Vol. 00, No. 00, April 2015, 1–21

Approximations to the p-values of tests for a change-point under non-standard conditions.

(*Received 00 Month 20XX; final version received 00 Month 20XX*)

Three test statistics for a change-point in a linear model, variants of those considered by Andrews and Ploberger [1]: the sup-LR statistic; a weighted average of the exponential of LR-statistics and a weighted average of LR-statistics, are studied. Critical values for the statistics with time trend regressors, obtained via simulation, are found to vary considerably, depending on conditions on the error terms. The performance of the parametric bootstrap in approximating p-values of the distributions is assessed in a simulation study. A sample approximation to asymptotic analytical expressions extending those of Kim and Siegmund [2] in the case of the sup-LR test is also assessed. The approximations and bootstrap are applied to the Quandt data [3] and real data concerning a change-point in oxygen uptake during incremental exercise testing and the bootstrap gives reasonable results.

Keywords: change-point; regression; likelihood ratio test; Lagrange multiplier test; sup-F test; time trend regressor; correlated errors; heteroskedastic errors; simulation; bootstrap;

AMS Subject Classification: 62F05; 62F10; 62F40

Word count: 7778

1. Introduction

Many interesting change-point problems in medicine, physiology, economics and other disciplines involve testing the constancy of regression relations over time. Change-point methodology is also used in disease surveillance and monitoring systems, where there have been rapid developments in the last three decades, as described in Höhle and Mazick [4]. In testing for a change-point in a linear model, it is customary to first propose a test for a change-point at a particular point e.g. a likelihood ratio, Wald or Lagrange multiplier test. The point at which the change occurs is then allowed to vary and the sup-likelihood ratio (LR) statistic, which is the sup of LR statistic at each possible change-point, was first proposed by Quandt [3] to test for a change-point. The sup-Wald (equivalently sup-F) or sup-Lagrange multiplier statistics are also commonly used. There is a large literature on the behaviour of the sup-F and sup-LR statistics. Kim and Siegmund [2] derived an approximation to the asymptotic distribution of the sup-LR statistic in the time trend model where errors are assumed to be i.i.d and there is no change in variance. Andrews [5] considered the sup-Wald and sup-LR tests, and derived their common asymptotic distribution under general conditions i.e. correlation structure of the errors, number of regressors, pure structural change or just structural change. Bai and Perron [6] also explain how their sup-F statistic can be modified to account for heteroscedasticity and serially correlated errors. However, in the final sentence of their paper they state 'the limiting distributions of the various tests for structural change remain to be studied in the presence of trending regressors'. They also state [7, p. 14] that although asymptotic distributions differ in the case of non-trending and trending regressors, the two cases

are fairly similar, especially in the tail where critical values are obtained. Hence, one can safely use the same critical values i.e. those given in [6]. Kim and Siegmund [2] however, found this not to be the case, with large discrepancies in the p-values for the two cases. Andrews and Ploberger [1] show the classical optimality properties of sup-Lagrange multiplier, sup-Wald and sup-LR tests do not hold for change-point problems. They proposed three test statistics: the sup-F statistic which is the sup of F statistics at each possible change-point; a weighted average of the exponential of the F-statistics 'Exp' and a weighted average of F-statistics 'Ave'. They derived analytical expressions for the distributions and evaluated these by numerical methods. However, Andrews, Lee and Ploberger [8, p.31] state the asymptotic critical values presented are not appropriate for the time trend model.

Here the change-point problem is examined for linear regression with time trend where both the regression coefficients and error variance structure may change simultaneously and the errors may be correlated. For purposes of clarity, results are derived for these somewhat simple models but the extension of the results to more complicated models e.g. more than one regressor, ARMA errors, is straightforward.

In Section 2, definitions and commonly used approaches to detect change-points are introduced, in particular the three test statistics of Andrew and Ploberger [1], with sup-F replaced by sup-LR. A sample approximation to an asymptotic analytical expression for the p-value, extending that of Kim and Siegmund [2], is constructed in the case of the sup-LR test. Section 3 describes two simulation studies, the first in which empirical critical values of the three test statistics are derived (under normality assumptions on the errors). These critical values are compared to asymptotic approximations in the literature [1, 2, 6], where the latter two are based on the sup-F statistic. In a separate simulation study, the nonparametric bootstrap is used to approximate p-values of the distributions of the test statistics under non-standard conditions. Bootstrap and analytic p-values are compared to true probabilities. In both these approximations the p-value depends on the data.

In Section 4, the approximations are applied to real data concerning a change-point in oxygen uptake during incremental exercise testing give in Kelly [9], as well as to the classical data set of Quandt [3]. Computational issues are discussed in Section 5 and the R code used to estimate the change-point is supplied in an Appendix. Conclusions are given in Section 6.

2. Definitions

Change-point data are frequently correlated over time and often also exhibit a change in variance at the change-point. We therefore consider the general linear regression change-point model

$$y_i = \alpha_0 + \beta_0 x_i + u_i, \ i = 1, \ldots, \tau \tag{2.1a}$$

$$y_i = \alpha_1 + \beta_1 x_i + u_i, \ i = \tau + 1, \ldots, m \tag{2.1b}$$

where

$$u_i = \rho u_{i-1} + \epsilon_i \text{ and } u_1 = \epsilon_1 \tag{2.1c}$$

and the ϵ_i are i.i.d. $N(0, \sigma_1^2)$ for $i = 1, \ldots, \tau$ and $N(0, \sigma_2^2)$ for $i = \tau + 1, \ldots, m$ where τ denotes the unknown change-point. $|\rho|$ is assumed to be < 1 so that the u_i form a stationary process AR(1) process. The model can also be written in matrix form in the

obvious way as

$$Y = X\beta + u, \quad \text{Cov}(Y) = V \tag{2.2}$$

The relationship between V and the model parameters in equation (2.1) is given in Appendix A for the case of AR(1) errors. In what follows, the x's are assumed equally spaced and V can in fact be any covariance matrix, such that if it changes at the change-point, a parametric form needs to be provided. The simulation study presented is restricted to AR(1) errors, as are the numerical examples.
Kim and Siegmund [2] examined this model under the assumptions that the covariate x was of the form $x_j = j/m$, no correlation and no change in variance i.e. $\sigma_1^2 = \sigma_2^2$. They derived asymptotic expressions for the p-value of the likelihood ratio (LR) test of no change, $H_0 : \beta_0 = \beta_1 = \beta$ and $\alpha_0 = \alpha_1 = \alpha$ against the alternative H_A: there exists a τ $(1 \leq \tau < m)$ such that $\beta_0 \neq \beta_1$ or $\alpha_0 \neq \alpha_1$ i.e. there is a change in either the intercept or slope. Here their results are extended to model (2.2). We note compared to simulation, analytic approximations can be calculated more efficiently and importantly, help to understand properties of the test statistic.
Methods exist for much more general settings on the covariance structure [10], but there are particular applications, in which assumptions such as those above in model (2.1) are satisfied, to which the approximation developed here can be reliably applied.

2.1. *Kim and Siegmund [2] approximation to the p-value of the sup-LR test*

This section gives the form of the approximation to the p-value of the LR statistic for a change-point in simple linear regression with i.i.d. errors derived by Kim [11] in an unpublished thesis and reported for the case of equally spaced x's in Kim and Siegmund [2].
Firstly, consider the model given in equation (2.1) where $V = \sigma^2 I$, i.e. there is no correlation and no change in variance $\sigma_1^2 = \sigma_2^2$. We introduce the following notation: let $A_1' = (1, -1, 0)$, $A_2' = (0, 1, 0, -1)$, $Y' = (y_1, \ldots, y_m)$

$$X_{1i} = \begin{bmatrix} 1 & 0 & x_1 \\ \vdots & \vdots & \vdots \\ 1 & 0 & x_i \\ 0 & 1 & x_{i+1} \\ \vdots & \vdots & \vdots \\ 0 & 1 & x_m \end{bmatrix}, X_{2i} = \begin{bmatrix} 1 & x_1 & 0 & 0 \\ \vdots & \vdots & \vdots & \vdots \\ 1 & x_i & 0 & 0 \\ 0 & 0 & 1 & x_{i+1} \\ \vdots & \vdots & \vdots & \vdots \\ 0 & 0 & 1 & x_m \end{bmatrix}$$

and for $\mu = 1$ or 2 let

$$\begin{aligned} U_{\mu,m}(i) = {} & A_\mu'(X_{\mu,i}'X_{\mu,i})^{-1}X_{\mu,i}'Y \\ & /[A_\mu'(X_{\mu,i}'X_{\mu,i})^{-1}A_\mu]^{1/2} \end{aligned} \tag{2.3}$$

