Piecewise Regression Models: Estimation Theory and Applications

Lianfen Qian
Department of Mathematical Sciences, Florida Atlantic University
Boca Raton, FL 33431, USA
E-mail: Lqian@fau.edu

Abstract

When classical linear regression models are not adequate in modeling real data, nonlinear regression models are the natural option. Piecewise regression models are the most popular utilized nonlinear regression models in various applications. In this paper, we survey applications and theoretical research results for piecewise regression models, typically focusing on two phase linear regression models with an unknown change point. For the purpose of easy accessing and motivating reading, we first present a few concrete examples for different types of piecewise regression models. Then we continue to present hypothesis tests, estimations and limit distributions of the parameters in the piecewise regression models. Finally, we address some computational aspects and give final remarks for recent developments.

1 Introduction and Examples

In the last two decades, a considerable body of literature appears on the hypothesis tests, parameter estimation and relative computing program on change points for piecewise regression models. Various practical applications for such models have been appeared in diversified research areas such as biology (Gaylor et al. 1988; Vieth, 1989; Pieporsch and Baier, 1997), medical science (Smith and Cook, 1986; Cox, 1987; Gaylor, et. al. 1988; Berman at. el. 1996; Piepho and Ogunto, 2003; Muggeo, 2003), epidemiology (Ulm, 1991; Pastor and Guallar, 1998), environmental science (Anderson and Nelson, 1975; Gbur, et. al. 1979; Qian and Ryu, 2003; Toms and Lesperance, 2003; Pieporsch and Baier, 2005), software engineering (Qian and Yao, 2002), finance and econometrics...
(Chow, 1960; Koul and Qian, 2002; Fiteni, 2004; Zeileis, 2006) to name a few.

One special piecewise regression model is called a two-phase linear regression model. A two-phase linear regression model is defined by two linear models over two different domains of the design variable. To be precise, let \( Y \) be the response variable and \( X \) be an univariate predictor so that \((X,Y)\) be a bivariate random vector with \( E|Y| < \infty \). The conditional expectation is denoted by \( m(x) = E(Y|X = x) \).

Let \( \mathbb{R} \) be the real line and \( \overline{\mathbb{R}} = \mathbb{R} \cup \{-\infty, \infty\} \). Then the set \( \overline{\mathbb{R}} \) is compact under the metric \( d(x,y) = |\arctan x - \arctan y|, \quad x, y \in \mathbb{R} \).

Throughout we assume that \( \theta \) is an interior point of the parameter space \( \Theta = \mathbb{K} \times \overline{\mathbb{R}} \) for a known compact set \( \mathbb{K} \) in \( \mathbb{R}^4 \). A typical point in \( \Theta \) will be denoted by \( \theta' = (\beta_{10}, \beta_{11}, \beta_{20}, \beta_{21}, s) = (\theta'_1, s) \). We define

\[
m(x, \theta) = (\beta_{10} + \beta_{11}x)I(x \leq s) + (\beta_{20} + \beta_{21}x)I(x > s), \quad x \in \mathbb{R},
\]

and consider a set of independent observations \((X_i, Y_i), \ i = 1, \cdots, n\), such that for some \( \theta' = (\theta'_1, \tau) = (b_{10}, b_{11}, b_{20}, b_{21}, \tau) \in \mathbb{K}^5 \),

\[
Y_i - m(X_i, \theta) = \varepsilon_i, \quad i = 1, \cdots, n, \tag{1.1}
\]

are independent identically distributed (i.i.d.) random variables and \( \tau \) is called the change point.

Bhattacharya (1990, 1994) gave a model classification. That is, there are two types of two-phase regression models: continuous (also named restricted and segmented) and discontinuous (unrestricted). By continuous, the two phases meet at the change point \( \tau \) (also called joint point). By discontinuous, the two phases are separated at the change point \( \tau \) with some jump. Within discontinuous models, there are further classification of fixed jump and contiguous jump. By fixed jump, the jump size at \( \tau \) is a fixed number, neither equal to zero nor dependent on the sample size indicating an abrupt change, while by contiguous jump, the jumps size goes to zero when the sample size goes to infinity indicating a gradually change.

Among the first studies, the hypothesis tests and parameter estimation problems under regression setting for change point problem were introduced by Quandt (1958, 1960) using maximum likelihood method under normal error. When the data pairs are observed over a discrete time points, the model changes at a change point of time. Before this change point of time, the relationship follows one linear model, while after the change point of time, the relationship follows another linear model.

