Journal of Data Science 13(2015),157-186

Heteroscedastic Change Point Analysis and Application to Footprint Data

Stephen J. Ganocy and Jiayang Sun

Case Western Reserve University

Abstract: Analysis of footprint data is important in the tire industry. Estimation procedures for multiple change points and unknown parameters in a segmented regression model with unknown heteroscedastic variances are developed for analyzing such data. Our approaches include both likelihood and Bayesian, with and without continuity constraints at the change points. A model selection procedure is also proposed to choose among competing models for fitting a middle segment of the data between change points. We study the performance of the two approaches and apply them to actual tire data examples. Our Maximization–Maximization–Posterior (MMP) algorithm and the likelihood–based estimation are found to be complimentary to each other.

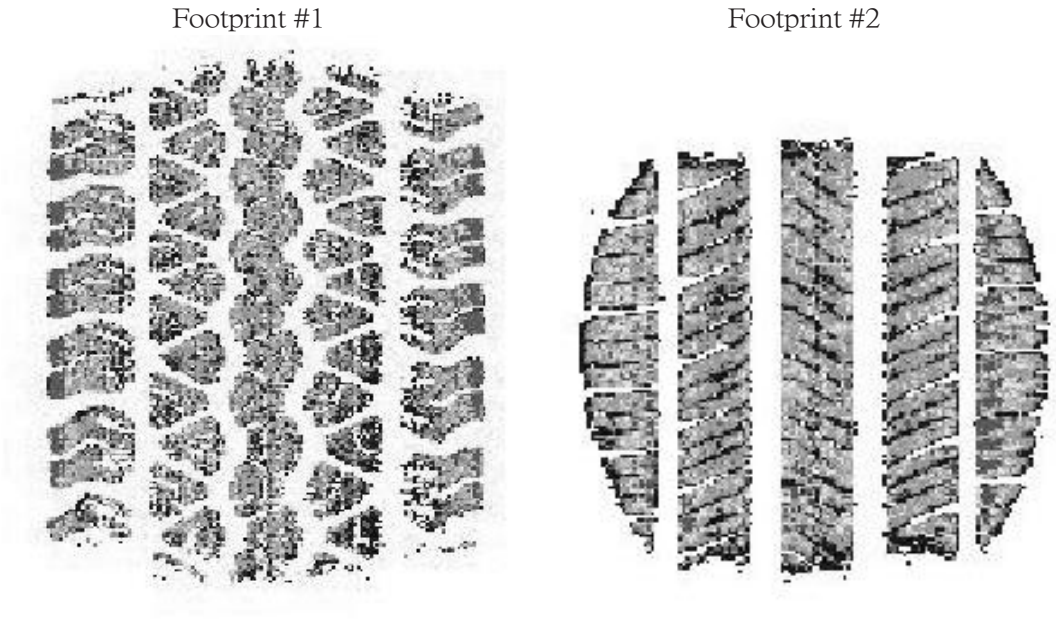
Key Words: Change point analysis, segmented regression, piecewise or joinpoint regression, heteroscedastic variance, likelihood and Bayesian methods, algorithms, model selection, footprint and tire data.

*This research is supported in part by the National Science Foundation.

1. INTRODUCTION

Footprint data in the tire industry is important to engineers in ascertaining many important features of the tire that affect its performance and endurance. These in turn have a significant impact on fuel economy and other environmental issues. In order to obtain an image of a tire footprint, a tire is either placed on, or rolled slowly over a device which contains sensors that measure the contact of the tire with a surface (Sakai, 1995). This procedure produces a pixelated image of the contact region for the tire either in color or in shades of gray (as illustrated in Figure 1). The differences in the gray scale of the footprint image indicate the nonuniformity of the contact pressure of the tire with the surface on which it is resting or rolling. Due to the irregular nature of the geometry of the tire and inherent uncertainty in the measuring device the boundary of the surface of contact pressure is not known exactly. This poses a challenge to determine the footprint "length" which is used to measure how much of the tire material actually makes contact with the

surface on which it is riding. Footprint length is used in conjunction with other tire measurements to compare the characteristics of different designs of tires or to compare the characteristics of a specific tire design to itself under different load and inflation conditions. The characteristics of interest include rolling resistance, traction, wear, etc. As an example, a significant change in the ratio of the footprint width to footprint length of a newly designed tire after testing for say 1,000 hours at 28 PSI inflation and at a constant load indicates that the tire is wearing unevenly, which will lead to safety and performance issues. This dictates a change in some design feature of the tire, e.g. tread geometry or tread compound. This paper is concerned with developing techniques for determining the length of the footprint.



,

Figure 1: Tire footprint images

Footprint length can be determined by a change point analysis as given below. Since the density of the pixels in the footprint image can be converted into equivalent pressure values measured in pounds per square inch (PSI), we then have the footprint pressure curve (FPC) as shown in Figure 2. The horizontal axis represents rectangular slices of the tire footprint, measured at .05 inch intervals, starting at the top of the footprint. The vertical axis corresponds to the average pressure calculated within each slice. Thus the first point on the FPC (at $x = 0$) plots the average PSI for the first (top) slice and the last point plots the average PSI for the last (bottom) slice. Since a tire footprint has finite length the footprint pressure curve must include two points, say τ_1 and τ_2; one for the beginning of the footprint (leading edge) and one for the end of the footprint (trailing edge) - see Figure

3. These two points are change points of the expected FPC in which the middle segment is measured with more random noise than the two tails. The middle segment appears as a polynomial curve which needs to be determined by data. Even though some FPC's indicate that there may be more change points in the middle region, these are not of interest to the tire engineers for the estimation of the footprint length (FPL). The behavior of the FPC in this region may indicate other important features of the tire, however, which could relate to traction properties, for example. The *footprint length* is the difference of the two change points, $\tau_2 - \tau_1$.

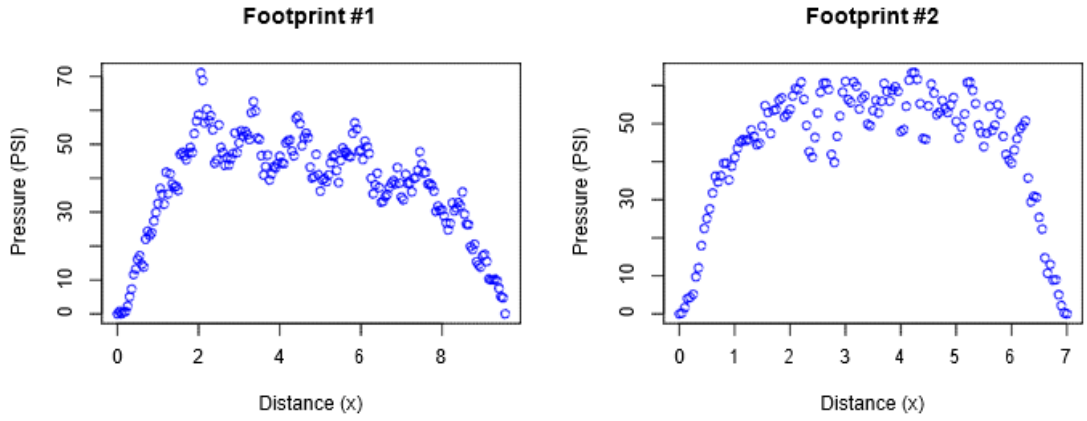


Figure 2: Examples of footprint pressure curves associated with the tire footprint images

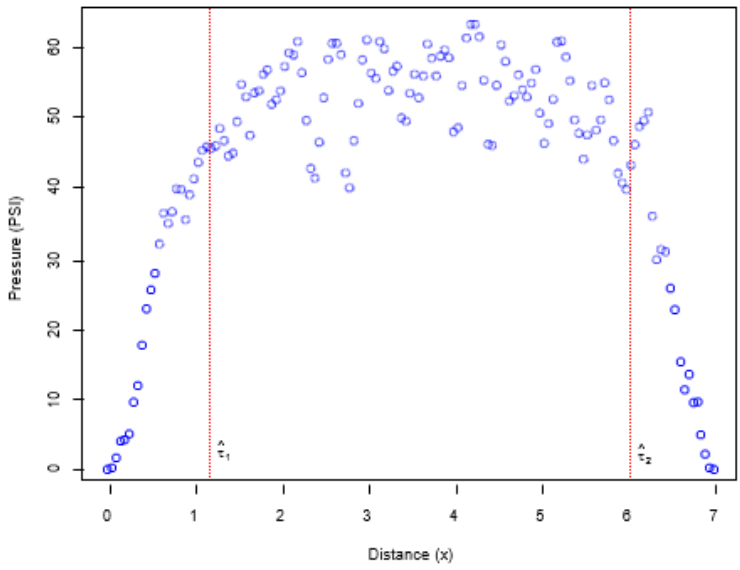


Figure 3: Footprint pressure curve with change points

If we let x denote the horizontal distance along the footprint pressure curve and Y be the footprint pressure in pounds per square inch (PSI) then a reasonable model for the data from a FPC is one consisting of 3 regression segments:

$$Y = \mu_i(x; \beta_i, \tau_i) + \varepsilon^i, \text{for } x \in I_i, \quad (1)$$

where $I_1 = [a, \tau_1]$, $I_2 = (\tau_1, \tau_2]$ and $I_3 = (\tau_2, b]$, and ε^i is distributed with mean 0 and variance σ_i^2 for $i = 1, 2, 3$. Therefore, one way to determine the footprint length is by estimating the parameters β_i, τ_i from the change–point model in (1).