Note $U_{1,m}(i)$ is the test statistic for a change of intercept only at i. For $\mu = 1$ or 2, $\lambda = 1$ or 2, let

$$C_{\lambda,\mu}(i,k) = \sigma^{-2}\text{cov}[U_{\lambda,m}(i), U_{\mu,m}(k)] \tag{2.4}$$

THEOREM 2.1 $-2\log$*(likelihood ratio) statistic for* H_0 *versus* H_A *(generalized slightly) is*

$$\hat{\sigma}^{-2} \max_{m_0 \le i \le m_1} [U_{1,m}^2(i) + U_{2,m}^2(i)] \tag{2.5}$$

where $2 \le m_0 < m_1 \le m-2$ *and* $\hat{\sigma}^2$ *is the mle of* σ^2 *under the null model.*

THEOREM 2.2 *Kim [11] shows that taking limits in equation (2.4) gives the covariances of the process*

$$(U_1(t), U_2(t)) = \lim_{m\to\infty} (U_{1,m}([mt]), U_{2,m}([mt])),\ 0 < t < 1.$$

These covariances will be denoted by

$$\lambda_{11}(t,s), \lambda_{12}(t,s), \lambda_{21}(t,s), \lambda_{22}(t,s).$$

THEOREM 2.3 *Kim and Siegmund [2] shows the probability* p_2 *that the random variable in equation (2.5) exceeds* b^2 *is given by, in the case* σ^2 *unknown for* $V = \sigma^2 I$ *and* $x_i = i/m$, $(i = 1, \ldots, m)$

$$\begin{aligned} p_2 &\approx (2\pi)^{-1} b^2 (1 - b^2/m)^{(m-6)/2} \\ &\times \int_{t_0}^{t_1} \int_0^{2\pi} \mu(t,\theta)\nu\left[\left(\frac{2c^2\mu(t,\theta)}{(1-c^2)}\right)^{(0.5)}\right] d\theta dt \end{aligned} \tag{2.6}$$

where $c = b/\sqrt{m}$,

$$\nu(x) = 2x^{-2} exp\left[-2\Sigma_{n=1}^{\infty} n^{-1}\Phi(-1/(2x\sqrt{n}))\right], x > 0, \tag{2.7}$$

Φ *denotes the standard normal distribution function, and*

$$\begin{aligned} \mu(t,\theta) &= .5 + [1 - 6t(1-t)]\sin^2(\theta) - \sqrt{3}(2t-1)\cos(\theta)\sin(\theta) \\ &/\ [t(1-t)(1-3t(1-t))] \end{aligned} \tag{2.8}$$

THEOREM 2.4 *The probability* p_2 *that the random variable in equation (2.5) exceeds* b^2

for the case $Cov(Y) = V$, is as in Theorem 2.3 with $\mu(t,\theta)$ replaced by

$$\mu(t,\theta) = \left[\frac{-d}{ds}\lambda_{11}(t,s)|_{s=t} + sin^2(\theta)A_1(t)\right] - [cos(\theta) \times sin(\theta)A_2(t)] \tag{2.9a}$$

where

$$A_1(t) = -\left[\frac{d}{ds}\lambda_{22}(t,s)|_{s=t} - \frac{d}{ds}\lambda_{11}(t,s)|_{s=t}\right] \text{ and} \tag{2.9b}$$

$$A_2(t) = \left[\frac{d}{ds}\lambda_{12}(t,s)|_{s=t} + \frac{d}{ds}\lambda_{21}(t,s)|_{s=t}\right] \tag{2.9c}$$

with $\lambda_{ij}(t,s)$ defined in Theorem 2.2.
Proof.
This follows from the fact that for arbitrary matrices C, D with $Z_1 = CY$ and $Z_2 = DY$, $\text{cov}(Z_1, Z_2) = CVD^T$ *and applying it to equation (2.3).*

When variance is unknown but common in the two regimes, the LR test is given by

$$-2\log(LR) = -m \times \log(1 - U/Q) = -m \times \log(1 - V^*)$$

where Q is the residual sum of squares under H_0, and $U = \max_k(U_k)$ where U_k is the (reduction in) residual sum of squares under H_A: change-point at k, that is of course smaller than Q. Kim and Siegmund's [2] expression for the LR test can then be expressed in terms of the F-statistic, since Worsley [12] showed the LR test was equivalent to the maximum F statistic via the relationship

$$\begin{aligned} F &= \frac{V^*/p}{(1-V^*)/(m-2p)} \\ \Rightarrow V^* &= \frac{pF}{m-2p+pF} \\ \Rightarrow 1 - V^* &= \frac{m-2p}{m-2p+pF} \end{aligned}$$

and thus

$$-2\log(LR) = -m\log(1-V^*) = -m\log(\frac{m-2p}{m-2p+pF})$$

The statistic used by Kim and Siegmund [2], denoted by KS, for which Tables are given is $KS^2/m = V^*$.

2.2. *Approximation to the p-value of the sup-LR test: non-constant variance and correlated errors*

We now wish to estimate the p-value of the LR statistic for a change-point in simple linear regression under more general conditions on the error terms i.e. non-constant variance and correlated observations. We do this by adapting the equations of section (2.1).
We consider model (2.1) where covariances are not necessarily 0 and σ_1^2 not necessarily equal to σ_2^2. We denote the variance-covariance matrix of the vector of observations y by V.

THEOREM 2.5 *The likelihood ratio statistic under the assumption of an arbitrary variance-covariance V where V_i denotes its form assuming a change-point at i is given by equation(2.5) with $U_{\mu,m}(i)$, $\mu = 1$ or 2, given by*

$$U_{\mu,m}(i) = A'_\mu(X'_{\mu,i}\hat{V_i}^{-1}X_{\mu,i})^{-1}X'_{\mu,i}\hat{V_i}^{-1}Y \\ /[A'_\mu(X'_{\mu,i}\hat{V_i}^{-1}X_{\mu,i})^{-1}A_\mu]^{1/2}. \tag{2.10}$$

It is assumed the form V_i remains constant over i e.g. V_i corresponding to an AR(1) process for the error terms with a change of variance at the ith observation.
Proof.
It can be shown that it is possible to find a unique nonsingular symmetric matrix P such that

$$P'P = PP = P^2 = V.$$

Note $(P^{-1})' = P^{-1}$. Writing $f = P^{-1}u$ then $f \sim N(0, I)$. If we pre-multiply equation (2.2) by P^{-1} we obtain a new model

$$Z = P^{-1}Y = P^{-1}X\beta + P^{-1}u = Q\beta + f. \tag{2.11}$$

We apply equations (2.5-2.9) to this new model with $X_{\mu,i}$ replaced by ${P_i}^{-1}X_{\mu,i}$ and Y replaced by ${P_i}^{-1}Y$ in (2.5) where P_i is P assuming the change-point is at i. Note equation (2.5) then involves V_i^{-1} only (not ${P_i}^{-1}$) where V_i corresponds to V assuming a change-point at i. The likelihood under the null and an alternative assuming a change-point at i, is that of a multivariate normal with variance-covariance matrix V and V_i respectively. Parameters are then estimated by maximum likelihood and $\hat{V_i}$ denotes the mle of V_i assuming a change-point at i and is substituted into equation (2.5) and the likelihood ratio statistic calculated.

THEOREM 2.6 *A consistent approximation to the p-value of the statistic given by equation (2.10) is given by replacing the covariances in equation (2.4) by estimated values and the derivatives in equation (2.9) by discrete sample approximations as follows:*

$$\text{Let } t = i/m, \text{ then}$$
$$\frac{d}{ds}\lambda_{11}(t,s)|_{s=t} \sim \frac{(C_{11}(i,i+1) - C_{11}(i,i))}{(x_{i+1} - x_i)} \tag{2.12}$$

where C is given in equation (2.4) with $\hat{V_i}$ replacing V_i, $i = 1, \ldots, m$, and with similar expressions for $\frac{d}{ds}\lambda_{12}(t,s)|_{s=t}$, $\frac{d}{ds}\lambda_{21}(t,s)|_{s=t}$ and $\frac{d}{ds}\lambda_{22}(t,s)|_{s=t}$.
The integral in equation (2.6) is then evaluated using Riemann sums over the intervals $i/m, (i+1)/m$, $i = 0, \ldots, m-1$. Note as before $\hat{\sigma^2}$ is the mle of σ^2 under the null model of no change.
Proof
The approximate p-value derived is consistent since $\hat{V_i}$ is consistent for V_i, the discrete approximations (2.12) are consistent for the derivatives in equation (2.9) and since Riemann sums approximate an integral. The degrees of freedom of (2.6) i.e. (m-6) is adjusted to $(m - 5 - s)$ where s is the number of estimated parameters in V_i using arguments as in Kim [11].

