}}_2);\sigma, \hat{F}_n(\hat{\boldsymbol{\beta}}_1;\hat{\boldsymbol{\beta}}_2))$. The proposed statistic F_{wle} may be written as

$$F_{wle} = \frac{\hat{\sigma}^2_{wle(0)} - \hat{\sigma}^2_{wle(a)}}{\hat{\sigma}^2_{wle(a)}} \frac{\sum_{i=1}^n \hat{\omega}(r_i) - 3}{1}, \tag{11}$$

with the estimators of scale parameter for models under the null and alternative hypotheses being

$$\hat{\sigma}^2_{wle(0)} = \frac{1}{\sum_{i=1}^n \hat{\omega}(r_i)} \sum_{i=1}^n r_i(\hat{\boldsymbol{\beta}}_1)^2,$$

$$\hat{\sigma}^2_{wle(a)} = \frac{1}{\sum_{i=1}^n \hat{\omega}(r_i)} \sum_{i=1}^n r_i(\hat{\boldsymbol{\beta}}_1;\hat{\boldsymbol{\beta}}_2)^2.$$

Under the assumption of the change point location being known, asymptotic equivalence of (11) and the classical F, that is, the null distribution of $F_{wle} \sim F_{1,n-3}$, follows from Theorem 1 in [3].

For two change points, the proposed F_{wle} is defined analogously to (10), but the quantities 1 and $\sum_{i=1}^n \omega(r_i(\hat{\boldsymbol{\beta}}_1;\hat{\boldsymbol{\beta}}_2)) - 3$ therein are replaced by 2 and $\sum_{i=1}^n \omega(r_i(\hat{\boldsymbol{\beta}}_1;\hat{\boldsymbol{\beta}}_2;\hat{\boldsymbol{\beta}}_3)) - 4$, respectively, with $\hat{\boldsymbol{\beta}}_3 = (\hat{\beta}_{30}, \hat{\beta}_{31})$.

3.3 Unknown change point location case

When the location of the change point is unknown, and therefore must be estimated along with the regression parameters, (3) and (8) impose non-linear constraints on the parameters and change point/s. Thus, in practice we need to estimate the parameters through numerical optimization [19], using techniques as described in [17] or [19]. [23]'s discrete grid search is another commonly cited technique used in such a situation. Here, we propose using a grid search in concert with the hypothesis testing approach described in the preceding section. Consider G equi-spaced grid points spanning the range of values of the independent variable, indexed by $g = 1, \ldots, G$. We omit the first and last grid points from our search, as a change point at the extremes of the grid cannot reasonably be posited given the available data. For the case of a single putative change point, we consider each of the values in the grid as the potential change point, and for each such candidate we calculate the F_{wle} statistic, denoted $F_{wle(g)}$ for $g = 2, \ldots, G-1$ for the set of G such F_{wle} values. Similarly, for the case of two putative change points, we would calculate the value of F_{wle}, denoted $F_{wle(g_1,g_2)}$ for each pair (g_1, g_2) for which $2 \leq g_1 \leq G-1$ and $g_1 < g_2 \leq G-1$ and imposing the requirement on (g_1, g_2) that there be at least 3 data points positioned between the grid points indexed by g_1 and g_2. This latter requirement, for which the value 3 is an arbitrary choice that could be raised as appropriate to the specific situation, is to allow for reasonable estimation precision of regression parameters between the two change points. Our rationale in computing the values of F_{wle} at each grid point is that at the grid points closest to "correct" change point values, the corresponding F statistic should be large compared to the other F values. Thus, we estimate as the change point the grid value for which $F_{wle(g)}$ (one change-point) or $F_{wle(g_1,g_2)}$ (two change points) is maximized across the grid values. Note that we use the WLE F-statistic as opposed to the usual OLS-based F-statistic to take advantage of the robustness of the former against outlying observations or non-normal error structure. By way of comparison, we can also estimate the change points using the maximum of the classical (OLS-based) F-statistics, and we denote the maximum F values for the WLE and classical cases as F_{wlemax} and F_{max}, respectively.

Considering the case of a single change point (the two change-point case proceeds analogously), once we have a candidate for a change point, we need to determine its significance. The distributions of each of the F_{wlemax} and F_{max} statistics are, respectively, the distributions of the maximum of $G-2$ correlated random variables. We propose using the nonparametric bootstrap [11] to estimate the sampling distributions of F_{wlemax} and F_{max}, with such resampling reflecting the null hypothesis assumption of no change point over the range of available data. There are two "standard" bootstrap approaches for linear regression models: case-based resampling, based on resampling (x, y) pairs; and residual-based resampling, based on resampling residuals from a model fit and then building back resampled cases using the fitted model and resampled residuals. In this paper, we consider residual-based resampling. Specifically, for each of the ordinary least squares and wle fitting methods, respectively, within this resampling we impose the null hypothesis by first fitting the no-change-point model (the correct model under the null hypothesis) to the original data, computing residuals from this fit, and then selecting B same-size random samples with replacement from these residuals before re-assembling residual-based bootstrap resamples using the original model fit applied to the x_i for each fitting method. For the WLE method, the resamples are (x_i, y_i^{*b}) for $b = 1, 2, \ldots, B$ where $y_i^{*b} = \hat{\beta}_0 + \hat{\beta}_1 x_i + e_i^{*b}$ where $\hat{\beta}_0$ and $\hat{\beta}_1$ are the WLE regression coefficient estimates from the fit to the original data and the $\{e_i^{*b}\}$ are the resampled WLE residuals. Similarly, for the OLS fitting method, we use resamples $(x_i, \tilde{y}_i^{*b})$ for $b = 1, 2, \ldots, B$ where $\tilde{y}_i^{*b} = \tilde{\beta}_0 + \tilde{\beta}_1 x_i + \tilde{e}_i^{*b}$ where $\tilde{\beta}_0$ and $\tilde{\beta}_1$ are the OLS regression coefficient estimates from the fit to the original data and the $\{\tilde{e}^{*b}\}$ are the resampled OLS residuals. Because the original no-change-point model parameters are included in the construction of resamples, this approach explicitly imposes the null hypothesis within the resampling. For each of those B resamples, we calculate the F_{max} and F_{wlemax} statistics, designating F^{*b}_{max} and F^{*b}_{wlemax} as the respective values of the two statistics arising from the b'th bootstrap resample ($b = 1, \ldots, B$). For an α-level test, we use as the critical value the $(1-\alpha)$'th percentile of the bootstrap sampling distributions of the two F statistics, denoted $F^*_{max_{(1-\alpha)}}$