Change point problems occur in many areas of application. Examples of recent change point applications include the following: Arani, Soong, Weiss, Wood, Fiddian, Gnann and Whitley (2001) and Desmond, Weiss, Arani, Soong, Wood, Fiddian, Gnann and Whitley (2002) used maximum likelihood estimation to find the change points in a multistage hazard function model of pain associated with herpes zoster and applied bootstrapping to estimate the standard errors of the change–points. Change point estimation via permutation tests is used by Kim, Fay, Feuer and Midthune (2000) to model US prostate cancer rates, while Pauler and Finkelstein (2002) and Slate and Turnbull (2000) provided Bayesian change point methodology to predict time to prostate cancer recurrence or onset. There have been applications in animal science (Bergfelt, Sego, Beg and Ginter, 2003), intervention (Shope and Molnar, 2003), and bioinformatics (Zhang and Siegmund, 2007), where Zhang and Siegmund (2007) also developed a modified Bayes information criteria (BIC) for Gaussian change point models to determine the number of changed chromosome copies at each genome location of a given cell sample. In applications to financial data, Ray and Tsay (2000) applied a Bayesian change point analysis to the duration between trades of IBM stock and the volatility of daily Coca Cola stock returns; Loschi, Cruz, Iglesias and Arellano–Valle (2003) used Gibbs sampling to analyze Brazilian stock market data; and most recently Spokoiny (2009) offered a local change point detection approach to modeling and forecasting of nonstationary time series with applications to volatility modeling for financial data.

A change point model can be specified in several different ways based on the following factors:

the *number* of change points (known or unknown), the *location* of the change points (τ_i),

often unknown in statistics, while the Chow's test frequently used in economics (Chow (1960)) tests if a change has occurred at a known location), the *type* of the change points (assumed abrupt or continuous), the *error structure* (*E*, homoscedastic, heteroscedastic, independent or correlated) and error distribution (often assumed Gaussian, but can be other distributions such as a heavy tailed distribution), the *type of regression function* μ (assumed parametrically up to some finite unknown parameters or nonparametrically as a piecewise smoothed function) if it is a change point regression problem, the *change scheme* (changes in the mean or parameters of) regression function, or error structure or both), the *detection process* (finding change points from a given data set of fixed size *n* - may be called *looking-back process*, or signaling a change as soon as possible as the data points come in - may

be called *looking-forward process*, used frequently in *sequential analysis*).

The wide applicability of change point analysis and its versatile modeling has motivated extensive research in the subject. See the five texts (Csörgö and Horváth, 1997; Chen and Gupta, 2000; Brodsky and Darkhovsky, 1993; Siegmund, 1985; Cohen, 2008) and the references therein. Asymptotic properties of change points and hypothesis testing involving change points can be found in the text by Csörgö and Horváth (1997). Parametric change point analysis techniques are provided by Chen and Gupta (2000). Some nonparametric methods in change point problems are in Brodsky and Darkhovsky (1993). Connections of change point problems to boundary crossing problems and a practical correction to the approximation by Brownian motion are given in Siegmund (1985). Applied "turning point" analysis in the behavioral and sociological sciences by Cohen (2008) is also related to change point analysis. There was also a proceedings devoted to change point problems edited by Carlstein, Müller and Siegmund (1994). Recent research papers include Zhang, Siegmund, Ji and Li (2009) for detecting recurrent copy number variants (i.e. simultaneous change points) in multiple genome samples, Kim, Yu and Feuer (2009) for selecting the number of change points in segmented line regression, Spokoiny (2009) for multiscale local change point detection, and Zhao, Wu and Zhou (2009) for change point hazard model for survival data with long-term survivors, and Chen and Raimondo (2008) for change point estimation in derivatives, among others. The literature listed here is by no means exhaustive. The change point analysis methodology we develop will be limited to that relevant to our footprint data analysis, although the methodology should be applicable to data from other application fields.

In our case of the footprint data (see Figures 2 and 3), the number of interesting change points are known to be two, the location of changes are unknown and to be estimated, the errors are independent to each other but are heteroscedastic from one segment to the another, the reasonable regression functions are segmented polynomials with unknown coefficients in which the polynomial order in the middle segment may also need to be determined (based on the data or the type of tires under study), the changes can be in both the parameters of the regression function and the error variances while the change type appears to be continuous (however we shall develop estimation procedures with and without continuity, under the likelihood approach), and the detection process is the "looking-back" process (i.e. not those used in sequential analysis). So, this paper is concerned with estimation of parameters and location of change points in a change point regression model.

Articles on the change point problem in a regression setting can also be found under the heading of "segmented", "piecewise", "joinpoint" or "bent–cable" regression. We shall use the terminology of segmented regression. Gallant and Fuller (1973) provided least squares type estimators for segmented polynomial regression models with homoscedastic errors. In the field of animal science, Bergfelt, Sego, Beg and Ginter (2003) used segmented regression for modeling follicle growth in cattle; they estimated a single change point in simple linear regression by incremental searching and basing the final value on that which gives the highest R^2. Montiani–Ferreira, Petersen–Jones, Cassotis, Ramsey, Gearhart and Cardoso (2003) investigated the changes in corneal thickness of canine eyes as the animals mature,

utilizing two different regression functions with a specified value of a single change point. The asymptotic distribution theory of the least squares estimators of the parameters in homoscedastic segmented

regression models is studied in Feder (1975a) and Feder (1975b). Maximum likelihood estimation of change points in two–phase linear regression models with homoscedastic error is covered in Hinkley (1969), Hinkley (1971) and Hudson (1966). Most existing papers for change point regression consider a change point with independent or dependent error, or multiple change points with homoscedastic error. There are few change point papers (in comparison to other error structures) on change points with heteroscedastic errors. Those with heteroscedastic errors often assume that σ^2 is a continuous function of x or of $\lambda = \lambda(x)$, *e.g.* a Poisson intensity parameter used in a Poisson regression model (Kim, Fay, Feuer and Midthune, 2000); hence the changes in the error variance can be absorbed into the changes in the mean (regression) function. Similarly, Chen, Choi and Zhou (2008) utilize a wavelet smoother in nonparametric regression models with robust estimators for conditional heteroscedastic variance which are also functions of time, so do Loschi, Cruz, Takahashi, Iglesias, Arellano–Valle and MacGregor Smith (2008) to shifts in variance of financial time series data with an inverse gamma prior for variance. Killick, Eckley, Ewans and Jonathan (2010) study the change in variability of significant wave heights of storm peak events across the Gulf of Mexico for the period 1900–2005. Killick and Eckley (2010) provide R software for detecting changes in mean and variance in time series based on the pruned exact linear time (PELT) algorithm (Killick, Fearnhead and Eckley (2012)).

We consider the multiple change point problem with heteroscedastic errors which places no specific assumptions between σ_i^2 and $\mu(x)$ other than $0 < \sigma_0^2 \leq \sigma_i^2 < \infty$, it is bounded below by zero. Chiu, Lockhart and Routledge (2006) describe a similar model with homoscedastic error. In §2, our model is formally described. In §3, two likelihood solutions are provided for the three segmented curve. One places no constraint on the juncture of the segments (the unconstrained method), another imposes constraints on the join points (the constrained method). The estimates of polynomial regression parameters and change points have simple closed-form solutions under the unconstrained model. Their counter parts under the constrained model can be computed quickly using our constrained iterative reweighted least squares (CHIRLS) algorithm. A simple test for a middle segment to be a linear function vs. a quadratic function is also developed. In §4, two Bayesian solutions are derived for the segmented curve without continuity constraints at the change points, one is a classical Bayesian solution and another is a "hybrid" Bayesian technique, which we call Maximization-Maximization-posterior, or MMP, method. The classical Bayesian procedure is computationally intractable while the MMP method is. Comparisons among the two likelihood solutions and the MMP solution and application to real tire data sets are given in §4 and §5. Discussion is in §7 and some additional details or justifications are provided in the Appendix.

Although the motivation for this work stems from a specific application in the tire industry, our modeling and estimation procedures may be applied or generalized to other application

areas. They include process control in engineering, disease outbreaks studied in epidemiology, the effects of an intervention that is of interest in medicine, and the behavior of an economical indicator over time in finance.

2. MODEL DESCRIPTION

Let (x_i,y_i), $i = 1,\ldots, n$ be data. Two typical regression models of (1) are considered in detail: (1) all three segments are straight lines (**LLL model**), and (2) the first and third segments are straight lines and the second, or middle, segment is quadratic (**LQL model**). These two models work reasonably well for estimating FPL; the procedures developed for these models can be generalized to other segmented polynomial regression models.

For the LLL model, the model (1) is reduced to:

$$Y_i = (a_0+a_1x_i+\varepsilon_{1,i})\mathbf{1}_{[0,\tau1)}(x_i)+(b_0+b_1x_i+\varepsilon_{2,i})\mathbf{1}_{[\tau_1,\tau_2)}(x_i)+(c_0+c_1x_i+\varepsilon_{3,i})\mathbf{1}_{[\tau_2,x_n]}(x_i). \quad (2)$$

For the LQL model , a $b_2x_i^2$ term is added to the middle segment in (2). Furthermore the errors $\varepsilon_{l,i}$ are assumed to be from $N(0,\sigma_l^2)$ for l = 1,2,3. In the context of the tire application, it is reasonable to assume that the variance in the first and third segments would be equal to each other but different from the variance in the second segment. Thus in this paper we take $\sigma_1^2 = \sigma_3^2$. The procedures for cases that $\sigma_1^2, \sigma_2^2, \sigma_3^2$ are arbitrary or $\sigma_2^2 = \sigma_3^2$ can be developed in a similar fashion. For convenience let the model by parameterized in terms of *precision*, $\eta_i = 1/\sigma_i^2$ instead of variance σ_i^2, for i = 1,2.

For the LLL model the set of unknown parameters is $\theta = (\beta, \eta, \tau, \iota)$ with $\beta = (a_0,a_1, b_0,b_1,c_0, c_1)$ being the linear regression coefficients, $\eta = (\eta_1, \eta_2)$ the unknown precisions , $\tau = (\tau_1, \tau_2)$ the change points and $\iota = (j, k)$ the indices of the x–values at the change points.:

$$\begin{aligned} &x_1 < \ldots < x_j \le \tau_1 \\ \tau_1 < x_{j+1} < \ldots < x_k \le &\tau_2 \\ &\tau_2 < x_{k+1} < \ldots < x_n. \end{aligned} \quad (3)$$

Then the log-likelihood under the LLL model is easily seen to be,

$$l(\theta|y,x) = -\frac{n}{2}(2\pi) + \frac{n-k+j}{2}\log(\eta_1) + \frac{k-j}{2}\log(\eta_2) - \frac{\eta_1}{2}(\psi_1+\psi_3) - \frac{\eta_2}{2}\psi_2 \quad (4)$$

where

$$\psi_1 = \sum_{i=1}^{j}(y_i - a_0 - a_1x_i)^2, \psi_2 = \sum_{i=j+1}^{k}(y_i - b_0 - b_1x_i)^2, \psi_3 = \sum_{i=k+1}^{n}(y_i - c_0 - c_1x_i)^2 \quad (5)$$

are the *residual sum of squares* for the three linear segments respectively.