2.3. *Andrews and Ploberger tests for a change-point*

Consider the model given by equations (2.1a)-(2.1c). Let $F_m(\tau)$ denote a Wald, Lagrange multiplier or LR statistic of the hypothesis of no structural change for given τ. When τ is known only to lie in the range $[k_1, k_2]$ the Quandt or "sup" test statistic is

$$\text{Sup } F_m = \sup_{k_1 \leq \tau \leq k_2} F_m(\tau) \tag{2.13}$$

The Andrews and Ploberger [1] "Exp" and "Ave" tests are

$$\text{Exp } F_m = \ln \left(\frac{1}{k_2 - k_1 + 1} \sum_{\tau=k_1}^{k_2} \exp(1/2 F_m(\tau)) \right) \text{ and} \tag{2.14}$$

$$\text{Ave } F_m = \frac{1}{k_2 - k_1 + 1} \sum_{\tau=k_1}^{k_2} F_m(\tau) \tag{2.15}$$

where $m > k_2 > k_1 > 0$ and both k_1 and k_2 known. The Exp F_m test is the optimal test for testing against distant alternatives in the class of all tests of asymptotic significance level α. The Ave F_m is designed for alternatives close to the null hypothesis. Andrews and Ploberger [1] state they show little difference in power. These test statistics have a Bayesian posterior odds ratio form and are easier to compute than an actual Bayesian posterior odds ratio. The sup F_m test is designed for distant alternatives but is of a more extreme form than the optimal exponential test. It has greater power when the true change-point is in the middle of the series. Thus the Kim and Siegmund [2] approximation to the p-value will not necessarily be good if the change-point occurs near the extremes of the distribution. All three test statistics involve a choice of $\pi_1 = k_1/m$ and $\pi_2 = k_2/m$ and asymptotic distributions depend on this choice as well as m. This choice can be represented by a single parameter π_0 given by $\frac{1}{1+\sqrt{\lambda_0}}$, where $\lambda_0 = \frac{\pi_2(1-\pi_1)}{\pi_1(1-\pi_2)}$. Note when the range $[k_1, k_2]$ is symmetric in the sample, $\pi_0 = \pi_1 = 1 - \pi_2$. The tables of Hansen [13] give p-values over many possible choices of π_0 but a change in π_0 from 0.02 to 0.15 does not change the critical values by much as can be seen in the tables of Andrews and Ploberger [1]. These authors suggest taking $\pi_0 = 0.02$ when there is no prior knowledge as to in which interval the change-point occurs (if it occurs). However, a choice of $\pi_0 = 0.02$ may not be practicable for small sample sizes and the number of unknown parameters, p, to be estimated also needs to be considered in this choice i.e. $\pi_0 \times m > p$. Bai and Perron [7, p.13] address this issue. They state when allowing different variances across segments or serial correlation, a higher trimming value should be used. If estimates of variances etc. are imprecise, the tests will show size distortions and be very variable, as one is estimating various quantities using very few observations and thus they also considered $\pi_0 = [0.1, 0.25]$. Zeileis *et al.* [14] constructed an R add-on package `strucchange`, obtainable from the Comprehensive R Archive Network (CRAN) where the R software is that of the R core team [15]. The package computes the three F statistics of Andrews and Ploberger and p-values are computed based on Hansen [13]. However, these p-values are not correct for trending regressors. Three test statistics, the sup-LR, exp(LR), ave(LR), given by equations (2.13)-(2.15) with the LR statistic replacing the F statistic (this will not change asymptotic properties by [5]) are considered here in a simulation study.

2.4. *Confidence intervals*

Using the p-values from section (2.1) or (2.2), approximate confidence intervals for the change-point can be found using Worsley's method [16]. This includes j in a $(1-\alpha)$ confidence interval if the likelihood ratio tests for no change in $[0, j-1]$ and in $[j, m]$ are both accepted at significance levels greater than $1-(1-\alpha)^{0.5} \approx \alpha/2$ i.e. the p-values $\geq \alpha/2$.
Kim and Siegmund [2, Section 4] describe a method for the construction of confidence intervals based on the likelihood ratio test for i.i.d. errors in their equations (4.2) and (4.3), but this cannot be readily modified to the non-i.i.d. case being considered here.
Multiple change-points are considered in our examples. Briefly, for arbitrary regressors, a simple method of estimating all the change-points in a sequence is the binary segmentation procedure. The sequence of observations is tested for a change, and if the change is judged significant at a predetermined level α then the two subsequences before and after the change-point are tested in the same way as above [16]. Alternatively, using a Schwartz criterion, Lee [17] for exponential families, proposed a consistent estimator k^* of the number of change-points as that k which maximized $\log g(y; \theta_k, j(k), k) - kd(n)$, where g is the joint likelihood of the data, $j(k)$ is the position of the k change points, $d(n)$ is a penalty term greater than $2(1+\epsilon_0)\log(n)$ for some known $\epsilon_0 > 0$ and $d_0^*(n)$ is the restriction for the minimum distance between successive true change locations, which satisfies $\frac{d(n)}{d_0^*(n)} \to 0$ as $n \to \infty$.

3. Simulation Study

A simulation study was conducted to check the accuracy of the asymptotic approximations to the p-value of the sup-LR test. In 10,000 repetition Monte Carlo experiments, sample sizes m of 20 and 40 were considered with trimming fraction $\pi_0 = 0.1$ to enable comparisons with Kim and Siegmund [2] and such that $\pi_0 \times m > p$. In addition, in many applications such as in economics, sample sizes may be small [7]. A time trend regressor was chosen $x_i = i/m$ $(i = 1, \ldots, m)$ and normally correlated AR(1) errors with AR(1) parameter $\rho = 0, 0.1, 0.4$ and 0.7 and $\sigma^2 = 1$.
Models were fitted assuming σ unknown and the mle's of the parameters and estimated log-likelihoods obtained for each simulation using the function `elliptic` from J. Lindsey's R public library called `growth`, available at `http://www.commanster.eu/rcode.html`. The R code [15] used for a single change-point is explained and given in the Appendix. The simulations were done on a 384core HPC cluster.

3.1. *Approximations to the p-value of the sup-LR test of section 2.2*

In each experiment, the 90th, 95th and 99th percentiles of the distribution of the statistics in equation (2.5) were estimated and the p-values of equations (2.6), with (2.8) or (2.9) evaluated at the estimated percentiles. For the case of $\rho \neq 0$ and a change in variance, the approximate p-value involves $\hat{V}_i$ (see section 2.2) in which the auto-correlation parameter ρ and variance parameters σ_1 and σ_2 are estimated from the data, assuming the change-point is at i. The results are shown in Table 1.

3.2. *Bootstrap approximations to the p-value*

The bootstrap offers an alternative method for computing the p-values of test statistics for a change-point and has been documented by several authors for the case of the sup-LR test with i.i.d. errors [18,19]. The more general change-point problem considered here can be classified as a non-stationary time series for which bootstrap methods are the subject of on-going research [20,21]. One such method, now known as the residual-based sieve bootstrap procedure [22] and later modified Smeekes [21], is used here. The basic idea is to compute marginal residuals (innovations) from which a bootstrap sample can be constructed. 5,000 simulations were carried out for data sets as above with sample size 40. For each configuration of data, the 90th, 95th and 99th percentiles of the distribution of the test statistics of section (2.3) were calculated and are shown in Tables 2 and 3. Then, separately, for each data configuration, 200 simulations were carried out with 500 bootstrap iterations in each. In each simulation, it was noted whether the observed test statistic exceeded the relevant percentile of the bootstrap distribution. The bootstrap p-value is then the proportion of times out of 200 that the statistic exceeded the estimated percentile. For example, for the sup-LR statistic, a linear regression is fit to the data assuming AR(1) errors, and the observed value of the statistic calculated. Fitted values are obtained. The marginal residuals are calculated. These are resampled, recursive residuals computed and a bootstrap sample obtained by adding these residuals to the fitted values. A sup-LR statistic is calculated from the bootstrap sample. The resampling is repeated B=500 times and a bootstrap percentile obtained based on the 500 values. It is then recorded if the observed statistic exceeds this percentile. The proportion of times this occurs in 200 simulations is the bootstrap p-value and results are shown in Tables 2 and 3. The study was limited to 200 simulations due to computation time required. Each simulation in Table 2 takes approximately 144 hours while those in Table 3 take 192 hours. Therefore, they were run in parallel on a cluster.