Notice that when the design points \( \{X_i\} \) are ordered already in model (1.1), then Quandt's model is equivalent to two-phase separate regression model, which switches at an unknown time \( k \) such that \( X_k = \tau \). Hence both \( k \) and \( \tau \) are called the change point.

Under a strong assumption of known change point \( \tau \), Sprent (1961) discussed hypothesis tests of two-phase continuous (segmented) regression models for several data sets with application in biometry. Under continuity constraint, Hudson (1966) obtained the estimated parameters using method of least squares. Hasselblad el. al (1976) applied Hudson's method to the specific case of a hockey-stick function, which assumes that the left regime of the data set is fitted by a linear regression with slope zero. The evolution of the names of hockey-stick models is very colorful. For the diversified real applications of the models, it can be very differently named. It is called linear-plateau model by Anderson and Nelson (1975), and threshold response model by Gaylor et al. (1988) and Ulm (1991), in the analysis of environmental and biological data in predicting crop yield based on the amount of nitrogen in the soil. For a left linear plateau model, we write

\[
m(x, \theta) = b_0I(x \leq \tau) + [b_0 + b_1(x - \tau)]I(x > \tau), \quad x \in \mathbb{R}, \tag{1.2}
\]

where \( \theta' = (b_0, b_1, \tau) \).

Example 1.1: (Determination of Osmotic Threshold) Vieth (1989) fitted the continuous two phase linear model (1.1) to plasma osmolality against arginine vasopressin (AVP) concentration in plasma of conscious dogs to determine the osmotic threshold. The detected threshold is 33.37 plasma osmolality (mOsm/kg) using least squares (LS) fit shown in Figure 1.1.

As one can see that the plasma osmolality data in Example 1.1 can also be fitted by the left plateau model (1.2). Piez cos and Bailer (2005, p. 45) fit the data reported by Braña and Ji (2000) on embryogenic response in wall lizards (Podarcis muralis) by manipulating selected environmental conditions associated with their embryonic development using the left linear plateau model (1.2). For the data set and detail modeling, please refer to Piez cos and Bailer (2005).

In reality, data often show patterns with more than one change point. One of the extended model of (1.2) is a multiple plateau model. For example, a continuous model with both left and right plateaus can be expressed as

\[
m(x, \theta) = \begin{cases} 
b_0, & \text{if } x \leq \tau_1, \\
b_0 + b_1(x - \tau_1), & \text{if } \tau_1 < x \leq \tau_2, \\
b_0 + b_1(\tau_2 - \tau_1), & \text{if } x > \tau_2, 
\end{cases}
\tag{1.3}
\]

where \( \tau_1 \) and \( \tau_2 \) are the change points and \( \theta' = (b_0, b_1, \tau_1, \tau_2) \).
Example 1.2: (Pond Water Levels) Small and isolated bodies of water within larger ecosystems often exhibit substantial variation in their water levels, due ostensibly to factors associated with their supply sources (data from Piegorsch and Bailer, 2005). Blood et al. (1997) studied this phenomenon through data collected from 59 small ponds in Georgia State in USA, where the water levels of those ponds are affected by the nearby deep Floridian ponds. The variables measured are $Y=$pond water level vs. $x=$aquifer level. Visualizing the data indicates that the model (1.3) with left and right plateau is adequate (Figure 1.2). The detected change points using LS fit are 37.893 and 41.079. The corresponding PROC NLIN SAS code is included in Figure 1.3 for convenience of reproducing the fits.

Abrupt (fixed jump size) change or graduate (contiguous) change may happen in real life because of policy changes, or management skill changes or the nature of the problem changes. One of the example is given below for modeling software effort using abrupt change estimation procedure.

Example 1.3: (Software Effort Cost) As Shepperd, M. and Schofield, C. (1997) pointed out, an important aspect of any software development project is to know how much it costs. In most cases the major cost factor is labor. For this reason, estimation of software project effort is one of the most important empirical modeling tasks in software engineering, as indicated by the large number of models developed over the past two decades. Shin and Goel (SG,2000) studied the modeling of the software project effort using radial basis functions. They used the well-known NASA software project data set to validate the model. The method proposed in SG is complicated and contains many parameters to be estimated.
Figure 1.3: Sample SAS program to fit continuous three segment model to pond water level data.