and $F^*_{wlemax_{(1-\alpha)}}$, respectively. The estimated p-values for the test based on the original sample statistics F_{max} and F_{wlemax} are also simply calculated as

$$p\text{-value}_{(wlemax)} = \frac{1}{B}\sum_{b=1}^{B} I(F^{*b}_{wlemax} > F_{wlemax}),$$

$$p\text{-value}_{(max)} = \frac{1}{B}\sum_{b=1}^{B} I(F^{*b}_{max} > F_{max}).$$

We note that from a practical perspective, change point detection close to the endpoints of the data range may be simply infeasible as potential change points at the extremes of the data may not be distinguished from more benign outlier behaviour. Of course, edge effects are common artefacts of most model validation exercises.

4. NUMERICAL RESULTS

4.1 Simulation study

In this section, we report on a simulation study to explore the finite sample properties of the statistics F_{max} and F_{wlemax}, for a simple linear regression model versus one change point and two change points models, respectively. First, we consider a single change point alternative. We consider three low to moderate sample sizes $n = 20, 40, 60$. For convenience, we choose $x = 1, \ldots, n$, $\boldsymbol{\beta_1} = (0, 1)$ and $\boldsymbol{\beta_2} = (\gamma(1-\beta_{21}), \beta_{21})$ with $\beta_{21} = 4, 5, 6$ and change point value γ. We simulate data from four different scenarios:

1 Normally distributed errors: $\epsilon \sim N(0,1)$,
2 Errors having heavy-tailed but symmetric distribution: $\epsilon \sim t(5)$,
3 Normal errors with 10% contamination: $N(0,1)$ with probability 0.9 and $N(20,1)$ with probability 0.1,
4 Normal errors ($\epsilon \sim N(0,1)$) but with one deliberately placed outlier in terms of its y value.

The values on which Figures 1–3 are based were calculated using 10,000 Monte-Carlo simulations. They illustrate the detection rates – that is, the proportion of the cases for which the statistics correctly detect the change point. Since the performance of the methods are very similar across the considered β_{21} values, we only present the results for $\beta_{21} = 5$. For the case of standard normal errors without any contamination, both statistics perform very similarly, as expected. For the case of heavy-tailed errors, $\epsilon \sim t(5)$, the classical F_{max} performs slightly better for the smallest sample size, but that difference vanishes as the sample size gets larger. We also consider the normal errors case with one deliberately-placed outlier as the $n/2$'th data point. Unlike F_{max}, the proposed F_{wlemax} proves robust against the inclusion of this outlier. For the case of errors with 10% contamination, the proposed F_{wlemax} was able to detect the change point successfully except when the change point was

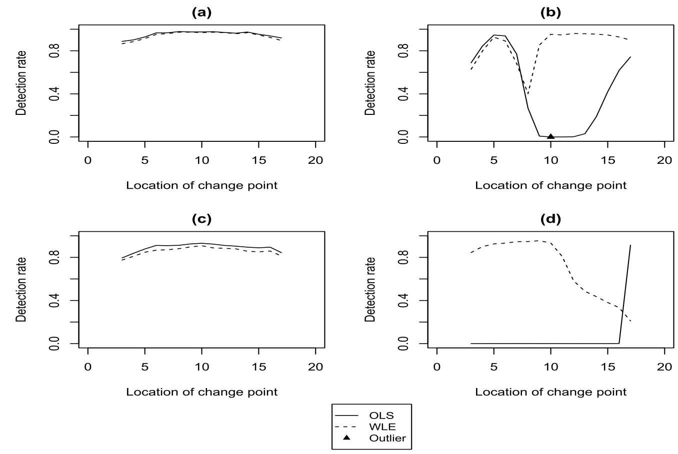

Figure 1. Estimated detection rate when $n = 20$ and $\beta_{21} = 5$ (a) $\epsilon \sim N(0,1)$ (b) $\epsilon \sim N(0,1)$ with one outlier y_i at $i = 10$ (c) $\epsilon \sim t(5)$ (d) $\epsilon \sim N(0,1)$ w.p. 0.9 and $N(20,1)$ w.p. 0.1.

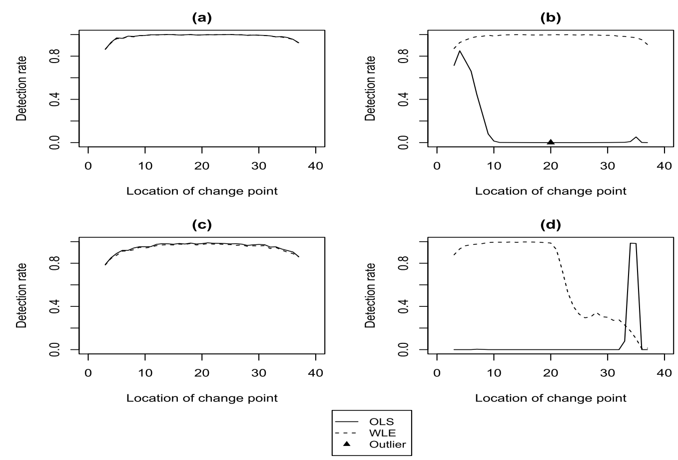

Figure 2. Estimated detection rate when $n = 40$ and $\beta_{21} = 5$ (a) $\epsilon \sim N(0,1)$ (b) $\epsilon \sim N(0,1)$ with one outlier y_i at $i = 20$ (c) $\epsilon \sim t(5)$ (d) $\epsilon \sim N(0,1)$ w.p. 0.9 and $N(20,1)$ w.p. 0.1.

located towards the right extreme of the x-data range, reflecting the difficulty that the test statistic faces in distinguishing between the effect of a systemic change point in the relationship and of a large outlying observation. This phenomenon suggests that, in general, if a change point is too close to the edges of the range over which data is gathered, the capability of algorithms to detect system change is diminished. Indeed, this effect can be seen in all cases examined, with the detection rate declining at each end of the data range.