For the LQL model, the log-likelihood has the same expression as (4) except that β and ψ_2 need to be changed to

$$\beta = (a_0, a_1, b_0, b_1, c_0, c_1), \qquad \psi_2 = \sum_{i=j+1}^{k} (y_i - b_0 - b_1 x_i - b_2 x_i^2)^2.$$

Under the model described above, two cases can occur with respect to the placement of the change points in relation to the estimated straight line segments.

1. **Unconstrained case:** No restriction is placed on the intersections of the 3 lines (allowing for an *abrupt* or a smoothed change)

2. **Constrained case:** The change points occur at the 2 intersection points of the 3 lines (requiring changes to be *continuous*, i.e. $lim_{x\to\tau_i^+}\mu(x) = lim_{x\to\tau_i^-}\mu(x), i = 1,2$).

See an example of the resulting estimates in Figure 4 where $\mu(x)$ is the mean function in (2) with β and τ suppressed.

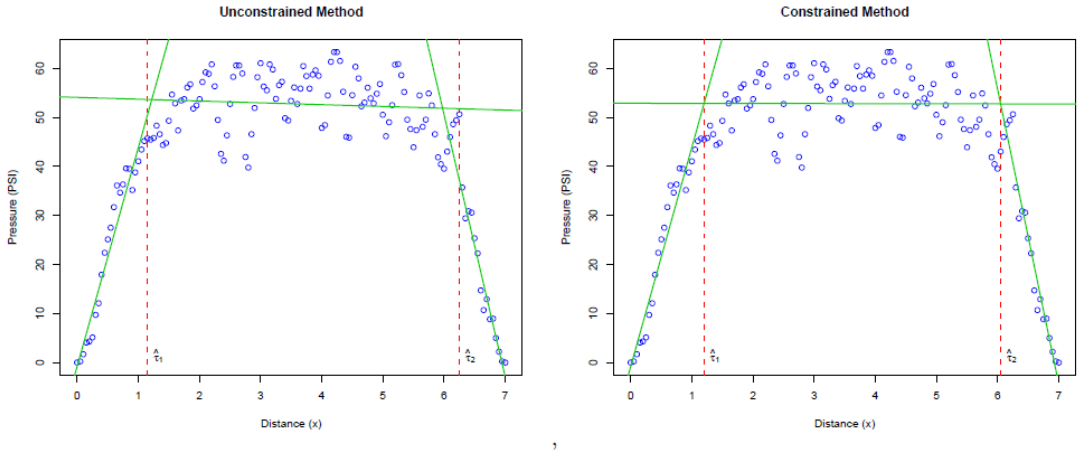


Figure 4: Footprint pressure curve with two fits

3. LIKELIHOOD ESTIMATION

3.1 Unconstrained Method

Let $\bar{u}^{(r,s)} = \frac{1}{s-r+1}\sum_{i=r}^{s} u_i$, $S_{uv}^{(r,s)} = \sum_{i=r}^{s}\left(u_i - \bar{u}^{(r,s)}\right)\left(v_i - \bar{v}^{(r,s)}\right)$, (6)

Proposition 3.1 *Under the unconstrained LLL model and the assumption that* $k - j > 3$, *for each* (j,k) *the MLE's of the regression coefficients* β *and precisions* η

are

$$\hat{a}_0 = \bar{y}^{(1,j)} - \hat{a}_1\bar{x}^{(1,j)},\ \hat{a}_1 = \frac{S_{xy}^{(1,j)}}{S_{xx}^{(1,j)}}$$
$$\hat{b}_0 = \bar{y}^{(j+1,k)} - \hat{b}_1\bar{x}^{(j+1,k)},\ \hat{b}_1 = \frac{S_{xy}^{(j+1,k)}}{S_{xx}^{(j+1,k)}}$$
$$\hat{c}_0 = \bar{y}^{(k+1,n)} - \hat{c}_1\bar{x}^{(k+1,n)},\ \hat{c}_1 = \frac{S_{xy}^{(j+1,k)}}{S_{xx}^{(j+1,k)}}$$
$$\hat{\eta}_1 = \frac{n-k+j}{\hat{\psi}_1+\hat{\psi}_3},\ \hat{\eta}_2 = \frac{k-j}{\hat{\psi}_2},\ \hat{\tau}_1 = x_j,\ \hat{\tau}_2 = x_k$$

(7)

where $\hat{\psi}_1 = \sum_{i=1}^{j}(y_i - \hat{a}_0 - \hat{a}_1 x_i)^2, \hat{\psi}_2 = \sum_{i=j+1}^{k}\left(y_i - \hat{b}_0 - \hat{b}_1 x_i\right)^2, \hat{\psi}_3 = \sum_{i=k+1}^{n}(y_i - \hat{c}_0 - \hat{c}_1 x_i)^2$, are the values of $(\hat{j}, \hat{k}) = arg\, max_{2<j<k-3<n-2}\, l(\hat{\beta}, \hat{\eta}, j, k)$.

Proposition 3.2 *For the unconstrained LQL model and the assumption that k – j* > 4*, the MLE's are*

$$\left(\hat{\beta}_0, \hat{\beta}_1, \hat{\beta}_2\right)' = (X'X)^{-1}X'Y, \qquad \hat{\psi}_2 = \sum_{i=j+1}^{k}\left(y_i - \hat{b}_0 - \hat{b}_1 x_i - \hat{b}_2 x_i^2\right)^2$$

where $Y = \left(y_{j+1}, \ldots, y_k\right)'$ *and*

$$X = \begin{pmatrix} 1 & x_{j+1} & x_{j+1}^2 \\ . & . & . \\ 1 & x_k & x_k^2 \end{pmatrix}.$$

Remark 1: Given fixed values of j and k, the values for $\hat{\beta} = \left(\hat{a}_0, \hat{a}_1, \hat{b}_0, \hat{b}_1, \hat{c}_0, \hat{c}_1\right)$ defined in (7) are simply the least squares estimates obtained by separate linear regression analyses of the three distinct straight line segments. That $k - j > 3$ for the LLL model ($k - j > 4$ for the LQL model) ensures that there are at least 3 data points for the LLL model (4 points for the LQL model) available to perform the regression in order to avoid singularity in the matrix inversion during the solution of the regression parameter estimates. These constraints on *k* and ***j*** are sufficient for the "identifiable condition" required by Feder (1975a).

3.2 Constrained Method

By imposing the continuity constraint in the second case the first change point, τ_1, must be common to the fitted lines of segments 1 and 2 and the second change point, τ_2, must be common to the fitted lines of segments 2 and 3, that is,

$a_0 + a_1 x_j = b_0 + b_1 x_j, \qquad b_0 + b_1 x_k = c_0 + c_1 x_k$

where $x_j = \tau_1$ and $x_k = \tau_2$, as the MLE's have to be at the data points.

Therefore imposing the continuity constraint has reduced β from 6 dimensional in the LLL model to 4 dimensional $\tilde{\beta} = (a_0, a_1, b_1, c_1)$, thus the parameter space is of 8 dimensional $\tilde{\theta} = (\tilde{\beta}, \eta, \tau)$. The log-likelihood in (4) is valid after ψ_2 and ψ_3 in (5) are replaced by:

$$\psi_2 = \sum_{i=j+1}^{k} \left[y_i - a_0 - a_1 x_j - (x_i - x_j) b_1 \right]^2 \tag{8}$$

$$\psi_3 = \sum_{i=k+1}^{n} \left[y_i - a_0 - a_1 x_j - (x_k - x_j) b_1 - (x_i - x_k) c_1 \right]^2. \tag{9}$$

After making these substitutions into the log-likelihood (4) and differentiating it with respect to the parameters $\tilde{\beta}, \eta$ we obtain the following system of six equations:

$$r_1 = w_{11} a_0 + w_{12} a_1 + w_{13} b_1 + w_{14} c_1 \tag{10}$$

$$r_2 = w_{12} a_0 + w_{22} a_1 + w_{23} b_1 + w_{24} c_1 \tag{11}$$

$$r_3 = w_{13} a_0 + w_{23} a_1 + w_{33} b_1 + w_{34} c_1 \tag{12}$$

$$r_4 = w_{14} a_0 + w_{24} a_1 + w_{34} b_1 + w_{44} c_1 \tag{13}$$

$$\eta_1 = \frac{n-k+j}{\psi_1+\psi_3} \tag{14}$$

$$\eta_2 = \frac{k-j}{\psi_2} \tag{15}$$

where

$$w_{pq} = w_{pq}(\eta_1, \eta_2), \quad rp = rp(\eta_1, \eta_2); \; p,q = 1,2,3,4 \tag{16}$$

are linear functions in the η_i's for fixed values of j and k, with their detailed expressions in the Appendix. Since the residual sum of squares terms in (14) and (15) are functions of a_0, a_1, b_1 and c_1 through (5), (8) and (9), $\eta_1 = \eta_1(a_0,a_1,b_1,c_1,k,j)$ and $\eta_2 = \eta_2(a_0,a_1,b_1,c_1,k,j)$ are non-linear functions. So this leads to a non-linear system of 6 equations with 6 unknowns given k and j.

Rather than solving this non–linear system with a brute force numerical method, we propose an iterative linear solution as in Barham and Drane (1972). Note that the first four equations (10)–(13) may be written: $W\tilde{\beta} = r$, which has a simple solution,

$$\tilde{\beta} = W^{-1} r, \tag{17}$$

given $\mathbf{W}$, η_1, η_2.

So, for fixed change–points *j, k*, our *constrained heteroscedastic iterative reweighted least squares (CHIRLS)* algorithm is:

Setup. Let $\gamma > 0$ be the desired precision in parameter estimation, and let j, k be fixed.