The scatterplot (Figure 1.4) of the software effort cost versus the line of code (in thousands) of the projects appears to be an abrupt change in the effort. Qian and Yao (2002) re-fitted the well-known NASA software project data set. The discontinuous model (1.1) with fixed jump size is used to detected change point at 31.1 lines of codes (in thousands) shown in Figure 1.4. It is wise to do not make the continuity assumption since there is an apparent discontinuity after the value of 31.1 lines of codes (in thousands). If a continuity assumption were to imposed, either it would need to impose an unwarranted constraint on the regression function leading to a bad fit, or we would have to use more complicated functional forms that would otherwise be necessary. Furthermore, Qian and Yao showed that the two-phase linear regression using least absolute deviation estimation method is the most robust model comparing with linear model, COCOMO model (Boehm 1981) and SG method.

For much earlier research applications, more examples can be found from Seber and Wild (1989, Chapter 9) and Piegorsch and Bailer (1997, section 7.1.2) which gives a good summary on threshold models and other nonlinear forms with applications in environmental sciences.

In the fixed design and restricted case, Feder (1975) gave a rigorous treatment covering the two phase regression model under i.i.d errors. Feder (1975) provided a proof that the likelihood ratio statistic is asymptotically χ² based on the assumption that the parameter vector remains identified under null hypothesis for i.i.d errors. It is obvious that Feder’s (1975) results can not be used for testing of no change point.

Two comprehensive survey were given by Shaban(1980) and Schulze (1987). Shaban (1980) simply quoted a list of the author’s summaries (abstracts) except in a few cases and classified the papers published before 1980 depending on the approach adopted by the author(s) for change-point problem and the related problem in segmented regression models. Schulze (1987) provided a collection of existing methods mainly focusing on the least squares estimation, testing of hypotheses and test-
ing of model stability for analyzing data using multiphase segmented regression models majoring on two-phase models. Park (1978) discussed some design issues when the change points are known. Omikogu (1984) discussed question: when is a piecewise regression really necessary? In another word, it is important to test the hypothesis of having change versus no change since under no change hypothesis the change point is not identifiable. Hence the standard likelihood theory is not applicable.

Kim and Siegmund (KS, 1989) and Kim (K, 1993) studied the likelihood ratio test for a change point in simple linear regression with homogeneous and non-homogeneous errors, respectively. They consider two models: change only happens for intercept (step-type), changes on both intercept and slope.


Within regression setting, the two types of models, in terms of continuous and discontinuous, require distinctly different statistical methodologies in proving limiting property of the change point. There are three important problems to address:

(a) testing simple linear regression versus two-phase linear regression with one single unknown change point.

(b) making inference about the change point.

(c) making inference about the regression parameters and the error variances for the two phases.

The next portion of the paper is organized as follows. Section 2 addresses hypothesis testing issues based on likelihood ratio and informational based approaches. Section 3 focuses on reviewing the up-to-date research activities on asymptotic limiting theory for maximum likelihood estimators including change point estimator for continuous case. Section 4 reviews estimation methods for change points and regression parameters for discontinuous case. Section 5 addresses some computational aspects. Section 6 concludes the paper with remarks on recent developments and extensions to model other types of data such as time series and longitudinal data.

2 Hypothesis Testing

For continuous two phase regression model, Davies (1977, 1987) derived test statistics for both normal and chi-squared cases when a nuisance parameter is present only under the alternative hypothesis. In this section, we focus on the hypothesis tests for discontinuous models utilizing two approaches: Likelihood ratio based approach and informational based approach.

2.1 Likelihood Ratio Test

A likelihood ratio approach was first considered by Quandt (1958) to detect a change in a simple linear regression. Kim and Siegmund (1989) discussed analytic difficulties associated with the likelihood ratio test. Kim and Cai (1993) studied the robustness of the likelihood ratio test for a change in simple linear regression and concluded that the test statistics converge to same distribution regardless of the underlying distribution.

The hypothesis test for a single linear regression versus two-phase linear regression (1.1) with a change at a fixed observation index, say, $k$ between 2 and $n-2$ can be formulated to test null hypothesis:

$$H_0 : E(Y_i|X_i = x_i) = b_0 + b_1 x_i, \text{ for all } i = 1, \ldots, n$$

versus the alternative hypothesis

$$H_1 : E(Y_i|X_i = x_i) = \begin{cases} b_{10} + b_{11} x_i, & i = 1, \ldots, k \\ b_{20} + b_{21} x_i, & i = k + 1, \ldots, n. \end{cases}$$