3.3. *Results*

We refer to Table 1 for a summary of the main results concerning analytic approximations. Note the statistic tabulated is as in Kim and Siegmund [2] to facilitate comparison with that study. The table shows that in the independence and no change in variance case, the accuracy of the approximation of equation (2.6) with (2.9)-(2.12) is as good as that of equation (2.8) (derived in [2]) where the form of $\mu(t, \theta)$ is assumed known. For $m = 40$ the former gave values 0.11, 0.05 and 0.01, very close to the true p-values. However, the approximation (2.8) significantly underestimates the true p-values when there is autocorrelation. This becomes more marked as the auto-correlation magnitude goes up. The table also shows that the percentiles of the square root of the sup-LR statistic increase as the correlation parameter ρ gets larger. This is to be expected, as with correlated data we have less information and therefore less power. Similar results have been found for the sup-LR test for a change in mean Kim [23]. Table 1 also shows for correlated data, agreement with true values using the method of section 2.2 is good especially for $m = 40$ and smaller probabilities. Larger probabilities are over-estimated particularly for large values of ρ. Kim and Siegmund [2] in a simulation study for the i.i.d. case, also obtained less accurate approximations for larger probabilities than smaller ones. We note the approximations are meant for tail probabilities and assume small probabilities. Moreover, smaller probabilities are more important in hypothesis testing and the construction of confidence regions. The general overestimation observed by Kim [23, Remark 3] was not replicated here.

There is some discrepancy between critical values obtained by analytic methods by different authors. For example, for $\alpha = 0.1$, $m = 20$, $x_i = i$ Worsley's [12] critical value

Table 1. Accuracy of the approximations to the p-value of the sup-LR statistic for $y_i = x_i + u_i, i = 1, \ldots, m$, $u_i \sim$ normal AR(1) with autocorrelation parameter ρ, $\sigma^2 = 1.0$, $m_0 = 0.1m, m_1 = 0.9m, x_i = i/m$.

m	True tail Probability	Estimated tail percentile	Approximation KS[a]	Approximation using method of section (2.2)[b]
$\rho = 0.0$				
20	0.10	2.967	0.096	0.105
	0.05	3.133	0.046	0.050
	0.01	3.433	0.008	0.009
40	0.10	3.104	0.110	0.115
	0.05	3.308	0.053	0.054
	0.01	3.695	0.010	0.010
$\rho = 0.1$				
40	0.10	2.539	0.030	0.119
	0.05	3.264	0.004	0.056
	0.01	5.293	0.000	0.011
$\rho = 0.4$				
40	0.10	4.252	0.000	0.121
	0.05	5.339	0.000	0.057
	0.01	9.255	0.000	0.011
$\rho = 0.7$				
40	0.10	6.671	0.000	0.152
	0.05	8.297	0.000	0.071
	0.01	13.372	0.000	0.013
$\rho = 0.0$	change in variance			
20	0.10	2.967	0.096	0.105
	0.05	3.133	0.046	0.050
	0.01	3.433	0.008	0.009
40	0.10	3.104	0.110	0.115
	0.05	3.308	0.053	0.054
	0.01	3.695	0.010	0.010
$\rho = 0.1$				
40	0.10	2.539	0.030	0.119
	0.05	3.264	0.004	0.056
	0.01	5.293	0.000	0.011
$\rho = 0.4$				
40	0.10	4.252	0.000	0.121
	0.05	5.339	0.000	0.057
	0.01	9.255	0.000	0.011
$\rho = 0.7$				
40	0.10	6.671	0.000	0.152
	0.05	8.297	0.000	0.071
	0.01	13.372	0.000	0.013

[a]In all tables KS is the approximation evaluated by equation (2.23) of Kim and Siegmund [2] and given by equation (2.6) with (2.8).

[b]The method of section (2.2) uses equations (2.6-2.9) with (2.12) and variance matrix V replaced by its mle.

for the F statistic with one change-point is 6.1, that from Kim and Siegmund is 6.24 [2]. Bai and Perron [6, p.58] report a value of 5.51 (non-trending x) and Andrews [5, p.840] 5.25. Similarly, for $\alpha = .05$, $m = 20$, Worsley's critical value is 7.5 and that of Kim and Siegmund is 7.77847, Bai and Perron is 6.445 and Andrews is 6.13. For $\alpha = 0.01$, $m = 20$, the Kim and Siegmund value for the F statistic is 11.59 while Bai and Perron is 8.32 and Andrews is 8.02. It is clear that the results differ considerably depending on whether the regressor is trending or not.

Other configurations of x's that were not equally spaced as in Beckman [24] were also simulated, and gave similar results to Table 1, except that the approximation of section 2.2 overestimated the larger probabilities by about 120-200% as the variance of the x's increased. If the x_i's are effectively random i.e. are a sample from some empirical distribution, then as noted by Kim and Siegmund [2], the limiting distribution under H_0 of

Table 2. Accuracy of the approximations to the p-value of the sup-LR, exp(LR) and ave(LR) statistics using the bootstrap for $y_i = x_i + u_i, i = 1, \ldots, m$, $u_i \sim AR(1)$ with autocorrelation parameter ρ, $m_0 = 0.1m, m_1 = 0.9m, x_i = i/m$ and m=40. Percentiles are based on 5000 simulations while bootstrap approximations are based on 200 simulations each with 500 bootstrap iterations.

Test Statistic	True Probability	Estimated tail percentile	Median bootstrap percentile	Bootstrap p-value
$\rho = 0.0$				
sup-LR	0.10	10.646	10.424	0.12
	0.05	12.531	12.139	0.055
	0.01	16.277	15.939	0.015
exp(LR)	0.10	3.014	2.943	0.10
	0.05	3.782	3.667	0.05
	0.01	5.444	5.323	0.01
ave(LR)	0.10	4.049	4.006	0.075
	0.05	4.990	4.969	0.03
	0.01	7.088	7.274	0.01
$\rho = 0.1$				
sup-LR	0.10	7.029	6.152	0.110
	0.05	8.179	7.574	0.055
	0.01	10.649	29.020	0.0
exp(LR)	0.10	1.584	1.331	0.137
	0.05	1.934	1.779	0.041
	0.01	2.879	12.240	0.000
ave(LR)	0.10	2.049	1.874	0.096
	0.05	2.581	2.262	0.041
	0.01	3.644	5.972	0.000
$\rho = 0.4$				
sup-LR	0.10	7.941	7.330	0.054
	0.05	9.044	8.412	0.021
	0.01	11.824	10.860	0.011
exp(LR)	0.10	1.741	1.626	0.042
	0.05	2.114	1.966	0.021
	0.01	3.249	2.834	0.010
ave(LR)	0.10	2.028	1.985	0.090
	0.05	2.409	2.332	0.031
	0.01	3.387	3.273	0.000
$\rho = 0.7$				
sup-LR	0.10	11.175	10.832	0.077
	0.05	12.919	12.193	0.022
	0.01	16.664	15.038	0.011
exp(LR)	0.10	3.009	2.974	0.077
	0.05	3.778	3.567	0.046
	0.01	5.325	4.849	0.000
ave(LR)	0.10	3.709	3.764	0.077
	0.05	4.480	4.572	0.031
	0.01	5.828	6.226	0.015

the likelihood ratio statistic is the same as when there is no covariate. They noted, the p-values assuming random x's are substantially smaller than those assuming equally spaced x's, particularly for larger probabilities. It could be conjectured that the true p-values for the non-equally spaced configurations of x's considered here, fall between the case of equally spaced x's and random x's, making the results obtained here consistent with those already in the literature. We note there are real-life situations, where the covariate is non-equally spaced, is non-random and is not a function of i, say $x_i = f(i/m)$. Theoretical approximations to the p-value of the likelihood ratio statistic for such conditions have as yet proved intractable.