Choose Initial Values. Compute a regression fit from each of the three segments to obtain $\tilde{\beta}^{(0)}$, pool the variance estimates in segments 1 and 3 to obtain $\hat{\sigma}_1^0$. This $\hat{\sigma}_1^0$ with $\hat{\sigma}_2^0$ from segment 2 lead to initial values $\eta^{(0)} = (\eta_1^0, \eta_2^0)$. Using these values compute $\mathbf{W^{(0)}}$, $\mathbf{r^{(0)}}$ from (16).

Compute Regression Estimates. Repeat the following until

$$\left|\frac{\eta^{(m)} - \eta^{(m-1)}}{\eta^{(m-1)}}\right| < \gamma \text{ and } \left|\frac{\beta^{(m)} - \beta^{(m-1)}}{\beta^{(m-1)}}\right| < \gamma$$

component-wisely.

1. Compute $\eta_1^{(m)} = \eta_1^{(m)}(a_0, a_1, b_1, c_1)$ and $\eta_2^{(m)} = \eta_2^{(m)}(a_0, a_1, b_1, c_1)$ from (14) and (15).
2. Compute $\mathrm{W}^{(m)} = \mathrm{W}^{(m)}(\eta_1^{(m)}, \eta_2^{(m)})$ and $\mathrm{r}^{(m)} = \mathrm{r}^{(m)}(\eta_1^{(m)}, \eta_2^{(m)})$ from (16)
3. Solve $\tilde{\beta}^{(m)} = [\mathrm{W}^{(m)}]^{-1}\mathrm{r}^{(m)}$.
4. Increment $m \to m+1$.

Let $\hat{\beta}(j,k)$ and $\hat{\eta}(j,k)$ be the final output $\hat{\beta}(m)$ and $\eta^{(m)}$ from the above loop.

Compute Change Points. Find $(\hat{j}, \hat{k}) = \arg\max_{2<j\le k-3<n-2} l(\hat{\beta}(j,k), \hat{\eta}(j,k), j, k)$.

This strategy provides an algorithm which quickly converges to the solution.

A CHIRLS algorithm for the LQL solution can be similarly written out as that for the LLL model above.

Remark 2: The estimators in the unconstrained case are simpler but are not as efficient as the constrained estimators if the true model has continuous change points.

Remark 3: Siegmund and Zhang (1994) and Kim, Fay, Feuer and Midthune (2000) provide an alternative way of writing the constrained model described above. Using their notation our model becomes

$$E[y|x]\ \beta_0 + \beta_1 x + \delta_1 (x - \tau_1)^+ + \delta_2 (x - \tau_2)^+, \qquad (18)$$

where τ_1 and τ_2 are the change points which have to be estimated, $\beta_0, \beta_1, \delta_1$ and δ_2 are unknown regression parameters which also have to be estimated, and $u^+ = u$, for $u > 0$ and 0 otherwise. This is also the way the bent–cable model in Chiu, Lockhart and Routledge (2006) is described.

3.3 Model Selection

Zhang and Siegmund (2007) provide a modified Bayes information criterion (BIC) procedure for model selection in change point problems which is used not only to determine the number of change points but also the number of terms to be included in a change point model. However, since our case is simpler (two change points) we employ the following standard approach. Testing for the significance of the b_2 coefficient could be used to chose between the LLL and LQL models. Since we are working in a regression situation we can use the same type of hypothesis testing as is done in multiple regression analysis to assess the significance of added terms or the addition of higher order terms in polynomial regression. If we let the null hypothesis be that b_2 is equal to zero and the alternative hypothesis that it is nonzero, then an *approximate F–test* can be used. Formally,

$$H_0 : b_2 = 0, \qquad H_a : b_2 \neq 0.$$

A reasonable test statistic is then

$$F = \frac{SSR(H_0) - SSR(H_a)}{SSR(H_a)/(n-7)}$$

where $SSR(H0)$ and $SSR(Ha)$ are the residual sum of squares of the models postulated under $H0$ and Ha respectively. See the Appendix for further details.

Unfortunately, if $\sigma_1 \neq \sigma_2$, the null distribution of F is not even approximately $F_{1,n-7}$. However, its P-value can be approximated by simulation. A simpler test statistic is

$$F' = \frac{S_2'' - S_2'}{S_2'/(k-j-3)},$$

Which has an exact $F_{1,k-j-3}$ distribution if k' and j' are given, and has an approximate $F_{1,k-j-3}$ distribution if k' and j' are to be estimated. Here S_2'' is the same as S_2 but with $k = k$' and $j = j$' and S_2' (estimated under H_a) is given in the Appendix.

4 BAYESIAN METHODS

Most papers solve the change point problem using either the likelihood approach or a Bayesian approach. In this section we also provide a Bayesian solution and its variant and do a compariative study in §5. In the change point regression problem the linear regression

coefficients, β, the two variances σ_1^2 and σ_2^2 and the two change points τ_1 and τ_2 can be treated as having distributional properties of their own. Furthermore, an alternative to working with τ_1 and τ_2 themselves is to consider the estimation of the indices of the *x*-values which correspond to these values, under suitable conditions which can be described by $x_j = \tau_1$ and $x_k = \tau_2$. Thus we can work with the indices, say *j* and *k*, of the *x*-values. A similar approach can be found in Chen and Gupta (2000) in their analysis of simple linear regression models which contain a single point. In this section we provide two Bayesian procedures for the LLL model.

4.1 A Formal Procedure

For the analysis we assign general priors to all the unknown parameters (β, η, j, k). Let $\pi_0(\cdot)$ be the vague notation for prior density. Let *C* stand for an arbitrary finite constant,

$$\pi_0(a_0, a_1, b_0, b_1, c_0, c_1) \propto C, \quad -\infty < a_0, a_1, b_0, b_1, c_0, c_1 < \infty;$$

$$\pi_0 \propto \frac{1}{\sigma_1^2}, 0 < \sigma_1^2 < \infty; \ \pi_0 \propto \frac{1}{\sigma_2^2}, 0 < \sigma_2^2 < \infty;$$

$$\pi_0 \propto \frac{2}{n(n-1)}, 2 < j < k-2 < n-4.$$

Note that the priors for β are improper.

Proposition 4.1 *Given the above prior distributions the posterior distribution of the changepoint indices* (j, k) *derived by standard Bayesian treatment can be shown to be,*

$$\pi_1(j,k) \propto \left[j(k-j)(n-k)S_x^{(1,j)}S_x^{(j+1,k)}S_x^{(k+1.n)}\right]^{-\frac{1}{2}} D_1^{-m_1} D_2^{-m_2} B(m_1, m_2),$$

where $B(m_1, m_2)$ *denotes the Beta function and* D_1 *and* D_2 *are functions of the residual sums of squares,* $S_u^{(a,b)}$, $for\ \mathcal{R} = \{(j,k): 2 < j < k-2 < n-3\}$. *See the Appendix for the definition of* D_1 *and* D_2.

We then say that a change point is located at $(\hat{j}, \hat{k})$ if $\pi_1(\hat{j}, \hat{k}) = max_{\mathcal{R}}\pi_1(j,k)$. This is a

MAP (Maximize A Posterior) estimate. The Bayes estimate with a quadratic loss function is $E_{\pi 1}[(j,k)]$, the posterior expectation of j and k, which in this case is computationally intractable.

4.2 MMP Procedure

Instead of the vague priors for the precision parameters η_1 and η_2 suppose we assume that they come from Gamma $Ga(a,b)$ densities. This is reasonable since $\eta_i = 1/\sigma_i^2$ and the conjugate prior for the variance from a normal distribution is an Inverse-Gamma in Bayesian analysis. Accordingly, the priors for η_1, η_2 and (j,k) are,

$$\eta_1 \sim Ga(r,s),\ r,s>0, \tag{19}$$

$$\eta_2 \sim Ga(v,w),\ v,w>0, \tag{20}$$

$$(j,k) \sim \frac{2}{n(n-1)}. \tag{21}$$

where r, s, v and w are given and $2 < j < k-2 < n-4$.

This prior for the indices of the change points, (21), is not uniform in terms of distance but rather with respect to the set of possible points from which they may be drawn. Thus it is a *data–dependent* prior which can be considered "unorthodox" in a strict Bayesian sense. If we consider an alternative "orthodox" prior which is uniform on a finite interval, and if the x_i are uniformly distributed on that interval, then the limiting behavior of the posterior distribution of our method will be the same as one using the orthodox uniform prior on the continuous interval.

From the priors (19)–(21), we obtain the joint distribution of the parameters and the sample

$$f(\beta,j,k,\eta_1,\eta_2;y,x) \propto \eta_1^{\frac{n-k+j+2r-2}{2}} \eta_2^{\frac{k-j+2v-2}{2}} exp\left\{-\frac{\eta_1}{2}(2s+\psi_1+\psi_3) - \frac{\eta_2}{2}(2w+\psi_2)\right\} \tag{22}$$

First maximization: It is clear that

$$max_\beta f(\beta,j,k,\eta_1,\eta_2;y,x) = f(\hat{\beta},j,k,\eta_1,\eta_2;y,x) \equiv \tilde{f}(j,k,\eta_1,\eta_2|y,x)$$

where $\hat{\beta} = \hat{\beta}(j,k) = \left(\hat{a}_0,\hat{a}_1,\hat{b}_0,\hat{b}_1,\hat{c}_0,\hat{c}_1\right)$ is the least squares estimators of the regression coefficients in (7). In the following, we shall work with $\tilde{f}$, the (max-β) posterior distribution of j, k, η_1, η_2.