We also examine the observed type-I error probabilities (α) for each of the F_{max} and F_{wlemax} tests by using a bootstrap approach. Under the null hypothesis assumption of no change point, we carry out $S = 500$ Monte Carlo simulations, and for each of those simulated samples, we select

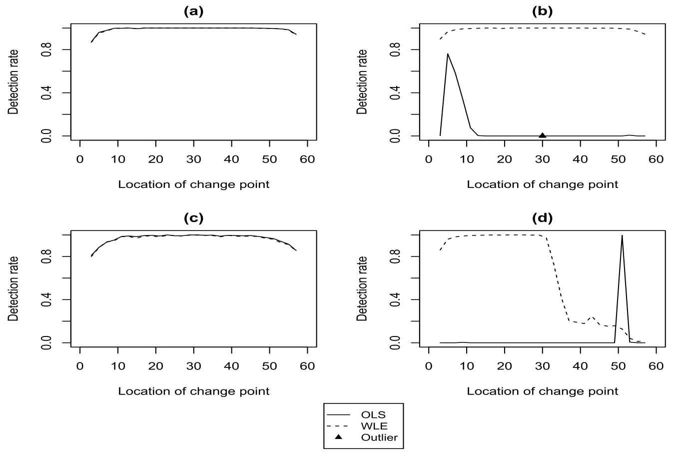

Figure 3. Estimated detection rate when $n = 60$ *and* $\beta_{21} = 5$
(a) $\epsilon \sim N(0, 1)$ *(b)* $\epsilon \sim N(0, 1)$ *with one outlier* y_i *at* $i = 30$
(c) $\epsilon \sim t(5)$ *(d)* $\epsilon \sim N(0, 1)$ w.p. 0.9 and $N(20, 1)$ w.p. 0.1.

$B = 300$ bootstrap resamples. The steps to calculate the observed α values are:

(1) Simulate a sample, and fit the no-change-point model to the sample using each of the OLS and WLE fitting methods, respectively, and calculate the residuals $(\tilde{e}_1, \tilde{e}_2, \ldots, \tilde{e}_n)$ and $(e_1, e_2, \ldots, e_n)$ to obtain OLS coefficient estimates $\tilde{\beta}_0$ and $\tilde{\beta}_1$ and WLE coefficient estimates $\hat{\beta}_0$ and $\hat{\beta}_1$.
(2) Calculate the corresponding statistics F_{max} and F_{wlemax}.
(3) Select B bootstrap resamples, $\tilde{\boldsymbol{e}}^{*\boldsymbol{b}} = (\tilde{e}_1^{*b}, \tilde{e}_2^{*b}, \ldots, \tilde{e}_n^{*b})^\top$ and $\boldsymbol{e}^{*\boldsymbol{b}} = (e_1^{*b}, e_2^{*b}, \ldots, e_n^{*b})^\top$ for $b = 1, 2, \ldots, B$, by same-size random sampling with replacement from each set of residuals calculated in Step (1).
(4) Re-assemble residual-based bootstrap resamples $(x_i, \tilde{y}_i^{*b})$ where $\tilde{y}_i^{*b} = \tilde{\beta}_0 + \tilde{\beta}_1 x_i + \tilde{e}_i^{*b}$, and (x_i, y_i^{*b}) where $y_i^{*b} = \hat{\beta}_0 + \hat{\beta}_1 x_i + e_i^{*b}$ for $b = 1, 2, \ldots, B$, where $(\tilde{\beta}_0,\ \tilde{\beta}_1)$ and $(\hat{\beta}_0,\ \hat{\beta}_1)$ are the original OLS and WLE coefficient estimates calculated in Step (1), and x_i is the i'th sample point from $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)$.
(5) Corresponding to each of the bootstrap resamples $(x_i, \tilde{y}_i^{*b})$ and (x_i, y_i^{*b}), calculate the bootstrap statistics F_{max}^{*b} and F_{wlemax}^{*b} for $b = 1, 2, \ldots, B$.
(6) Calculate the $(1-\alpha)$'th percentiles (call these the critical values), $F^*_{max_{(1-\alpha)}}$ and $F^*_{wlemax_{(1-\alpha)}}$, of the above two statistics based on their bootstrap sampling distributions.
(7) Determine if the original statistics exceed the critical values obtained in Step (6).
(8) To estimate the estimated α's via simulation, repeat steps (1)-(7) S times and calculate the proportion of simulated samples for which the original statistics exceed their corresponding bootstrap critical values:

$$\alpha_{F_{max}} = \frac{1}{S}\sum_{i=1}^{S} I(F_{max_i} > F^*_{max_{(1-\alpha)}})$$

$$\alpha_{F_{wlemax}} = \frac{1}{S}\sum_{i=1}^{S} I(F_{wlemax_i} > F^*_{wlemax_{(1-\alpha)}}),$$

where $I(\cdot)$ is the indicator function. These quantities estimate the probabilities, under the null hypothesis, of observing a value as or more extreme than those originally observed, respectively, for the two tests statistics – that is, the observed significance level, or type-1 error, in each case.

Figure 4 illustrates the calculated observed α values for $n = 20, 40, 60$ under four different error distributions, noting that the nominal α is 0.05. The results for the normal error case are equivocal, but in the other cases, the results reveal that the non-robust F_{max} is the more liberal of the two when the error distribution is heavy-tailed, and much more conservative in the case of normally distributed errors with one outlying point in the y direction. The contaminated error case reveals a significant failure of the test based on F_{max} (Figure 4(d-2)), yielding extraordinary error rates for each sample size.