Tables 2 and 3 display percentiles of the three test statistics of Andrews and Ploberger [1] under different conditions on the error terms. The estimated percentiles of the three test statistics are larger in Table 3 than Table 2, for the same AR(1) parameter, as to be expected. Percentiles also increased for the most part as the AR(1) parameter increased. The bootstrap approximations to the percentile and true probability perform very well

Table 3. Accuracy of the approximations to the p-value of the sup-LR, exp(LR) and ave(LR) statistics using the bootstrap for $y_i = x_i + u_i, i = 1, \ldots, m$, $u_i \sim AR(1)$ with autocorrelation parameter ρ, $m_0 = 0.1m, m_1 = 0.9m, x_i = i/m$ and m=40. It is assumed the variance may also change at the change-point. Percentiles are based on 5000 simulations while bootstrap approximations are based on 200 simulations each with 500 bootstrap iterations.

Test Statistic	True Probability	Estimated tail percentile	Median bootstrap percentile	Bootstrap p-value
$\rho = 0.0$ change in variance also				
sup-LR	0.10	13.463	13.521	0.08
	0.05	15.435	15.562	0.05
	0.01	19.814	19.884	0.005
exp(LR)	0.10	4.365	4.293	0.085
	0.05	5.208	5.191	0.045
	0.01	7.185	7.271	0.005
ave(LR)	0.10	5.961	5.790	0.08
	0.05	7.088	6.925	0.045
	0.01	9.905	9.577	0.005
$\rho = 0.1$ change in variance also				
sup-LR	0.10	10.556	14.212	0.047
	0.05	12.771	25.668	0.016
	0.01	21.667	34.573	0.004
exp(LR)	0.10	3.361	6.610	0.032
	0.05	4.324	11.539	0.010
	0.01	8.775	14.979	0.005
ave(LR)	0.10	4.390	11.080	0.015
	0.05	5.368	20.028	0.000
	0.01	8.523	28.602	0.000
$\rho = 0.4$ change in variance also				
sup-LR	0.10	11.463	11.080	0.096
	0.05	13.914	13.928	0.041
	0.01	22.223	21.652	0.014
exp(LR)	0.10	3.434	3.522	0.086
	0.05	4.596	4.632	0.038
	0.01	8.762	8.708	0.014
ave(LR)	0.10	3.912	3.912	0.082
	0.05	4.627	4.943	0.048
	0.01	6.385	7.727	0.010
$\rho = 0.7$ change in variance also				
sup-LR	0.10	14.086	14.833	0.077
	0.05	16.094	17.638	0.029
	0.01	22.326	23.812	0.003
exp(LR)	0.10	4.481	5.024	0.067
	0.05	5.370	6.068	0.029
	0.01	8.361	9.629	0.003
ave(LR)	0.10	5.342	6.085	0.069
	0.05	6.320	7.952	0.034
	0.01	8.241	12.299	0.000

when there is no correlation in the data. The performance for correlated data and for correlated data when there is a change in variance is less good. It can be seen that the median bootstrap percentile over the 200 simulations does not differ greatly from the estimated percentiles, except for $\rho = 0.1$ for the 99th percentile. There was considerable variation in the bootstrap percentiles over the simulations (results not shown). ρ was constrained to be between 0 and 1, and near the boundaries of the parameter space estimates are more variable. The standard deviation of the bootstrap percentiles for $\rho = 0.4$ was smaller than for other values of ρ, and this is reflected in Table 2. However, even in this case the bootstrap p-values are too small. In Table 3, the bootstrap p-values are also too small. Many bootstrap iterations for $\rho = 0.1$ or 0.7 resulted in non-convergence, sometimes more than 40%. Thus, the percentile may be based on a relatively small sample, and the standard deviation of the bootstrap percentiles was larger than in Table 2. However, all p-values are consistently too small. Note also, for the bootstrap p-value to

Table 4. Comparison of methods for estimating the change-point in Quandt's simulated data [3]. The independent variable is a random permutation of the first 20 integers. The change-point is at 12.

	KS[a]	method of section (2.2)	Bootstrap
p-value	0.047	0.048	0.046
95% C. I.[b]	(5,15)	(5,15)	-

[a]KS : method of Kim and Siegmund, section 2.1, assuming i.i.d. errors.

[b]C.I.: confidence interval.

agree exactly with the true tail probability of 0.01, the number of exceedances in 200 simulations has to be exactly 2. However, the simulation results were trustworthy in that bootstrap percentiles and p-values for the first 100 simulations agreed with those for the second 100 simulations.

4. Examples

4.1. *Quandt data*

Quandt [3] simulated data according to a simple linear regression model with a change in both intercept and slope after observation 12 in a sample of size 20. Worsley [12] found the exact p-value for the LR test for a change-point was 0.045, assuming σ unknown, using simulations under H_0. Table 4 shows the methods of sections 2.1 and 2.2 applied to these data give good agreement to the exact p-value. The bootstrap approximation gave a value of 0.046 that was very close. The 95% confidence intervals based on the analytic approximations are quite wide, as the true p-value of the LR test is close to 0.05, Wyse and Kelly [25] carried out a Bayesian analysis of these data and reported 95% credible intervals corresponding to three priors for the change-point, as in Ferreira [26]:
(i) uniform prior
(ii) prior that maximized the posterior distribution at the mle
(iii) prior that maximized the posterior distribution where the residual sum of squares was minimized.
The priors also satisfied the assumption that there are at least two observations in each regime. All other parameters α_1, α_2, β_1, β_2 and $\log(\sigma^2)$ were assumed to have independent uniform priors over the whole real line. We note applying both Worsley [12] and Lee's [17] methods the number of change-points is estimated to be one. It is computationally intensive to compute a confidence interval using Worsley's method based on bootstrap p-values, and therefore this was not carried out. A bootstrap interval was constructed conditional on there being one change-point. This was calculated by bootstrapping residuals from the two-segment model at the estimated change-point. The resulting interval was (2,18).

4.2. *Physiology data*

The assumption of i.i.d errors in change-point regression is often too restrictive as illustrated by the problem of assessing aerobic fitness of subjects in cardiology and physiology. This is frequently done non-invasively, using incremental work rate tests on a cycle ergometer Oxygen uptake ($\dot{V}O_2$) and carbon dioxide ($\dot{V}CO_2$) output is measured on a breath-by-breath basis during exercise. At a point, known as the gas exchange threshold (GET), the linear relationship between $\dot{V}CO_2$ and $\dot{V}CO_2$ becomes steeper, as the subject

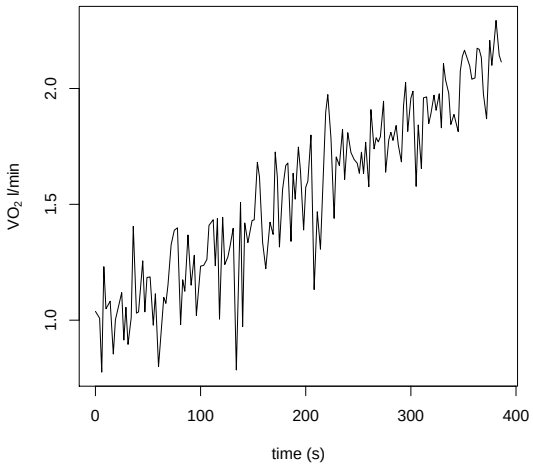

Figure 1. A plot of oxygen uptake (VO_2) in litres per minute versus time

Table 5. Comparison of methods for estimating the change-point in the physiology data for subject 1 from Kelly [9] assuming both correlated observations and a variance change at the change-point. The mle of the change-point is 195.

	KS[a]	method of section (2.2)	Bayes using uniform priors	bootstrap
p-value	0.001	0.0006		0.0000
95% C. I.[b]	(185,205)	(195)	(188,200)	

[a]KS : method of Kim and Siegmund [2], section 2.1, assuming i.i.d. errors.