Second maximization: From (22) the following conditional posterior distributions can be derived:

$$p_1(j,k|y,x,\eta_1,\eta_2) \propto \left(\frac{\eta_2}{\eta_1}\right)^{\frac{k-j}{2}} exp\left\{-\frac{\eta_1}{2}\left(\hat{\psi}_1+\hat{\psi}_3\right) - \frac{\eta_2}{2}\hat{\psi}_2\right\} \equiv h(j,k|y,x,\eta_1,\eta_2),$$

$$p_2(\eta_1, \eta_2 | y, x, j, k) \propto \eta_1^{\frac{n-k+j+2r-2}{2}} exp\left\{-\eta_1\left(s + \frac{\hat{\psi}_1 + \hat{\psi}_3}{2}\right)\right\} - \eta_2^{\frac{k-j+2v-2}{2}} exp\left\{-\eta_2\left(w + \frac{\hat{\psi}}{2}\right)\right\}$$

$$\equiv (\eta_1^{\alpha} exp\{-\gamma\eta_1\})(\eta_2^{\delta} exp\{-\xi\eta_2\})$$

where

$$\alpha = \frac{n-k+j+2r-2}{2}, \gamma = s + \frac{\hat{\psi}_1 + \hat{\psi}_3}{2}, \delta = \frac{k-j+2v+2}{2}, \xi = w + \frac{\hat{\psi}_2}{2}.$$

To solve for the unknown change point parameters j, k, η_1 and η_2 we propose to maximize $p_1(j,k|\ldots)$ and $p_2(\eta_1,\eta_2|\ldots)$ iteratively,

$$max_{(j,k)} h(j, k | y, x, \eta_1, \eta_2), \ k > j + 2, \quad (23)$$

$$max_{(\eta_1,\eta_2)} (\eta_1^{\alpha} exp\{-\gamma\eta_1\})\left(\eta_2^{\delta} exp\{-\xi\eta_2\}\right). \quad (24)$$

Since this process involves maximizing both conditional (max-β) posterior distributions it is called the *Maximization-Maximization-Posterior*, or MMP, method.

Optimization of (24) by simultaneously optimizing its two components is straightforward; we have

$$\hat{\eta}_1 = \frac{\alpha}{\gamma}, \qquad \hat{\eta}_2 = \frac{\delta}{\xi}$$

Optimization of (23) is quite fast with good initial values of j, k.

The Gibbs sampler, a popular way to analyze hierarchical data, was also applied to our problem. However attempts to use Gibbs sampling on the 3 straight line segment change point regression model were unsuccessful; it did not converge.

Alternatively priors may be used for the unknown parameters. For example, the regression coefficients (a_0, a_1, b_0, b_1, c_0, c_1) may be assigned normal priors. Conjugate priors could be given to the variance components, σ_i^2, and the change points τ_1 and τ_2 could be considered as being uniform in an interval $[a, b]$, thus involving the Dirichlet distribution. With these proper priors, it's possible to show the posterior consistency. We conjecture that the MMP procedure is asymptotically equivalent to the above proper Bayesian procedure as $n \to \infty$ when the normal prior for β is close to a data dependent prior implied by the MMP.

5. SIMULATION

In order to compare the unconstrained and constrained methods of maximum likelihood estimation and the Bayesian MMP method 3 simulations were run. Within each experiment 1,000 independent data sets consisting of 60, 120 and 240 data values based on change points $\tau_1 = 1$ and $\tau_2 = 5$ were generated. The data were generated from equation (2) using the parameter values,

$$\sigma_1^2 = 2, \sigma_2^2 = 5. a_0 = 0, a_1 = 20, b_0 = 20, b_1 = 0, c_0 = 120, c_1 = -20.$$

Additionally, for the MMP method it was assumed that $\frac{1}{\sigma_1^2} = \eta_1 \sim Ga(10{,}40)$ and $\frac{1}{\sigma_2^2} = \eta_2 \sim Ga(10{,}80)$.

The tabulated results from the simulation, Tables 1 – 3, show that, in general, as n increases the estimated parameter value approaches the true value. These values also indicate that the constrained method is superior to the other two in estimating the parameters. The MMP outperformed the other two at estimating both σ_1^2 and σ_2^2 in the 60 point simulation and at estimating σ_2^2 in the larger simulations. They also hint that the unconstrained method and MMP method perform very similarly as n increases. However, the MSE's for the regression coefficients and the change points give a slight edge in performance to the unconstrained method compared to the MMP method with increasing n. 90% and 95% confidence intervals for the change-point parameters $\tau_1$and τ_2 can be obtained from the values given in Table 4.

Table 1: Comparison of unconstrained, constrained and MMP estimation – 60 points

Parameter		Unconstrained			Constrained			MMP		
(Initial Value)		Mean	SEM	MSE	Mean	SEM	MSE	Mean	SEM	MSE
σ_1^2	(2)	1.465	0.018	0.778	1.622	0.019	0.702	2.590	0.009	**0.650**
σ_2^2	(5)	4.715	0.035	1.149	4.832	0.035	1.106	5.629	0.022	**0.935**
a_0	(0)	0.123	0.028	0.881	0.040	0.026	**0.838**	0.129	0.027	0.877
a_1	(20)	19.636	0.059	1.894	19.966	0.050	**1.592**	19.661	0.054	1.739
b_0	(20)	19.871	0.033	**1.051**	20.024	0.034	1.060	19.864	0.034	1.070
b_1	(0)	0.022	0.010	0.327	0.004	0.010	**0.326**	0.022	0.010	0.332
c_0	(120)	118.64	0.340	10.82	120.46	0.298	**9.427**	118.68	0.326	10.40
c_1	(-20)	-19.766	0.061	1.930	-20.081	0.054	**1.695**	-19.771	0.058	1.862
τ_1	(1)	1.008	0.005	0.158	1.005	0.002	**0.068**	1.054	0.005	0.153
τ_2	(5)	4.911	0.005	0.178	4.997	0.002	**0.069**	4.864	0.005	0.197

Notes: The smallest MSE in each row is in **bold** font, indicating the winning method in that case.

Table 2: Comparison of unconstrained, constrained and MMP estimation – 120 points

Parameter		Unconstrained			Constrained			MMP		
(Initial Value)		Mean	SEM	MSE	Mean	SEM	MSE	Mean	SEM	MSE
σ_1^2	(2)	1.720	0.014	0.519	1.90	0.013	**0.475**	2.404	0.009	0.491
σ_2^2	(5)	4.881	0.026	0.823	4.924	0.026	0.811	5.400	0.020	**0.748**
a_0	(0)	0.098	0.020	0.643	0.038	0.020	**0.623**	0.110	0.020	0.653
a_1	(20)	19.763	0.039	1.263	19.969	0.036	**1.145**	19.735	0.040	1.282
b_0	(20)	19.923	0.023	0.737	20.012	0.022	**0.710**	19.920	0.024	0.754
b_1	(0)	0.009	0.007	0.227	0.001	0.007	**0.218**	0.009	0.007	0.232
c_0	(120)	119.187	0.235	7.467	120.412	0.210	**6.645**	119.022	0.238	7.571
c_1	(-20)	-19.858	0.042	1.339	-20.072	0.038	**1.198**	-19.828	0.043	1.358
τ_1	(1)	1.032	0.004	0.118	1.002	0.001	**0.044**	1.057	0.004	0.127
τ_2	(5)	4.940	0.004	0.132	5.000	0.001	**0.045**	4.912	0.004	0.147

Table 3: Comparison of unconstrained, constrained and MMP estimation – 240 points

Parameter		Unconstrained			Constrained			MMP		
(Initial Value)		Mean	SEM	MSE	Mean	SEM	MSE	Mean	SEM	MSE
σ_1^2	(2)	1.841	0.010	0.349	1.885	0.010	**0.324**	2.241	0.008	0.344
σ_2^2	(5)	4.963	0.018	0.573	4.964	0.018	0.567	5.242	0.106	**0.557**
a_0	(0)	0.056	0.014	0.451	0.013	0.014	**0.438**	0.058	0.014	0.454
a_1	(20)	19.833	0.027	0.870	19.976	0.025	**0.792**	19.829	0.027	0.875
b_0	(20)	19.945	0.106	0.517	19.990	0.015	**0.489**	19.938	0.016	0.523
b_1	(0)	0.009	0.005	0.162	0.006	0.005	**0.153**	0.009	0.005	0.165
c_0	(120)	119.420	0.158	5.032	119.982	0.142	**4.485**	119.399	0.161	5.123
c_1	(-20)	-19.897	0.028	0.906	-19.996	0.026	**0.811**	-19.894	0.029	0.922
τ_1	(1)	1.027	0.003	0.091	1.001	0.001	**0.031**	1.043	0.003	0.097
τ_2	(5)	4.961	0.003	0.094	4.998	0.001	**0.031**	4.946	0.003	0.102

Table 4: Simulation study confidence intervals for the change-point parameters, τ_1 and τ_2. True values are $\tau_1 = 1$ and $\tau_2 = 5$, respectively.

Method	Points	Parameter	2.5%	5%	Mean	95%	97.5%
Unconstrained	60	τ_1	0.70	0.70	1.01	1.20	1.30
		τ_2	4.60	4.70	4.91	5.20	5.20
	120	τ_1	0.80	0.80	1.03	1.20	1.25
		τ_2	4.70	4.75	4.94	5.15	5.15
	180	τ_1	0.83	0.88	1.03	1.18	1.20
		τ_2	4.80	4.83	4.96	5.10	5.15
Constrained	60	τ_1	0.90	0.90	1.01	1.10	1.10
		τ_2	4.90	4.90	5.00	5.10	5.10
	120	τ_1	0.90	0.95	1.00	1.05	1.10
		τ_2	4.90	4.95	5.00	5.05	5.10
	180	τ_1	0.95	0.95	1.00	1.05	1.08
		τ_2	4.95	4.95	5.00	5.05	5.05
MMP	60	τ_1	0.70	0.80	1.05	1.30	1.30
		τ_2	4.60	4.60	4.86	5.10	5.20
	120	τ_1	0.80	0.85	1.06	1.25	1.25
		τ_2	4.70	4.70	4.91	5.10	5.15
	180	τ_1	0.85	0.88	1.04	1.18	1.20
		τ_2	4.78	4.80	4.95	5.10	5.13

A box and whisker plot, shown in Figure 5, reveals the superiority of the constrained method in obtaining estimates of the change point parameters τ_1 and τ_2. It also confirms the similarity in performance of the unconstrained and MMP methods, especially as the number of data points increased. Dotplots for bias (Figure 6) and MSE (Figure 7) for all the simulation parameters provide evidence to prefer the constrained method. The dotplots suggest that the constant, c_0, in the third straight line segment appears to be an outlier in many of the plots. However, this parameter was much larger in absolute value than the other 7 in the simulation and should have a larger bias and MSE.