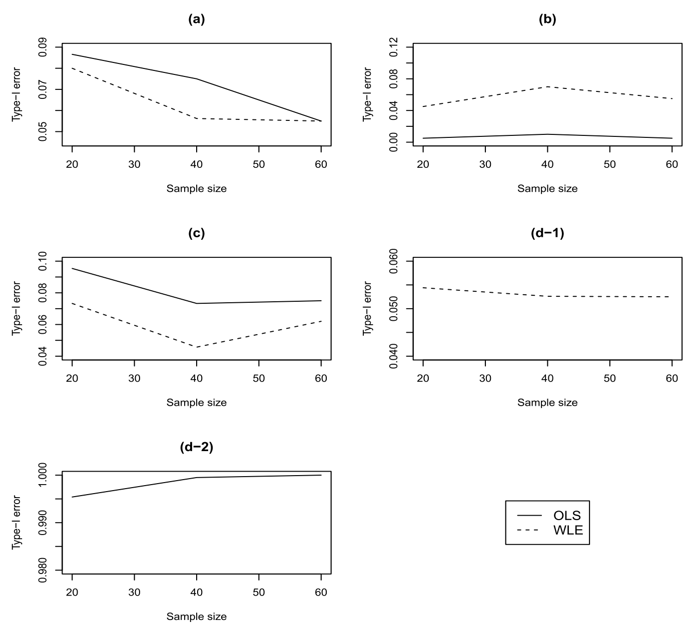

Figure 4. Type-I errors
(a) $\epsilon \sim N(0, 1)$ *(b)* $\epsilon \sim N(0, 1)$ *with one outlier* y_i *at* $i = n/2$
(c) $\epsilon \sim t(5)$ *(d)* $\epsilon \sim N(0, 1)$ w.p. 0.9 and $N(20, 1)$ w.p. 0.1.

For the alternative model with two change points, we study two different scenarios. First, we consider a design

$$y_i = \begin{cases} 5 + 1.2x_i + \epsilon_i, & x_i \le 20, \\ 21 + 0.5x_i + \epsilon_i, & 20 < x_i \le 40, \\ 49 - 0.3x_i + \epsilon_i, & x_i > 40, \end{cases}$$

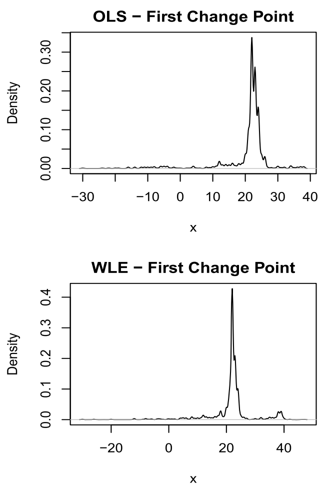

Figure 5. Density of the estimated change points by the statistics F_{max} and F_{wlemax} for $\epsilon \sim N(0,1)$.

where $i = 1, \ldots, 60$, $\epsilon_i \sim N(0, 0.25)$ and the covariate x is generated using $x_i = 4.5 + 0.9x_{i-1} + v_i$ with $v_i \sim N(0, 10)$, similar to the data generation process used by [10] for two change points in a continuous-mean, constant-variance alternative hypothesis model. Figure 5 shows the densities of the estimated change points determined using each of the two statistics F_{max} (first row of plots) and F_{wlemax} (second row). Since the generated data include no abnormal points, the two approaches perform roughly equivalently, identifying the change points at 20 and 40 reasonably accurately. However, when we intentionally add two outliers to the generated data, as Figure 6 shows, the density of the estimated change points estimated by the test based on F_{wlemax} (second row of plots) has considerably smaller variation in the identified change points compared to its non-robust counterpart F_{max} (first row), supporting the contention that use of F_{wlemax} offers a test that is more robust against outlying observations.

For the second two change-point design, we consider $n = 60$, $x = 1, \ldots, n$, $\boldsymbol{\beta_2} = (20(1 - \beta_{21}), \beta_{21})$ with $\beta_{21} = 4$, and $\boldsymbol{\beta_3} = (\beta_{20} + 40(\beta_{21} - \beta_{31}), \beta_{31})$ using the same four different error distributions as for the one change design case considered earlier. Figures 7 and 8 present the proportion of cases for which the change points are correctly detected with respect to the different values of β_{31} for the various error distributions for each of the two change points, respectively. When the errors are normal or $t(5)$-distributed with no unusual points inserted, there is no significant difference between the performance of the tests based on the two F statistics. However, when both of the $15th$ and $35th$ points positioned close to the change points are replaced with outlying points or when the error distribution is contaminated (90% standard normal, 10% shifted normal), the test based on F_{max} fails entirely to detect the change points, with a detection rate of zero observed across the simulation. As we

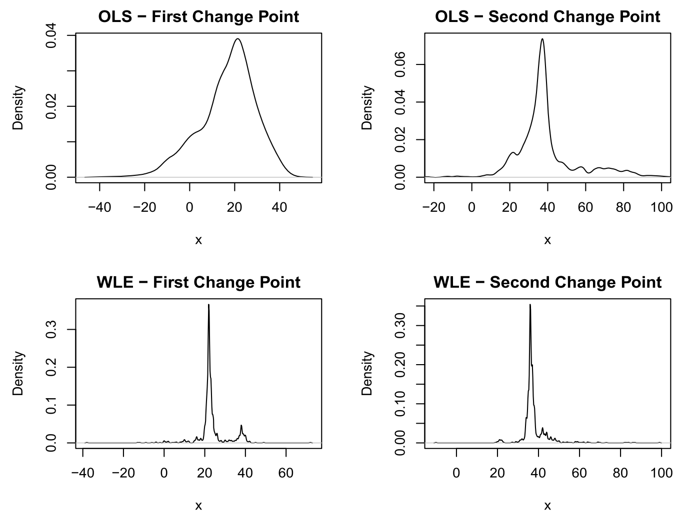

Figure 6. Density of the estimated change points by the statistics F_{max} and F_{wlemax} when there are two outliers inserted into the generated data set.