[b]C.I.: confidence interval except for the Bayes methods where it denotes credible interval.

switches from aerobic to a mixture of aerobic and anaerobic metabolism. Change-point methodology can be used to estimate the GET, which in turn can be used to monitor changes in fitness and heart functioning [27]. It is currently enjoying great popularity among athletes, with many companies offering metabolic analysis systems with professional software for determination of the GET [28]. The software methods generally are valid only under the restrictive assumption of i.i.d. errors. A more general model is where the u_i is an ARMA process of order (p, q) with the variance of the innovations possibly changing at the change-point. One then has a 2-segment transfer function model. This allows in estimating GET, for example an unusually large $\dot{V}CO_2$ value to have a carry-over effect on to subsequent $\dot{V}CO_2$ values, which cannot be explained by regression on $\dot{V}O_2$. It also allows for random effects, such as the subject's temperature, to affect the calibration of the immediate response and subsequent responses. Moreover, unless standard errors or confidence limits are attached to estimates, it is difficult to assess when a real change in GET has occurred. Standard transfer function models assume equally spaced time intervals. The values of $\dot{V}O_2$ are not strictly monotone in time and because intervals between breaths are not equally spaced, the change-point is found on breath number and converted to time. Kelly *et al.* [29] describe data from healthy subjects undergoing incremental ramp exercise (20 W$\cdot$ min^{-1}) on a bicycle to the limits of tolerance. Here we consider subject 1 from that data and Figure 1 illustrates the data.

A model assuming both correlated observations and a variance change at the change-

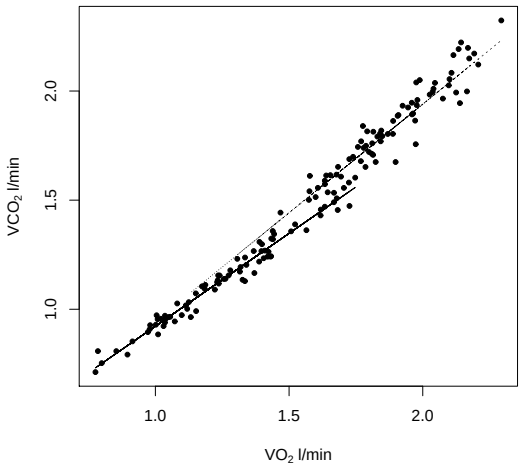

Figure 2. The data oxygen uptake (VO_2) and carbon dioxide exhaled (VCO_2) and estimated regression lines.

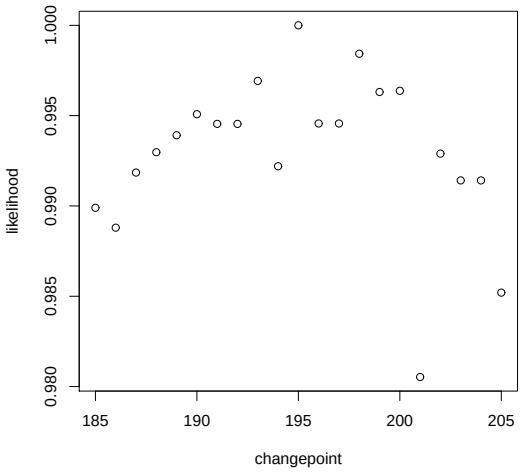

Figure 3. Normalized likelihood for the changepoint in Figure 2.

point is the preferred one both from a statistical and biological viewpoint [9]. Under this latter model, the mle of the change-point is 195 and from Table 5 the associated 95% LR confidence interval is the single point (195). Figure 2 shows the data with the estimated regression lines exhibiting a dis-continuous mean function. Figure 3 gives a graphical display of the normalized likelihood. (Note the usual cut-off of 0.15 based on the χ^2 approximation to the likelihood ratio does not apply here since the likelihood is not regular.)

Figure 2, Figure 3 here

In contrast, the Bayes [25] method gave wide credible intervals, as shown in Table 5. The differing results indicate that some methodological problems remain even with respect to simple change-point estimation.

We note the Bayesian method gave an extremely large pseudo-Bayes factor [25] i.e. strong evidence for a change-point. Using both the methods of Worsley [12] and Lee [17], the

estimated number of change-points is one.

5. Computational issues

Both for the simulated data and examples, the function `elliptic` used for model fitting, uses some of the same arguments that control the commonly used R function `nlm`. The function `elliptic` when fitted to data returns a code - an integer indicating if the optimization process terminated successfully. This was incorporated into the simulations, so that if optimization failed at a particular change-point value i, then this value i was excluded when taking the maximum in equation (2.5) in computing the likelihood ratio statistic for that simulation. Failure occurred in some simulations where the change-point being fitted i, was either close to 1 or the sample size m and thus few data were available to fit one of the line segments, leading to unstable estimates.
The function `elliptic` requires starting values for models and values close to the true were used in the case of simulated data. Optimization failure occurred if the starting values were poor and it was not possible to automate 'good' starting values. In the case of the examples, values close to empirical values were used.
The analytical approximation to the tail probability of the sup-LR test involves a double integral - see equation (2.6) above. In the simulation study, when this was evaluated empirically using Riemann sums, the argument in the quantity in square brackets in some terms of the sums was negative, and of course one cannot take the square root of a negative number. These terms were set to zero in the Riemann sum. The accuracy of the approximations remains good in the simulations (similar to those of [2] and close to the true values). As noted in [2] the approximations are only valid for small tail probabilities and this led to the negative quantities.
When fitting Bayesian models, the simple form for the variance-covariance matrix $\hat{V^{-1}}$ given in the Appendix was used to compute posterior densities without matrix inversion.
Computational intensity was a feature of this simulation study and results presented are preliminary before undertaking a larger study. In a changepoint-like problem arising in genetic linkage analysis considered by Manichaikul *et al.* [30] the performance of the bootstrap was investigated. For some aspects of their simulation studies, C code was adapted from R code to improve computational speed. The modifications using C code resulted in a 10− to 15−fold improvement in speed. 10,000 simulation replicates were performed with a 1,000 bootstrap replicates at each. The total computer time for their simulations was ∼450 days. The simulations were split across multiple processors on a fast cluster, but still required ∼2 months of constant computation.

6. Conclusions

We have extended Kim and Siegmund's [2] solution to the p-value of the sup-LR test for the change-point in the simple linear regression to the case of non i.i.d. errors. The simulation results verify that the approximations derived are sufficiently accurate as to be very useful. In addition, it is clear, the approximations to the p-value derived can be applied to any covariance structure V and not just that of model (2.1). It is not clear if the method is sensitive to the mis-specification of V and this is a subject for further work.
Using an application of Worsley's method [16], we showed how confidence intervals corresponding to our p-value approximations might be derived, and it can be argued the Worsley interval is akin to a sup-LR interval [31]. Applying Worsley's method to p-values

obtained by the bootstrap is computationally intensive, and thus analytic approximations continue to be useful.

Siegmund [31] noted that in general the likelihood in change-point models will be multimodal but if sharply peaked the confidence set may be small. In a simulation study he showed the Bayes solution makes too small an adaptation to the departure of the log-likelihood from its expected shape (under the prior). This is perhaps what has occurred here with subject 1 (see Figure 3) for the physiology data. Our finding of a confidence interval of a single point (based on our p-value analytical approximations) was also found in Kelly [29] using a different time series model and this is reassuring. A study by Chen *et al.* [32], also considered some gas exchange data, referred to as metabolic pathway data, where the variance could also change at the change-point but errors were assumed independent. The change-point was estimated using a grid-search (yielding an estimate close to the mle), Bayesian and some other methods. By the grid-search method, the estimated variance in the second regime is very small (in comparison to the change in slope), perhaps indicating a sharply peaked likelihood at the change-point, that would explain the difference with the Bayesian estimate. When methods were compared by leave-one-out predictive measures, Bayesian methods performed best and the grid-search method performed poorly since, because, as noted by the authors, it produced larger residuals when observations close to the estimated change-points were left out of the data. This however, may in fact be an indication that the estimated change-point by the grid-search is very accurate, as standard diagnostics are not applicable. As noted in [9], estimates of the change-point are extremely sensitive to changes in variance parameter estimates.

Höhle and Mazick [4] use algorithms in the R add-on package **surveillance**, available from CRAN, to prospectively detect a change-point, but use is limited to count data time series. A CUSUM statistic is recommended and interestingly, as here, the authors use Monte Carlo estimation to estimate critical values. However, its properties are not evaluated.

This study was restricted to tests for a change-point and the three considered were because of their optimality properties as well as their having a Bayesian formulation. In a simulation study, bootstrap p-values of test statistics proved accurate for the case of i.i.d. errors, and also worked well in an example where errors were not assumed to be i.i.d.

For the cases of non i.i.d errors considered the p-values tended to underestimate the true probability. This was a relatively small simulation study because it was computationally very intensive. The poor performance of the bootstrap may be due to the relatively small sample size. However, in the changepoint-like problem considered in [30], the bootstrap performed poorly. This was related to the discreteness of the distribution of the mle. Similarly here, the mle of the change-point occurs at at discrete observed values $x_i, k_1 \geq i \leq k_2$ and may also explain the results. The authors [30] concluded modifications to the bootstrap are necessary for problems that are not classically regular, as discussed in Beran [33].