Remark 4: In a preliminary simulation study the parameters for the Gamma distributions for η_1 and η_2 were set to values such that $E(\eta_1)$ and $E(\eta_2)$ were the exact inverse of the known values of σ_1 and σ_2. Under these ideal conditions for the MMP algorithm, it slightly outperformed the unconstrained method in estimating the regression coefficients and the change points with respect to MSE. In summary, if the MMP starts at the values close to the true values, it is similar to the unconstrained likelihood procedure.

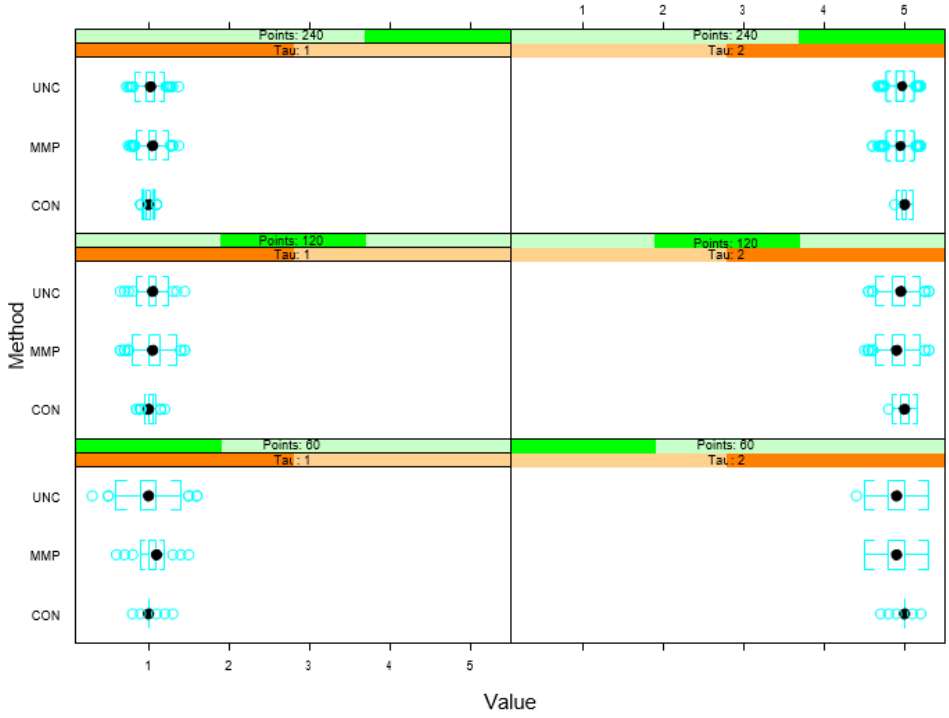


Figure 5: Box and Whisker Plots for τ_1 and τ_2

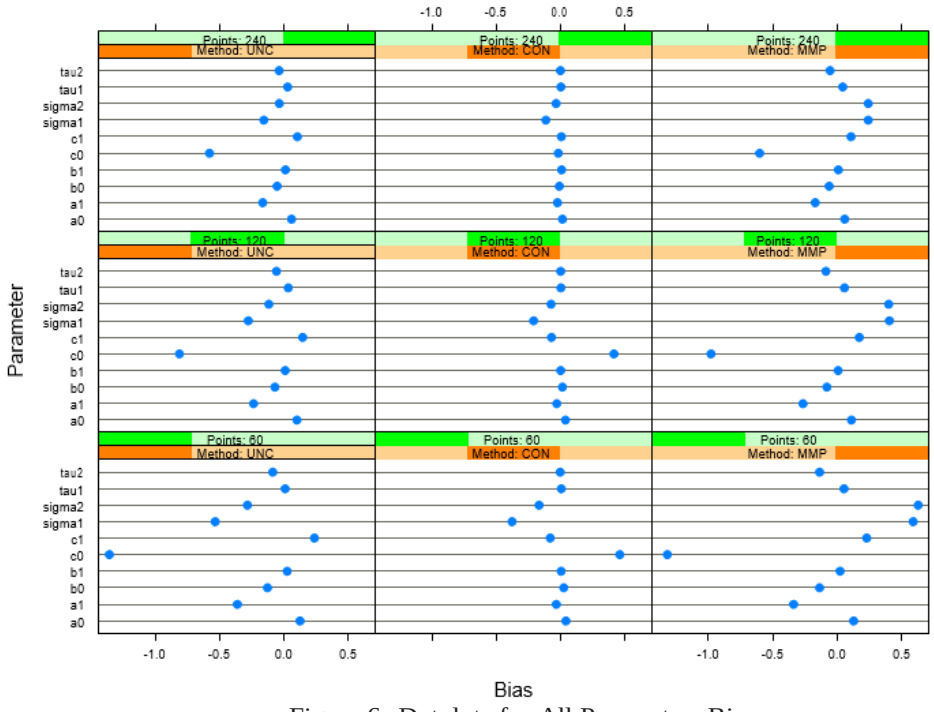


Figure 6: Dotplots for All Parameters Bias

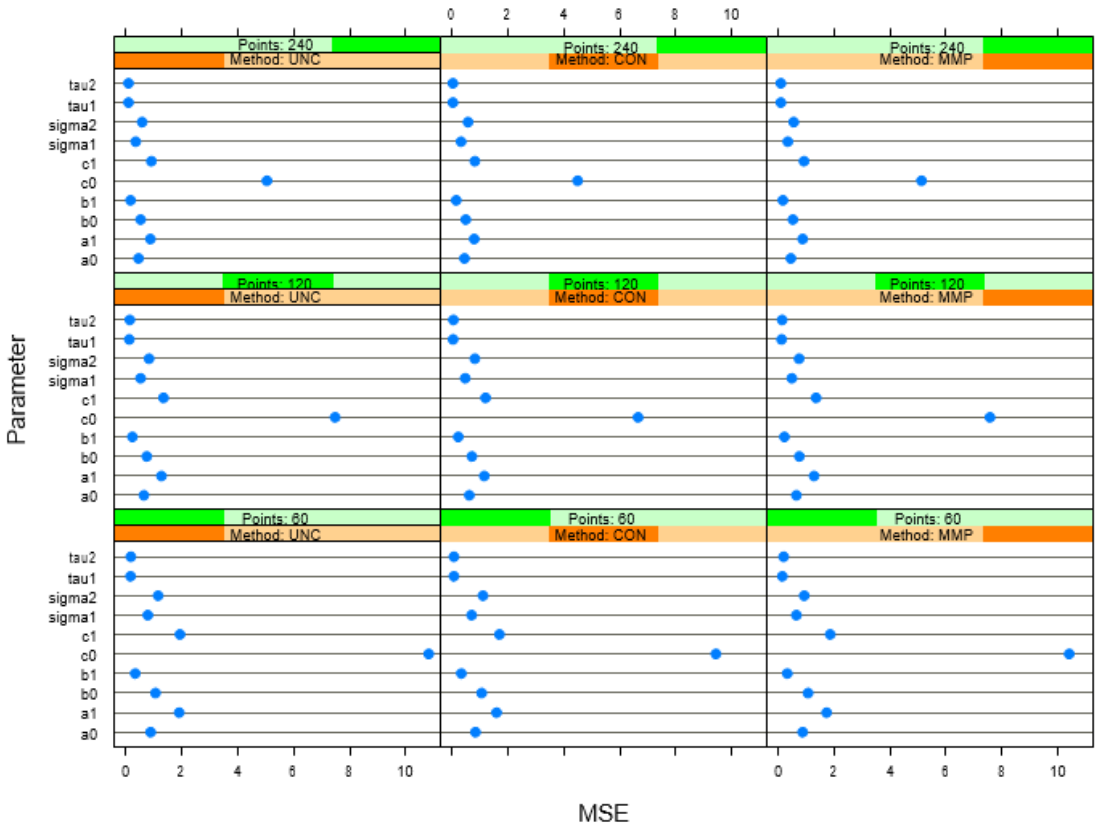


Figure 7: Dotplots for All Parameters MSE

6. EXAMPLES

In this section the methods are applied to two real data sets, corresponding to 2 typical footprints. Their footprint pressure curves are shown in Figure 2.

Results obtained by the four methods, unconstrained LLL (UNC), constrained LLL (LLL), extended constrained (LQL) and maximization–maximization–posterior (MMP) for the LLL fit are presented in Table 5.

Based on the tabulated values the best candidate should be either the LLL model or the LQL for the first FPC (Example 1). The reason for this is that they both provide change points which account for a greater footprint length (6.25 in.) than those estimated by the unconstrained and MMP methods (5.25 in.). The latter 2 models do not pick up a sufficient amount of the top and bottom portions of the footprint, an observation that requires engineering judgment. In choosing between the LLL and LQL

model the diagnostic test declares b_2 (p–value ≈ 1) not significant, therefore the final model selected is

the LLL. The solution to Example 1, illustrated by the first graph in Figure 8, shows the characteristic feature of the LLL method whereby the join points of the straight lines intersect with the estimated change points $\hat{\tau}_1$ and $\hat{\tau}_2$.

Results for the second footprint pressure curve (Example 2) indicate that the LQL model provides a slightly larger footprint length (5.45 in.) than the other 3 models. A test for significance of the b_2 term (F = 16.40, p–value < 0.001, n = 141) indicates that it is significant. A graphical representation of the LQL solution superimposed on the footprint pressure curve is shown by the third graph in Figure 8.