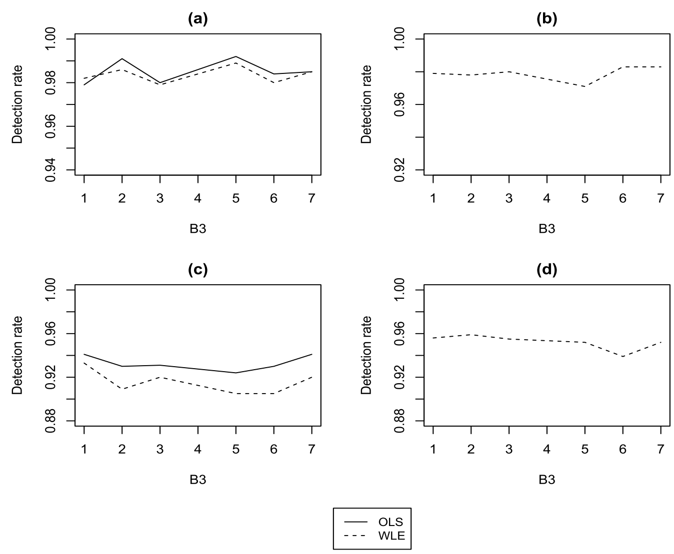

Figure 7. Detection rates for the first change point (a) $\epsilon \sim N(0,1)$ (b) $\epsilon \sim N(0,1)$ with two vertical outliers (c) $\epsilon \sim t(5)$ (d) $\epsilon \sim N(0,1)$ w.p. 0.9 and $N(20,1)$ w.p. 0.1.

showed for the first design, Figure 9 displays the densities of the estimated change points determined using each of the two statistics F_{max} (first row of plots) and F_{wlemax} (second row). In this case, even for standard normal error structure, the approach using F_{wlemax} results in smaller variation. Again, for the case of data generated with no abnormal data points, the two approaches perform roughly equivalently, identifying the change points at 20 and 40 reasonably accurately. However, when we intentionally add two outliers to the generated data, as Figure 10 shows, the density of

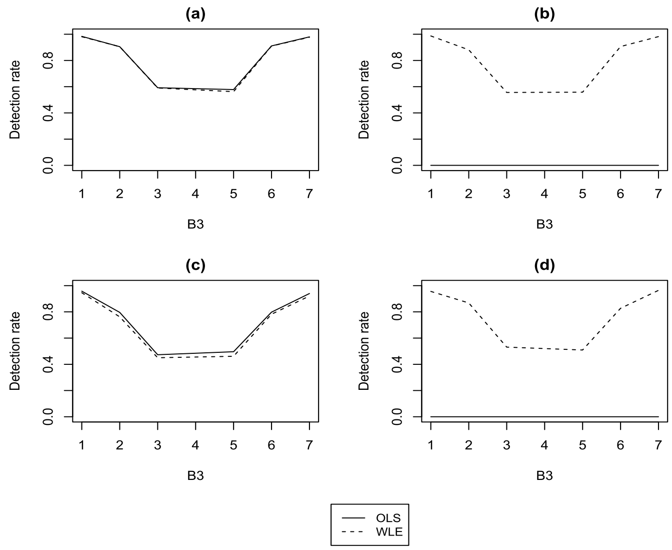

Figure 8. Detection rates for the second change point (a) $\epsilon \sim N(0,1)$ (b) $\epsilon \sim N(0,1)$ with two vertical outliers (c) $\epsilon \sim t(5)$ (d) $\epsilon \sim N(0,1)$ w.p. 0.9 and $N(20,1)$ w.p. 0.1.

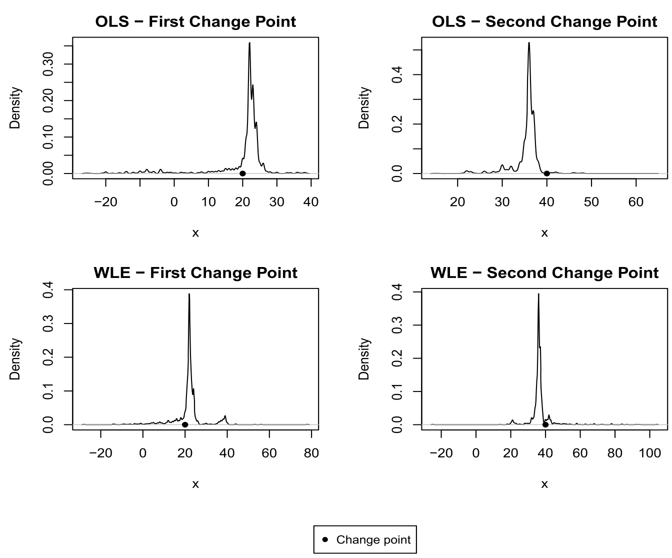

Figure 9. Density of the change points estimated by the statistics F_{max} and F_{wlemax}.

the estimated change points estimated by the test based on F_{wlemax} (second row of plots) again has considerably smaller variation in the identified change points compared to its non-robust counterpart F_{max} (first row), offering further evidence that use of F_{wlemax} offers a test that is more robust against outlying observations.

4.2 The stagnant band height data

This data set has also been studied by [10] and some references therein and examines how the stagnant surface layer

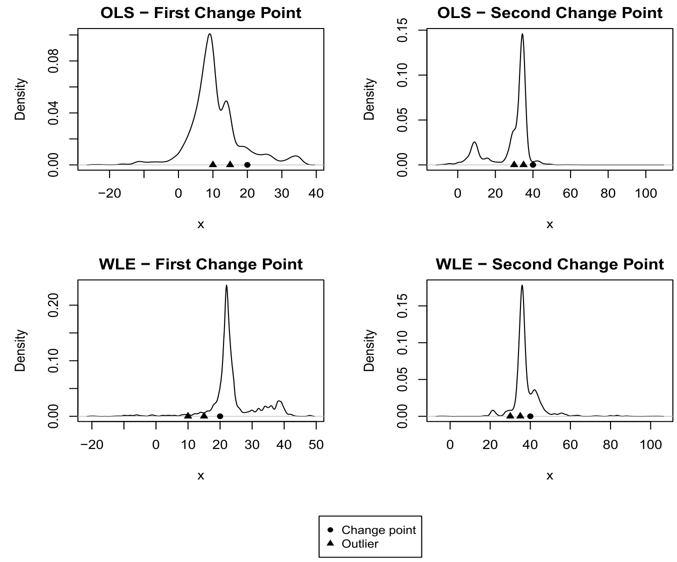

Figure 10. Density of the change points estimated by the statistics F_{max} and F_{wlemax} when there are two outliers.