Methods for obtaining bootstrap confidence intervals, that are computationally not too intensive, require further research. Analytical approximations therefore continue to be useful.

Acknowledgements

This research was partly supported by Science Foundation Ireland grant 06/RFP/MAT024. I would like to thank an Associate Editor and a referee for

helpful comments.

References

[1] Andrews DW, Ploberger W. Optimal tests when a nusiance parameter is present only under the alternative. Econometrica. 1994;62:1383–1414.

[2] Kim H-J, Siegmund D. The likelihood ratio test for a change-point in simple linear regression. Biometrika. 1989;76:409–423.

[3] Quandt RE. The estimation of a parameter of a linear regression system obeying two separate regimes. J Am Statist Assoc. 1958;53:873–880.

[4] Höhle M, Mazick A (2010). Aberration detection in R illustrated by Danish mortality monitoring. In T Kass-Hout, X Zhang editors. Biosurveillance: A Health Protection Priority. CRC Press; 2010. chapter 12, pp. 215–238.

[5] Andrews DWK. Tests for parameter instability and structural change with unknown change point. Econometrica. 1993;61:821–856.

[6] Bai J and Perron P. Estimating and testing linear models with multiple structural changes. Econometrica. 1998;66:47–78.

[7] Bai J and Perron P. Computation and analysis of multiple structural change models. J Appl Econometrics. 2003;18:1–22.

[8] Andrews DWK, Lee I, Ploberger W. Optimal changepoint tests for normal linear regression. Journal of Econometrics. 1996;70:9–38.

[9] Kelly GE, Lindsey JK, Thin AG. Models for estimating the change-point in gas exchange data. Phys Meas. 2004;25:1425–1436.

[10] Perron P. Dealing with structural breaks. In: Mills TC, Patterson K, editors. Palgrave Handbook of Econometrics. Vol. 1. Boston (MA): Kluwer Academic; 2006. p. 278–352.

[11] Kim H-J. Change Point Problems in Regression. [unpublished doctoral dissertation]. Stanford (CA): Stanford University; 1988.

[12] Worsley KJ. Testing for a two-phase multiple regression. Technometrics. 1983;25:35–42.

[13] Hansen B. Approximate asymptotic p values for structural-change tests. J Bus Econ Stat. 1997;15:60–67

[14] Zeileis A, Leisch F, Hornik K, Kleiber C. strucchange: An R package for testing for structural change in linear regression models. J Stat Soft. 2002;7(2):1-38.

[15] R Development Core Team. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing Vienna, Austria;2011. Available from: http://www.R-project.org

[16] Worsley KJ. Confidence regions and tests for a change-point in a sequence of exponential family random variables. Biometrika. 1986;73:91–104.

[17] Lee C-B. Estimating the number of change-points in exponential families distributions. Scand J Statist. 1997;24:201–210.

[18] Hinkley D. Bootstrap Methods. J R Statist Soc B. 1988;50:321–337.

[19] Julious S. Inference and estimation in a changepoint regression problem. J Roy Stat Soc D. 2001;50:51–61.

[20] Politis D. The Impact of Bootstrap Methods on Time Series Analysis. Stat Sci. 2003;18:219–230.

[21] Smeekes S. Bootstrapping Nonstationary Time Series. [Ph.D. thesis]. Universitaire Pers Maastricht; 2009.

[22] Psaradakis Z. Bootstrap tests for an autoregressive unit root in the presence of weakly dependent errors. J Time Ser Anal. 2001;22:577–594.

[23] Kim H-J. Change-point detection for correlated observations. Statist Sin. 1996;6:275–287.

[24] Beckman RJ, Cook RD. Testing for two-phase regressions. Technometrics. 1979;21:65–69.

[25] Wyse J, Kelly GE. A Bayesian analysis of a change-point model with two regimes. School of Mathematical Sciences, University College Dublin, Dublin, Ireland; 2008. Tech Rep.

[26] Ferriera PE. A Bayesian analysis of a switching regression model: known number of regimes. J Am Statist Assoc. 1975;70:370–374.

[27] Myers J, Goldsmith RL, Keteyian SJ, Brawner CA, Brazil DA, Aldred H, Ehrman JK, Burkhoff D. The Ventilatory Anaerobic Threshold in Heart Failure: A Multicenter Evaluation of Reliability. J Cardiac Failure. 2010;16:76–83.
[28] Ekkekakis P, Lind E, Hall E, Petruzzello S. Do regression-based computer algorithms for determining the ventilatory threshold agree? J Sports Sciences. 2008;26:967–976.
[29] Kelly GE, Lindsey JK, Thin AG. Estimation of the gas exchange threshold in humans: a time series approach. Eur J App Physiol. 2001;85:586–592.
[30] Manichaikul A, Dupuis J, Sen Ś, Broman KW. Poor performance of bootstrap confidence intervals for the location of a quantitative trait locus. Genetics. 2006;174:481-489
[31] Siegmund D. Confidence sets in change-point problems. Int Stat Rev. 1988;56:31–48.
[32] Chen CWS, Chan JSK, Gerlach G, Hsieh WYL. A comparison of estimators for regression models with change points. Stat Comput. 2011;21:395–414.
[33] Beran R. The impact of the bootstrap on statistical algorithms and theory. Stat Sci. 2003;18:175-184.
[34] Gómez E, Gómez-Villegas MA, Marin JM. A multivariate generalization of the power exponential family of distributions. Commun in Statist A. 1998;27:589–600.
[35] Lindsey JK. Nonlinear Models in Medical Statistics. Oxford (UK): Oxford University Press; 2001.

Appendix A.

The variance-covariance matrix V of the vector of observations y in equation (2.1) is as follows. We note $u_i = \sum_1^i \rho^{i-k}\epsilon_k$ and therefore $\text{Var}(u_i) = \sum_1^i \rho^{2(i-k)}\text{Var}(\epsilon_k)$ and for $i < j$ $\text{Cov}(y_i, y_j) = \rho^{j-i}\text{Var}(u_i)$. Therefore, for $i \leq j$

$$V[i,j] = \begin{cases} \left(\frac{1-\rho^{2i}}{1-\rho^2}\right)\sigma_1^2, & i = j \leq \tau \\ \rho^{j-i}\left(\frac{1-\rho^{2i}}{1-\rho^2}\right)\sigma_1^2, & 1 \leq i \leq \tau \\ \left(\frac{\rho^{2(i-\tau)}-\rho^{2i}}{1-\rho^2}\right)\sigma_1^2 + \left(\frac{1-\rho^{2(i-\tau)}}{1-\rho^2}\right)\sigma_2^2, & i = j > \tau \\ \rho^{j-i}\left[\left(\frac{\rho^{2(i-\tau)}-\rho^{2i}}{1-\rho^2}\right)\sigma_1^2 + \left(\frac{1-\rho^{2(i-\tau)}}{1-\rho^2}\right)\sigma_2^2\right], & \tau < i \end{cases}$$

and for $i > j$ we use the fact $V[i,j] = V[j,i]$. Note

$$V[\tau+1, \tau] = \frac{\rho(1-\rho^{2\tau})}{(1-\rho^2)}\sigma_1^2$$

$$V[\tau+1, \tau+1] = \left(\frac{\rho^2(1-\rho^{2\tau})}{1-\rho^2}\right)\sigma_1^2 + \sigma_2^2$$

$$V[\tau+1, \tau+2] = \frac{\rho^4 - \rho^{2(\tau+1)}}{1-\rho^2}\sigma_1^2 + \frac{1-\rho^4}{1-\rho^2}\sigma_2^2$$