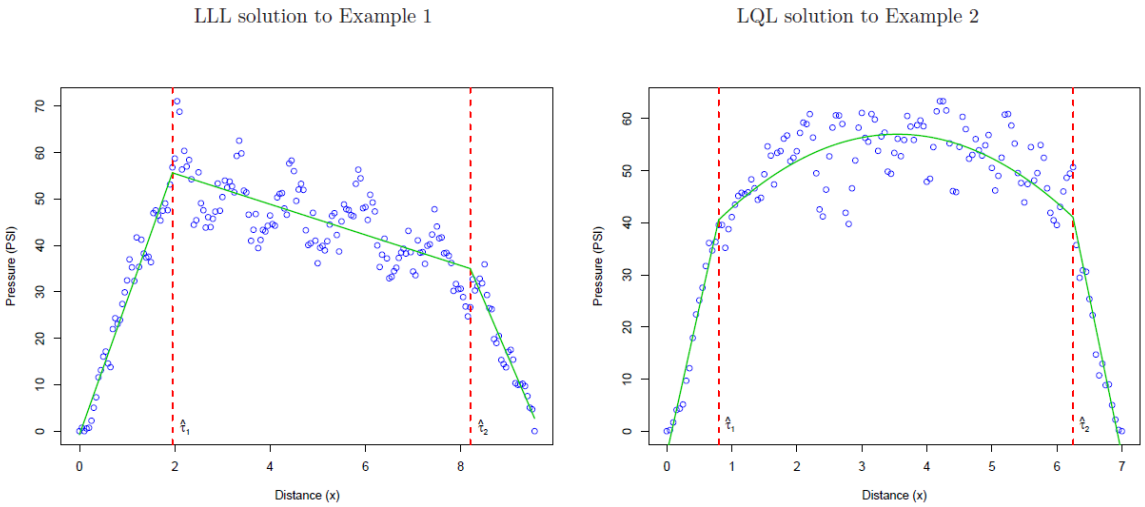


Figure 8: Graphs showing LLL/LQL solutions

7. DISCUSSION

We have provided a solution to a specific physical problem that occurs in the tire manufacturing industry, i.e. finding the change points in a footprint pressure curve. Solutions based on both maximum likelihood and Bayesian methodology were derived. A model selection procedure was established for discriminating between models in which the middle segment of the curve was either linear or quadratic. A new Bayesian-type procedure, the maximum–maximum–posterior was proposed. A simulation study indicates that applying the continuity constraint provides a better method of estimation than either the unconstrained method or the MMP procedure when the underlying model is continuous.

Analysis of two real data sets shows that a single model is insufficient to describe all possible footprint pressure curves. It also emphasizes the importance of using good engineering judgment to analyze physical problems.

We omitted asymptotic analysis in this paper. Interested readers can find the consistency results of our constrained and unconstrained estimators in Ganocy (2003).

The methods proposed in this paper can be extended further to models consisting of more than two change points, with polynomial segments and heteroscedastic errors.

APPENDIX

Proof of Propositions 3.1 and 3.2

Note that the maximum likelihood estimates (MLE) of the change points are at the data points under fairly general conditions (Worsley, 1986; Hinkley, 1971). This is reasonable because (1) there is no additional information between the data points without further assumptions; and (2) as the sample size $n \to \infty$, x_j, the change-point solution is consistent to the change points under some regularity conditions (Feder, 1975a).

For fixed τ, i.e. fixed (j,k) in (3) it is clear from the likelihood function (4) that the MLE's of (β, η) are the respective least squares solutions in each of the 3 segments under the condition that $\eta_1 = \eta_3$ are the same. Next, insert $\hat{\beta}(j,k), \hat{\eta}(j,k)$ and $\hat{\tau}(j,k)$ into the likelihood function (4) and maximize over all possible values of j and k to obtain the MLE's of $\hat{j}$ and $\hat{k}$. Hence the MLE's of β and η are $\hat{\beta}(\hat{j},\hat{k})$ and $\hat{\eta}(\hat{j},\hat{k})$.

Proof of Proposition 3.2 follows in a similar fashion to that of Proposition 3.1.

Likelihood derivation of the elements of W and r

By substituting the constraints into the log likelihood, taking derivatives and setting the results equal to zero we ultimately obtain the elements of matrix W,

$$w_{11} = (n-k+j)\eta_1 + (k-j)\eta_2$$

$$w_{12} = \eta_1\left[\sum_{i=1}^{j} x_i + (n-k)x_j\right] + \eta_2(k-j)x_j$$

$$w_{13} = \eta_1(n-k)(x_k - x_j) + \eta_2\sum_{i=j+1}^{k}(x_i - x_j)$$

$$w_{14} = \eta_1\sum_{i=k+1}^{n}(x_i - x_k)$$

$$w_{22} = \eta_1\sum \quad w_{22} = \eta_1\left[\sum_{i=1}^{j} x_i^2 + (n-k)x_j^2\right] + \eta_2(k-j)x_j^2$$

$$w_{23} = \eta_1(n-k)x_j(x_k - x_j) + \eta_2 x_j\sum_{i=j+1}^{k}(x_i - x_j)$$

$$w_{24} = \eta_1 x_j\sum_{i=k+1}^{n}(x_i - x_k)$$

$$w_{33} = \eta_1(n-k)(x_k - x_j)^2 + \eta_2 \sum_{i=j+1}^{n} (x_i - x_j)^2$$

$$w_{34} = \eta_1(x_k - x_j) \sum_{i=k+1}^{n} (x_i - x_k)$$

$$w_{44} = \eta_1 \sum_{i=k+1}^{n} (x_i - x_k)^2$$

and vector **r**,

$$r_1 = \eta_1 \left(\sum_{i=1}^{j} y_i + \sum_{i=k+1}^{n} y_i \right) + \eta_2 \sum_{i=j+1}^{k} y_i$$

$$r_2 = \eta_1 \left(\sum_{i=1}^{j} x_i y_i + x_j \sum_{i=k+1}^{n} y_i \right) + \eta_2 x_j \sum_{i=j+1}^{k} y_i$$

$$r_3 = \eta_1(x_k - x_j) \sum_{i=k+1}^{n} y_i + \eta_2 \sum_{i=j+1}^{k} y_i (x_i - x_j)$$

$$r_1 = \eta_1 \sum_{i=k+1}^{n} (x_i - x_k) y_i.$$

Model selection F-test

In order to construct the F-test first calculate

$$S_1 = \sum_{i=1}^{j} \left(y_i - \hat{a}_0 - a_1 x_i \right)^2,\ S_2 = \sum_{i=j+1}^{k} \left(y_i - \hat{b}_0 - b_1 x_i \right)^2,\ \ S_3 = \sum_{i=k+1}^{n} \left(y_i - \hat{c}_0 - c_1 x_i \right)^2$$

Based on the model under H_0, and

$$S_1' = \sum_{i=1}^{j'} \left(y_i - \tilde{a}_0 - \tilde{a}_1 x_i \right)^2,\ S_2' = \sum_{i=j+1}^{k'} \left(y_i - \tilde{b}_0 - \tilde{b}_1 x_i - \tilde{b}_2 x_i^2 \right)^2,\ \ S_3' = \sum_{i=k+1}^{n'} \left(y_i - \tilde{c}_0 - \tilde{c}_1 x_i \right)^2$$

for the model under H_a. The individual parameter estimates will have different values depending on which model is used to calculate them, hence the use of the notation $\widehat{\boldsymbol{\theta}}$ (estimates under H_0) and $\widetilde{\boldsymbol{\theta}}$ (estimates under H_a) to distinguish between the two.

From these values let us define:

$$SSR(H_0) = S_1 + S_2 + S_3,\ \ SSR(H_a) = S_1' + S_2' + S_3'.$$

It follows then that,

$$F = \frac{SSR(H_0) - SSR(H_a)}{SSR(H_0)/(n-7)}$$

provides the required F-test.

Standard Bayesian derivation of $\pi_1(j,k)$

Note that the following constitutes the proof of Proposition 4.1.

Using the precision notation in place of variance the joint distribution of parameters and sample is

$$f(\beta, j, k, \eta_1, \eta_2; y, x) = \left(\frac{1}{2\pi}\right)^{n/2} \eta_1^{(n-k+j)/2} \eta_2^{(k-j)/2} \exp\left\{-\frac{\eta_1}{2}(\psi_1+\psi_3) - \frac{\eta_2}{2}\psi_2\right\},$$

(25)

which leads to the joint posterior

$$\pi_1(\beta, j, k, \eta_1, \eta_2 \mid y, x) \propto \eta_1^{(n-k+j-2)/2} \eta_2^{(k-j-2)/2} \exp\left\{-\frac{\eta_1}{2}(\psi_1+\psi_3) - \frac{\eta_2}{2}\psi_2\right\}$$

(26)

Now the joint posterior distribution for the change point indices may be written,

$$\pi_1(j,k) \propto \int_0^\infty \int_0^\infty \int_{-\infty}^\infty \int_{-\infty}^\infty \int_{-\infty}^\infty \int_{-\infty}^\infty \int_{-\infty}^\infty \int_\infty^\infty \eta_1^{\frac{n-k+j-2}{2}} \eta_2^{\frac{k-j-2}{2}}$$
$$\cdot exp\left\{-\tfrac{\eta_1}{2}(\psi_1+\psi_3) - \tfrac{\eta_2}{2}\psi_2\right\} da_0 da_1 db_0 db_1 dc_0 dc_1 d\eta_1 d\eta_2. \qquad (27)$$

Denoting the multiple integral on the right hand side of (27) by I and utilizing the fact that it is a product of normal and gamma integrals, it reduces to,

$$I = \frac{(2\pi)^3}{\sqrt{j(k-j)(n-k)S_x^{(1,j)}S_x^{(j+1,k)}S_x^{(k+1,n)}}} \int_0^\infty \eta_1^{(n-k+j-6)/2} exp\left\{-\frac{D_1}{2}\eta_1\right\} d\eta_1$$
$$\cdot \int_0^\infty \eta_2^{(k-j-4)/2} exp\left\{-\tfrac{D_2}{2}\eta_2\right\} d\eta_2,$$

(28)

where $S_x^{(r,s)} = S_{xx}^{(r,s)}$ as defined in (6) and,

$$D_1 = S_y^{(1,j)} + S_y^{(k+1.n)} - (\hat{a}_1)^2 S_x^{(1,j)} - (\hat{c}_1)^2 S_x^{(k+1.n)} = \hat{\psi}_1 + \hat{\psi}_3,$$
$$D_2 = S_y^{(j+1,k)} - \left(\hat{b}_1\right)^2 S_x^{(j+1.k)} = \hat{\psi}_2,$$

with $\hat{a}_1, \hat{b}_1$ and $\hat{c}_1$ being the standard least squares estimates of a_1, b_1 and c_1 respectively as defined in (7) given j and k. Furthermore we note that the two integrals involving η_1 and η_2 are of the form,

$$\int_0^\infty u^m exp\{-\lambda u\} du = \frac{\Gamma(m+1)}{\lambda^{m+1}}, \qquad \lambda > 0, m > -1.$$