height of water (stagnant band height) behaves as it flows down an incline at differing rates. The response variable is the logarithm of the band height in centimeters and the explanatory variable is the logarithm of the flow rate of water down an inclined furrow in grams per centimeter per second. Each of the procedures based on the statistics F_{max} and F_{wlemax} estimate 0.04 as the change point value, with parameters of the change point model estimated by each of the OLS and the WLE methods, with very similar models fit using each approach - see also Figure 11(a):

$$\hat{y}_{i(ols)} = \begin{cases} 0.5450 - 0.4217x_i & x_i \leq 0.04, \\ 0.5689 - 1.0201x_i & x_i > 0.04; \end{cases}$$

$$\hat{y}_{i(wle)} = \begin{cases} 0.5452 - 0.4212x_i & x_i \leq 0.04, \\ 0.5692 - 1.0202x_i & x_i > 0.04. \end{cases}$$

The original data set does not include any unusual observations, so to examine the robustness of both approaches, we intentionally replaced the third data point of the response with an outlying y-value when the independent variable is -0.25. Under this contamination of the data set, the approach based on the statistic F_{max} falsely estimates -0.25 as the change point – it is overly influenced by the outlying value – but the proposed approach using the F_{wlemax} statistic successfully estimates the change point occurring at the "correct" original value of $x = 0.04$. Even for a model built with a false change point imposed at $x = -0.25$, the model with parameters estimated using the WLE approach fits the data better. Figures 11(b) and (c) illustrate the change point models built by OLS and the WLE methods when the change is at 0.04 and at -0.25, respectively. Table 1 includes

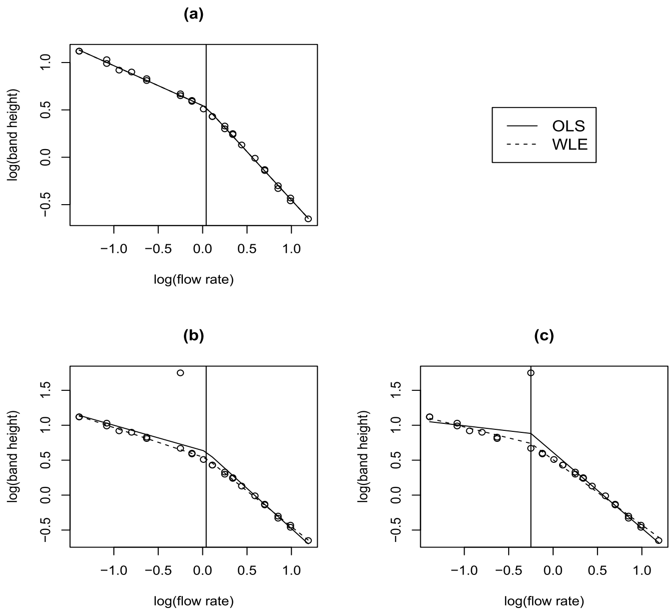

Figure 11. (a) Stagnant band height original data (b) Stagnant band height data with one outlying point (change point = 0.04) (c) Stagnant band height data with one outlying point (change point = -0.25).

Table 1. Scale estimates for Stagnant data

	$\hat{\sigma}_{WLE}$	$\hat{\sigma}_{OLS}$
Original data (no outliers)	0.01898	0.01952
Outlier included, $\gamma = 0.04$	0.01945	0.21710
Outlier included, $\gamma = -0.25$	0.03739	0.19455

the scale estimates for the various fitted models. The model standard errors are calculated as

$$\hat{\sigma}_{WLE} = \left\{ \frac{\sum_{i=1}^{n} \hat{\omega}(r_i) r_i(\hat{\boldsymbol{\beta}}_1; \hat{\boldsymbol{\beta}}_2)^2}{\sum_{i=1}^{n} \hat{\omega}(r_i) - 4} \right\}^{1/2},$$

$$\hat{\sigma}_{OLS} = \left\{ \frac{\sum_{i=1}^{n} r_i(\tilde{\boldsymbol{\beta}}_1; \tilde{\boldsymbol{\beta}}_2)^2}{n - 4} \right\}^{1/2},$$

where $\hat{\sigma}_{WLE}$ denotes the robust weighted likelihood standard error estimate and $\hat{\sigma}_{OLS}$ the usual least squares estimate. For the original data, the estimates $\hat{\sigma}_{WLE}$ and $\hat{\sigma}_{OLS}$ are very close to one another since there are no significant outliers. However, when there is an outlying observation, the robustness of $\hat{\sigma}_{WLE}$ is evident.

4.3 The plant data

This data set has been studied by [27] and [10] and describes three response attributes (denoted RWC, RKV and RKW) of a plant organ measured through time, which is the sole covariate. As pointed out by [10], two change points are indicated by biological theory for these data, for each of the three response variables. We fit the two change points model in (4) for each of these three attributes. Table 2 lists the change points estimated by the approaches based on F_{wlemax}, F_{max}, and by the analyses reported by [27] and [10] for each of the three attributes. The models fit by the WLE method are

Table 2. The estimated change points for the plant data

	RWC		RKV		RKW	
	CP-1	CP-2	CP-1	CP-2	CP-1	CP-2
F_{wlemax}	330	470	300	440	450	600
F_{max}	330	470	300	440	450	600
[27]	300	450	300	450	450	600
[10]	351.63	480.12	350.80	478.40	450.43	577.29