The determinant of V is given by, $|V| = \sigma_1^{2\tau}\sigma_2^{2(m-\tau)}$. The inverse of V has a special form with zeros everywhere except on the main and first minor diagonals and in addition it has a simple Cholesky decomposition $V^{-1} = A^T A$. V^{-1} and A are given by:

$$V^{-1} = \begin{bmatrix} \frac{1+\rho^2}{\sigma_1^2} & \frac{-\rho}{\sigma_1^2} & 0 & 0 & \dots & & 0 & & & \\ \frac{-\rho}{\sigma_1^2} & \frac{1+\rho^2}{\sigma_1^2} & \frac{-\rho}{\sigma_1^2} & & & & & & & \\ 0 & \frac{-\rho}{\sigma_1^2} & \frac{1+\rho^2}{\sigma_1^2} & \frac{-\rho}{\sigma_1^2} & & \dots & 0 & \dots & & \\ \vdots & & & \ddots & & & \vdots & & & \\ & & & & \frac{-\rho}{\sigma_1^2} & \frac{1+\rho^2}{\sigma_1^2} & \frac{-\rho}{\sigma_2^2} & 0 & 0 & \dots \\ & & & & 0 & \frac{-\rho}{\sigma_2^2} & \frac{1+\rho^2}{\sigma_2^2} & \frac{-\rho}{\sigma_2^2} & 0 & \\ & & & & 0 & 0 & \frac{-\rho}{\sigma_2^2} & \frac{1+\rho^2}{\sigma_2^2} & \frac{-\rho}{\sigma_2^2} & \\ & & & & \vdots & & & \ddots & & \\ 0 & & \dots & 0 & \dots & & & \frac{-\rho}{\sigma_2^2} & \frac{1+\rho^2}{\sigma_2^2} & \frac{-\rho}{\sigma_2^2} \\ & & & & & & & 0 & \frac{-\rho}{\sigma_2^2} & \frac{1}{\sigma_2^2} \end{bmatrix} A = \begin{bmatrix} \frac{1}{\sigma_1} & 0 & 0 & \dots & 0 & 0 \\ \frac{-\rho}{\sigma_1} & \frac{1}{\sigma_1} & 0 & \dots & 0 & 0 \\ 0 & \frac{-\rho}{\sigma_1} & \frac{1}{\sigma_1} & 0 & \dots & 0 \\ \vdots & \vdots & \ddots & \vdots & \vdots & \vdots \\ 0 & \dots & \frac{-\rho}{\sigma_2} & \frac{1}{\sigma_2} & 0 & \dots \\ \vdots & \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \dots & \frac{-\rho}{\sigma_2} & \frac{1}{\sigma_2} & 0 \\ 0 & 0 & \dots & 0 & \frac{-\rho}{\sigma_2} & \frac{1}{\sigma_2} \end{bmatrix}$$

Note $V^{-1}[\tau+1,\tau] = \frac{-\rho}{\sigma_2^2}$ and $V^{-1}[\tau+1,\tau+1] = \frac{1+\rho^2}{\sigma_2^2}$.
Using V^{-1} it is easy to show that maximum likelihood estimators (mle's) for the unknown parameters $\alpha_0, \beta_0, \alpha_1, \beta_1, \sigma_1, \sigma_2, \rho$ are given by minimizing the expression

$$\sum_{i=1}^{\tau} \frac{(y_i - \alpha_0 - \beta_0 x_i)^2(1+\rho^2)}{\sigma_1^2} + \sum_{i=\tau+1}^{m-1} \frac{(y_i - \alpha_1 - \beta_1 x_i)^2(1+\rho^2)}{\sigma_2^2} + \frac{(y_m - \alpha_1 - \beta_1 x_m)^2}{\sigma_2^2}$$
$$- \sum_{i=1}^{\tau-1} \frac{2\rho}{\sigma_1^2}(y_i - \alpha_0 - \beta_0 x_i)(y_{i+1} - \alpha_0 - \beta_0 x_{i+1}) - \sum_{i=\tau+1}^{m-1} \frac{2\rho}{\sigma_2^2}(y_i - \alpha_1 - \beta_1 x_i)(y_{i+1} - \alpha_1 - \beta_1 x_{i+1})$$
$$- (y_\tau - \alpha_0 - \beta_0 x_\tau)(y_{\tau+1} - \alpha_1 - \beta_1 x_{\tau+1})\frac{2\rho}{\sigma_2^2} + \tau \log(\sigma_1) + (m-\tau)\log(\sigma_2)$$

In the case $\sigma_1^2 = \sigma_2^2$, V^{-1} simplifies to

$$V^{-1} = \begin{bmatrix} \frac{1+\rho^2}{\sigma^2} & -\frac{\rho}{\sigma^2} & 0 & \dots & 0 & 0 \\ -\frac{\rho}{\sigma^2} & \frac{1+\rho^2}{\sigma^2} & -\frac{\rho}{\sigma^2} & \dots & 0 & 0 \\ 0 & -\frac{\rho}{\sigma^2} & \frac{1+\rho^2}{\sigma^2} & \dots & 0 & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots & \vdots \\ 0 & 0 & 0 & \dots & \frac{1+\rho^2}{\sigma^2} & -\frac{\rho}{\sigma^2} \\ 0 & 0 & 0 & \dots & -\frac{\rho}{\sigma^2} & \frac{1}{\sigma^2} \end{bmatrix}$$

and the (mle's) for the unknown parameters $\alpha_0, \beta_0, \alpha_1, \beta_1, \sigma, \rho$ are given by minimizing the expression

$$\sum_{i=1}^{\tau}(y_i - \alpha_0 - \beta_0 x_i)^2 \frac{1+\rho^2}{\sigma^2} + \sum_{i=\tau+1}^{m}(y_i - \alpha_1 - \beta_1 x_i)^2 \frac{1+\rho^2}{\sigma^2} + (y_m - \alpha_1 - \beta_1 x_m)^2 \frac{1}{\sigma^2}$$
$$- \sum_{i=1}^{\tau-1}(y_1 - \alpha_0 - \beta_0 x_1)(y_{i+1} - \alpha_0 - \beta_0 x_{i+1})\frac{2\rho}{\sigma^2} - \sum_{i=\tau+1}^{m-1}(y_1 - \alpha_1 - \beta_1 x_1)(y_{i+1} - \alpha_1 - \beta_1 x_{i+1})\frac{2\rho}{\sigma^2}$$
$$- (y_\tau - \alpha_0 - \beta_0 x_\tau)(y_{\tau+1} - \alpha_1 - \beta_1 x_{\tau+1})\frac{2\rho}{\sigma^2} + m\log(\sigma)$$

When there is no change-point this becomes

$$\sum_{i=1}^{m-1}(1+\rho^2)(y_i - \alpha_0 - \beta_0 x_i)^2 - 2\rho\sum_{i=1}^{m-1}(y_i - \alpha_0 - \beta_0 x_i)(y_{i+1} - \alpha_0 - \beta_0 x_{i+1}) + (y_m - \alpha_0 - \beta_0 x_m)^2$$
$$- m\log(\sigma^2)$$

Appendix B.

R code : Estimate the change-point for a single data set $(x_i, y_i), i = 1, \ldots, m$ where a regression line is fitted whose parameters may change at the change-point and where the variance may also change at the change-point.

```
library(MASS)
library(growth)
data<-read.table("data.txt")
colnames(data)<-c("x","y")
mu <- function(p){
ifelse(x > tau,p[1]+p[2]*x,p[3]+p[4]*x)}
varfn <- function(p)p[1]*(x <= tau)+p[2]*(x > tau)
###initial estimates for the parameters of the regression lines assuming a
###change-point at some value (here at tau <- 15)
tau <- 15
mdl <- elliptic(y,time=x,model=mu,preg=c(0.0,1.0,-0.1,0.9),
pvar=c(1.0,2.0),shfn=F,varfn=varfn,
par=.85, pell=1)
###estimates
mu <- function(p){
ifelse(x > tau,p[1]+p[2]*x, p[3]+p[4]*x)}
varfn <- function(p)p[1]*(x <= tau)+p[2]*(x > tau)
mdl <- elliptic(y,time=x,model=mu,preg=coef(mdl)[1:4],
pvar=coef(mdl)[5:6],shfn=F, varfn=varfn,
par=exp(coef(mdl)[7])/(1+exp(coef(mdl)[7])),
pell=1, dist="normal")
###- estimated log-likelihood
like<-mdl$maxlike
```

The function elliptic used here refers to the subclass of elliptically-contoured distributions which are a multivariate generalization of the power-exponential family introduced by Gómez [34]. They all have known covariance matrices and the covariance matrix can be structured for time series dependence and a variance change at the change-point. They include the multivariate normal as a special case and that is used in this paper. Here y is the outcome variable, x is the predictor variable with one unknown change-point, preg refers to the parameters of the regression lines, pvar the variance parameters and par to the AR(1) parameter of the correlated error terms. mdl contains all parameter estimates and the estimated log-likelihood. A more detailed description of the data objects and model formulae used in the Growth library can be found in the Appendix of Lindsey [35].