Since we need a minimum of 3 points to determine the parameters in each straight line

segment, 2 for the regression coefficients and 1 for the precision, we must have $2 < j < k - 2 < n - 3$. Setting $m_1 = (n - k + j - 4)/2$ and $m_2 = (k - j - 2)/2$,

$$\int_0^\infty \eta_1^{m_1 - 1} exp\left\{-\frac{D_1}{2}\eta_1\right\} d\eta_1 = \frac{\Gamma(m_1)}{\left(\frac{D_1}{2}\right)^{m_1}},$$

and

$$\int_0^\infty \eta_2^{m_2 - 1} exp\left\{-\frac{D_2}{2}\eta_2\right\} d\eta_2 = \frac{\Gamma(m_2)}{\left(\frac{D_2}{2}\right)^{m_2}}.$$

Inserting these results in equation (28) and noting $m_1 + m_2 = (n/2) - 3$ is independent of j and k, we obtain the expression for the posterior of the change point indices (j, k) as,

$$\pi_1(j,k) \propto \left[j(k-j)(n-k)S_x^{(1,j)}S_x^{(j+1,k)}S_x^{(k+1,n)}\right]^{-\frac{1}{2}} D_1^{-m_1} D_2^{-m_2} B(m_1, m_2),$$

where $B(m_1, m_2) = \Gamma(m_1)\Gamma(m_2)/\Gamma(m_1 + m_2)$ denotes the Beta function for $\mathcal{R} = \{(j,k): 2 < j < k - 2 < n - 3\}$.

References

[1] Arani, R.B., Soong, S.–J., Weiss, H.L., Wood, M.J., Fiddian, P.A., Gnann, J.W. and Whitley, R. (2001), "Phase specific analysis of herpes zoster associated pain data: A new statistical approach," *Statistics in Medicine*, **20**, 2429–2439.

[2] Barham, R. H. and Drane, W. (1972), "An algorithm for least squares estimation of nonlinear parameters when some of the parameters are linear," *Technometrics*, **14**, 757–766.

[3] Bellman, R. and Roth, R. (1969), "Curve fitting by segmented straight lines," *Journal of the American Statistical Association*, **64**, 1079–1084.

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

[5] Brodsky, B. E. and Darkhovsky, B. S. (1993), *Nonparametric Methods in Change-point Problems*, Mathematics and its Applications, 243. Kluwer Academic Publishers Group, Dordrecht.

[6] Carlstein, E., Mu¨ller, H.–G. and Siegmund, D., eds. (1994), *Change–point Problems, volume 23 of IMS Lecture Notes–Monograph Series*, Hayward, CA. Institute of Mathematical Statistics.

[7] Chen G., Choi, Y.K. and Zhou, Y. (2008), "Detection of changes in return by a wavelet smoother with conditional heteroscedastic volatility," *Journal of Econometrics*, **143**, 227–262.

[8] Chen, J. and Gupta, A. K. (2000), *Parametric Statistical Change Point Analysis*, Boston: Birkh¨auser.

[9] Cheng, M.Y., Raimondo, M. (2008), "Kernel methods for optimal change-points estimation in derivatives," *Journal of Computer Graphics and Statistics*. **17**, 56–75.

[10] Chiu, G., Lockhart, G. and Routledge, R. (2006), "Bent–cable regression theory and applications," *Journal of the American Statistical Association*, **101**, 542–553.

[11] Chow, G.C. (1960), "Tests of equality between sets of coefficients in two linear regressions," *Econometrica*, **28**, 591–605.

[12] Cohen, P., ed. (2008), *Applied Data Analytic Techniques for Turning Points Research*, New York: Routledge.

[13] Cs¨org¨o, M. and Horv´ath, L. (1997), *Limit Theorems in Change–Point Analysis*, New York: John Wiley & Sons.

[14] Desmond, R.A., Weiss, H.L., Arani, R.B., Soong, S.–J., Wood, M.J., Fiddian, P.A., Gnann, J.W. and Whitley, R.J. (2002), "Clinical applications for change-point analysis of herpes zoster pain," *Journal of Pain and Symptom Management*, **23**, 510–516.

[15] Ertel, J.E. and Fowlkes, E.B. (1976), "Some algorithms for linear spline and piecewise multiple linear regression," *Journal of the American Statistical Association*, **71**, 640–648.

[16] Farley, J.U. and Hinich, M.J. (1970), "A test for a shifting slope coefficient in a linear model," *Journal of the American Statistical Association*, **65**, 1320–1329.

[17] Feder, P. I. (1975a), "On asymptotic distribution theory in segmented regression problems – Identified case," *Annals of Statistics*, **3**, 49–83.

[18] Feder, P.I. (1975b), "The log likelihood in segmented regression," *Annals of Statistics*, **3**, 84–97.

[19] Gallant, A.R. and Fuller, W.A. (1973), "Fitting segmented polynomial regression models whose join points have to be estimated," *Journal of the American Statistical Association*, **68**, 144–147.

[20] Ganocy, S.J. (2003), "Estimation problems from data with change points," *unpublished Ph.D. dissertation*, Case Western Reserve University, Dept. of Statistics.

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

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

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

[24] James, B., James, K.L. and Siegmund, D. (1987), "Tests for a change–point," *Biometrika*, **74**, 71–83.

[25] James, B., James, K.L. and Siegmund, D. (1992), "Asymptotic approximations for likelihood tests and confidence regions for a change–point in the mean of a multivariate normal distribution," *Statistica Sinica*, **2**, 69–90.

[26] Killick, R., Eckley, I.A., Ewans, K. and Jonathan, P. (2010), "Detection of changes in variance of oceanographic time–series using changepoint analysis," *Ocean Engineering*, **37**, 1120–1126.

[27] Killick, R. and Eckley, I.A. (2010), "changepoint: Analysis of changepoint models," http://CRAN.R-project.org/package=changepoint.

[28] Killick, R., Fearnhead, P. and Eckley, I.A.(2012), "Optimal detecton of changepoints with a linear computational cost," *Journal of the American Statistical Association*, **107**, 1590–1598.

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

[30] Kim, H.–J., Yu, B. and Feuer, E. J. (2009), "Selecting the number of change-points in segmented line regression," *Statistica Sinica*, **19**, 597–609.

[31] Koul, H.L. and Qian, L. (2002), "Asymptotics of maximum likelihood estimator in a two-phase linear regression model," *Journal of Statistical Planning and Inference*, **108**, 99–119.

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

[33] Loschi, R.H., Cruz, F.R.B., Iglesias, P.L. and Arellano–Valle, R.B. (2003), "A Gibbs sampling scheme to the product partition model: An application to change–point problems," *Computers and Operations Research*, **30**, 463–482.

[34] Loschi, R.H., Cruz, F.R.B., Takahashi, R.H.C., Iglesias, P.L., Arellano–Valle, R.B. and MacGregor Smith, J. (2008), "A note on Bayesian identification of change points in data sequences," *Computers and Operations Research*, **35**, 156–170.

[35] Montiani–Ferreira, F., Petersen–Jones, S., Cassotis, N., Ramsey, D.T., Gearhart, P. and Cardoso, F. (2003), "Early postnatal development of central corneal thickness in dogs," *Veterinary Ophthalmology*, **6**, 19–22.

[36] Pastor–Barriuso, R., Guallar, E, and Coresh, J. (2003), "Transition models for change-point estimation in logistic regression," *Statistics in Medicine*, **22**, 1141–1162.

[37] 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.

[38] Pauler, D.K. and Laird, N.M. (2002), "Non-linear hierarchical models for monitoring compliance," *Statistics in Medicine*, **21**, 219–229.

[39] Poirer, D.J. (1973), "Piecewise regression using cubic splines," Journal of the American *Statistical Association*, **68**, 515–524.

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

[41] Quandt, R.E. (1960), "Tests of hypothesis that a linear regression system obeys two separate regimes," *Journal of the American Statistical Association*, **55**, 324–330.

[42] Quandt, R.E. (1972), "A new approach to estimating switching regressions," *Journal of the American Statistical Association*, **67**, 306–310.

[43] Ray, B.K. and Tsay, R.S. (2000), "Bayesian methods for change–point detection in long–range depen- dent processes," *Journal of Time Series Analysis*, **23**, 687–705.

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

[45] Sakai, H. (1995), "Measurement and visualization of the contact pressure distribution of rubber disks and tires," *Tire Science and Technology*, **23**, 238–255.

[46] Shope, J.T. and Molnar, L.J. (2003), "Graduated driver licensing in the United States: Evaluation results from the early programs," *Journal of Safety Research*, **34**, 63–69.

[47] Siegmund, D.O. (1985), *Sequential Analysis: Tests and Confidence Intervals*. New York: Springer.

[48] Siegmund, D.O. and Zhang, H. (1994), "C*onfidence regions in broken line regression," in Change–point Problems, volume 23 of Lecture Notes–Monograph Series*, pp 292–316, Hayward, CA: Institute of Mathematical Statistics.

[49] Slate, E.H. and Turnbull, B.W. (2000), "Statistical models for longitudinal biomarkers of disease onset," *Statistics in Medicine*, **19**, 617–637.

[50] Spokoiny, Vladimir(2009), "Multiscale local change point detection with applications to value-at-risk," *Annals of Statistics*. **37**, 1405–1436.

[51] Tishler, A. and Zang, I. (1981), "A new maximum likelihood algorithm for piecewise regression," *Journal of the American Statistical Association*, **76**, 980–987.

[52] Worsley, K. J. (1986), "Confidence regions and tests for a change-point in a sequence of exponential family random variables," *Biometrika*, **73**, 91–104, 1986.

[53] 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.

[54] Zhang, N.R., Siegmund, D.O., Ji, H., and Li, J. (2009), "Detecting simultaneous change-points in multiple sequences," Accepted for publication in Biometrika.

[55] Zhao, X., Wu, X., and Zhou, X. (2009), "A change-point model for survival data with long-term survivors," *Statistica Sinica*, **19**, 377–390.

Stephen J. Ganocy
Case Western Reserve Universit

Jiayang Sun
Case Western Reserve Universit