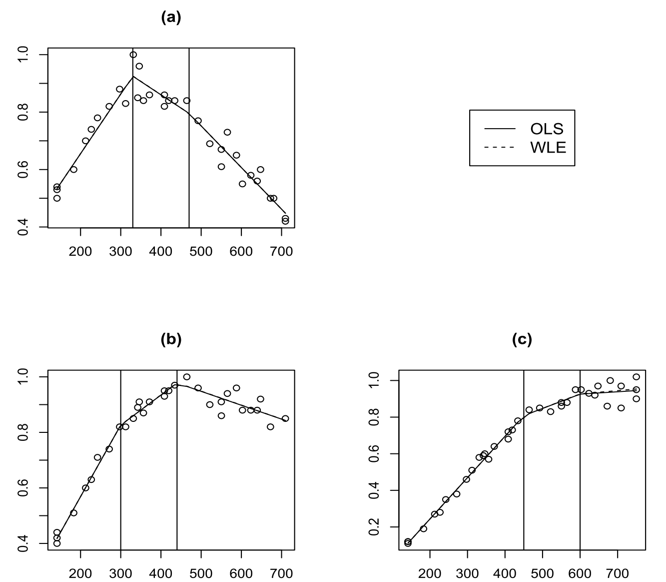

Figure 12. (a) RWC (b) RKV (c) RKW.

$$\hat{y}_{RWC(wle)} = \begin{cases} 0.2393 + 0.0021x_i, & 140 \le x_i \le 330, \\ 1.2293 - 0.0009x_i, & 330 < x_i \le 470, \\ 1.5113 - 0.0015x_i, & 470 < x_i \le 760; \end{cases}$$

$$\hat{y}_{RKV(wle)} = \begin{cases} 0.0534 + 0.0026x_i, & 140 \le x_i \le 300, \\ 0.5034 + 0.0011x_i, & 300 < x_i \le 440, \\ 1.2074 - 0.0005x_i, & 440 < x_i \le 710; \end{cases}$$

$$\hat{y}_{RKW(wle)} = \begin{cases} -0.2078 + 0.0023x_i, & 140 \le x_i \le 450, \\ 0.4672 + 0.0008x_i, & 450 < x_i \le 600, \\ 0.8272 + 0.0002x_i, & 600 < x_i \le 710. \end{cases}$$

As depicted by Figure 12, each of the approaches based on F_{wlemax} and F_{max} resulted in very similar fits (in most places, the dotted line, representing WLE is obscured by the solid line representing OLS, with only very slight variations visible).

In order to examine the robustness of the WLE and OLS approaches to change point estimation in this case, the response variable RKV was considered, with the 20th value of response variable for RKV replaced by an outlying value. As in the stagnant band height data example, the approach based on the classical F_{max} statistic is overly influenced by the outlier, and falsely estimates the position of the outlying data point as one of the change points suggesting the change points are $(400, 460)$, while the approach based on the robust F_{wlemax} statistic estimates the change points as $(310, 430)$ (as opposed to $(300, 440)$ for the original data – a mild change). Figure 13 illustrates the model fits for the change points estimated by the two competing approaches. For both of the change points, the models with the parameters estimated by the approach based on WLE clearly offer a better fit to the data, a finding that is also supported by the difference in scale estimates, $\hat{\sigma}_{WLE} = 0.03281$ and $\hat{\sigma}_{OLS} = 0.07949$ between the two methods in Figure 13(a), and $\hat{\sigma}_{WLE} = 0.0258$, $\hat{\sigma}_{OLS} = 0.0853$ between the two methods in Figure 13(b).

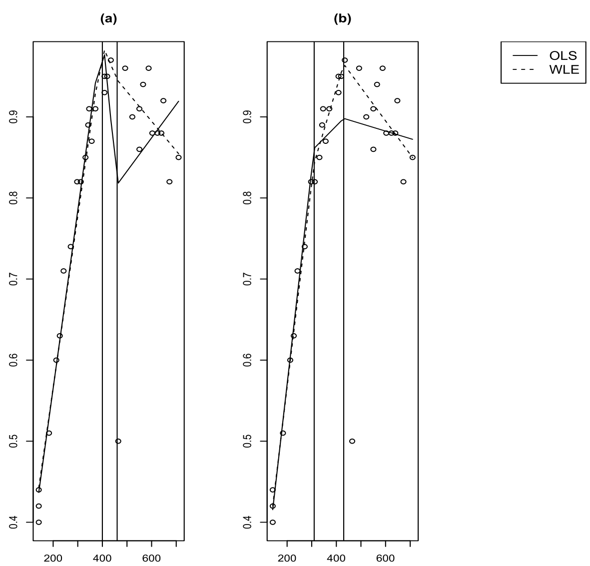

Figure 13. RKV with an outlying data; (a) The change points estimated using the F_{max} statistic are $(400, 460)$ (b) The change points estimated using the F_{wlemax} statistic are $(310, 430)$.

5. CONCLUSION

Detecting change points within linear regression models is essential for better model fit and interpretation when underlying theory or visualization suggest that "bent" lines reflect an appropriate model. However, the potential for outlying data points can result in poor fits to available data as these outlying points can be confused as change points in the relationship between the response and the covariates. In this paper, we propose a method based on grid search and hypothesis testing to estimate unknown change points in a linear regression context. The maximum of partial F_{wle} statistics, F_{wlemax}, based on weighted likelihood residuals is proposed as a test statistic to determine the appropriate position of change points. This approach offers a robust alternative to a test based on the maximum of classical F statistics, F_{max}, based on ordinary least squares residuals. We extensively studied the finite sample performance of the proposed approaches based on F_{wlemax} and F_{max} on simulated and real data for models incorporating one and two change points, respectively, using a bootstrap approach to characterise the sampling distributions of the competing test statistics. All of the considered numerical studies strongly supported the contention that the proposed statistic offers robustness against outlying observations.

We note that for the case of two change points, we proposed grid searching of each possible combinations of potential change points. We acknowledge that for situations for which there are more than two potential change points, this approach may not be practical. It may be possible to pursue grid-search methods that seek to sequentially determine multiple change points rather than searching over a high-dimensional grid, although such an investigation falls beyond the scope of the present investigation which was rather intended to establish the robustness properties of the WLE approach to change point estimation.

ACKNOWLEDGEMENT

Authors thank the anonymous reviewers and the associate editor for their comments and corrections which significantly improved the paper.

Received 10 April 2018

REFERENCES

[1] Agostinelli, C. (1998a). Inferenza statistica robusta basata sulla funzione di verosimiglianza pesata:alcuni sviluppi PhD thesis, Department of Statistics, University of Padova.

[2] Agostinelli, C. (1998b). Verosimiglianza pesata nel modello di regressione lineare. *Proceedings of the XXXIX Riunione scentifica of SIS*.

[3] Agostinelli, C. (2002). Robust model selection in regression via weighted likelihood methodology. *Statistics and Probability Letters* **56** 289–300. MR1892990

[4] Agostinelli, C. and Markatou, M. (1998). A one-step estimator for regression based on weighted likelihood reweighting scheme. *Statistics and Probability Letters* **37** 341–350. MR1624478

[5] Basu, A., Harris, I. R., Hjort, N. L. and Jones, M. C. (1998). Robust and efficient estimation by minimizing density based power divergence. *Biometrika* **85** 549–559. MR1665873

[6] Beckman, R. J. and Cook, R. D. (1979). Testing for two-phase regressions. *Technometrics* **21** 65–69.

[7] Bergfelt, D. R., Sego, L. H., Beg, M. A. and Ginter, O. J. (2003). Calculated follicle deviation using segmented regression for modeling diameter difference in cattle. *Theriogenology* **59** 1811–1825.

[8] Chen, J. (1998). Testing for a change point in linear regression models. *Communications in Statistics-Theory and Methods* **27** 2481–2493. MR1652636

[9] Chen, J. and Gupta, A. K. (2010). *Parametric Statistical Change Point Analysis and Finance With Applications to Genetics, Medicine*. Birkhäuser. MR3025631

[10] Chen, K. 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. MR2806617

[11] Efron, B. (1979). Bootstrap methods: Another look at the Jackknife. *The Annals of Statistics* **7** 1–26. MR0515681

[12] Feder, P. I. (1975). On asymptotic distribution theory in segmented regression problems-identified case. *The Annals of Statistics* **3** 49–83. MR0378267

[13] Ganocy, S. J. and Sun, J. (2015). Heteroscedastic change point analysis and application to footprint data. *Journal of Data Science* **13** 157–186.

[14] Hinkley, D. V. (1969). Inference about the intersection in two-phase regression. *Biometrika* **56** 495–504. MR0254946

[15] Hinkley, D. V. (1971). Inference in two-phase regression. *Journal of the American Statistical Association* **66** 736–743.

[16] Hofrichter, J. (2007). Change point detection in generalized linear models PhD thesis, Technische Universität Graz.

[17] Hudson, D. J. (1966). Fitting segmented curves whose join points have to be estimated. *Journal of the American Statistical Association* **61** 1097–1129. MR0210243

[18] Hušková, M. (2013). *Robustness and complex data structure* Robust change point analysis, 170–190. Springer. MR3135880

[19] Julious, A. (2001). Inference and estimation in a changepoint regression problem. *The Statistician* **50** 51–61. MR1821899

[20] Kim, H. J., Yu, B. and Feuer, E. J. (2008). Inference in segmented line regression:a simulation study. *Journal of Statistical Computation and Simulation* **78** 1087–1103. MR2518299

[21] Kim, H. J., Fay, M. P., Feuer, E. J. and Midthune, D. N. (2000). Permutation tests for change point regression with applications to cancer rates. *Statistics in Medicine* **19** 335–351.

[22] Lee., S. and Na, O. (2005). Test for parameter change based on the estimator minimizing density based divergence measures. *Annals of Institute of Statistical Mathematics* **57** 553–573. MR2206538

[23] Lerman, P. M. (1980). Fitting segmented regression models by grid search. *Applied Statistics* **29** 77–84.

[24] Lindsay, B. G. (1994). Efficiency versus Robustness: The case for minimum Hellinger Distance and related methods. *Annals of Statistics* **22** 1018–1114. MR1292557

[25] Markatou, M., Basu, A. and Lindsay, B. G. (1998). Weighted likelihood estimating equations with a bootstrap root search. *Journal of the American Statistical Association* **93** 740–750. MR1631378

[26] Martín, N. and Pardo, L. (2014). Comments on: Extensions of some classical methods in change point analysis. *Test* **23** 279–282. MR3210275

[27] Muggeo, V. M. R. (2003). Estimating regression models with unknown break-points. *Statistics in Medicine* **22** 3055–3071.

[28] Muggeo, V. M. R. (2016). Testing with a nuisance parameter present only under the alternative: a score based approach with application to segmented modelling. *Journal of Statistical Computation and Simulation* **86** 3059–3067. MR3523161

[29] Niu, Y. S., Hao, N., and Zhang, H. (2016). Multiple change point detection: a selective oveview. *Statistical Science* **31** 611–623. MR3598742

[30] Nosek, K. (2010). Schwarz Information Criterion Based Tests for a Change-Point in Regression Models. *Statistical Papers* **51** 915–929. MR2749007

[31] Osario, F. and Galea, M. (2006). Detection of a change-point in student-t linear regression models. *Statistical Papers* **47** 31–48. MR2253297

[32] Pauler, D. K. and Finkelstein, D. M. (2002). Predicting time to prostate cancer recurrence based on joint models for non-linear longitudinal biomarkers and event time outcomes. *Statistics in Medicine* **21** 3897–3911.

[33] Reeves, J., Chen, J., Wang, X. L., Lund, R. and Lu, Q. (2007). A review and comparison of Changepoint detection techniques for climate data. *Journal of Applied Meteorology and Climatology* **46** 900–915.

[34] Robison, D. E. (1964). Estimates for the points of intersection of two polynomial regressions. *Journal of the American Statistical Association* **59** 214–224. MR0182095

[35] Worsley, K. J. (1983). Testing for a two-phase multiple regression. *Technometrics* **25** 35–42. MR0694210

[36] Zhang, N. R. and Siegmund, D. O. (2007). A modified Bayes information criterion with applications to the analysis of comparative genomic hybridization data. *Biometrics* **63** 22–32. MR2345571

Aylin Alin
Department of Statistics
Dokuz Eylul University
Izmir
Turkey
E-mail address: aylin.alin@deu.edu.tr

Ufuk Beyaztas
Department of Statistics
Bartin University
Bartin
Turkey
E-mail address: ufuk.beyaztas@gmail.com

Michael A. Martin
Research School of Finance
Actuarial Studies and Statistics
Australian National University
Australia
E-mail address: Michael.Martin@anu.edu.au