In the fixed design case, where we use lower case letter for the design variable, $x_1 < \cdots < x_k \leq \tau < x_{k+1} < \cdots < x_n$ are non-stochastic and in (1.1), the coefficient vector changes after $k$th observations or $x_\tau = \tau$, so that either $\tau$ or $k$ can be regarded as the change point. Bhattacharya (1994) formulate the problem of testing simple linear regression versus two-phase linear regression with a single unknown change point into to test

$$H_0 : k = n$$

against the alternative

$$H_1 : 1 \leq k \leq n - 1.$$

In this section, we address the hypothesis tests under general setting of the design variable $X$ for the discontinuous models (no constraint on all parameters). That is, the design variable $X$ in general is a random variable. For simplicity, we assume that the error term in (1.1) is normal
distributed. Then for any parameter \( \theta \in \Theta \), the conditional likelihood function of the model (1.1) is

\[
L(\theta, \sigma^2) = (2\pi \sigma^2)^{-\frac{n}{2}} \exp \left( -\frac{1}{2\sigma^2} \sum_{i=1}^{n} (Y_i - m(X_i; \theta))^2 \right).
\]  

(2.1)

Let \( X(i) \) be the order statistics of \( \{X_i\} \) and \( Y(i) \) be the concomitant of \( X(i) \). For a fixed \( s \in \mathbb{R} \), denote \( i_s = \max\{i : X_i \leq s\} \), \( Y^{(1)} = (Y(1), \ldots, Y(i_s)) \) and \( Y^{(2)} = (Y(i_s+1), \ldots, Y(n)) \). The corresponding design matrices are denoted by \( X_s^{(1)} \) and \( X_s^{(2)} \). Then a two-step algorithm for likelihood based estimation is given below:

**Step 1:** For a given \( s \in \mathbb{R} \), define

\[
\hat{\theta}_{1j} = \left( X_s^{(j)} X_s^{(j)} \right)^{-1} X_s^{(j)} Y_s^{(j)}, \quad j = 1, 2
\]

and

\[
\hat{\sigma}^2 = \frac{S^{(1)}_s + S^{(2)}_s}{n},
\]

(2.3)

where \( S^{(j)}_s \) is the residual sum of squares for the \( j \)th phase of the model, \( j = 1, 2 \). Notice that for any \( 1 \leq k < n \), and \( s \in [X(k), X(k+1)) \), \( S^{(j)}_s \) is a constant function in \( s \) for \( j = 1, 2 \). We also denote \( S^{(j)}_s \) by \( S_k^{(j)} \) for \( s \in [X(k), X(k+1)) \). Hence the conditional likelihood function may not have unique maximizers (for discontinuous model), so we suggest to take the maximum likelihood estimator of \( \tau \) to be the smallest minimizer of \( \hat{\sigma}^2 \) in (2.3), denoted as \( \hat{\tau} \).

**Step 2:** Substitute the change point estimator \( \hat{\tau} \) into the equation (2.2), we obtain the estimated regression coefficient vector

\[
\hat{\theta}' = (\hat{\theta}_{11}', \hat{\theta}_{12}', \hat{\theta}_{22}').
\]

Then the maximum likelihood estimator (MLE) of regression coefficients

\[
\hat{\theta}' = (\hat{\theta}_{11}', \hat{\theta}_{12}', \hat{\theta}_{22}') = (b_{10}, b_{11}, b_{20}, b_{21})
\]

is denoted by

\[
\hat{\theta}' = (\hat{\theta}_{11}', \hat{\theta}_{12}', \hat{\theta}_{22}').
\]

Hence the maximum likelihood estimator of \( \theta' = (\theta_1', \tau) \) is denoted by

\[
\hat{\theta}' = (\hat{\theta}_{11}', \hat{\theta}_{12'}, \hat{\tau}').
\]

For a given \( k \) between 2 and \( n-2 \) or equivalently \( s = X(k) \), \( S_k^{(j)} = S_s^{(j)} \) for all \( s \in [X(k), X(k+1)) \) for \( j = 1, 2 \). Let \( S \) be the residual sum of squares with no change (simple linear regression model). One constructs an extra sum F-test

\[
F_k = \frac{[S - (S_k^{(1)} + S_k^{(2)})]/2}{(S_k^{(1)} + S_k^{(2)})/(n-4)},
\]

which has an \( F_{2, n-4} \) distribution when \( H_0 \) is true. Hence for unknown change point, one intuitive test statistic is

\[
F_{\text{max}} = \max_{2 \leq k \leq n-2} F_k.
\]

In the fixed design and discontinuous case, Beckman and Cook (1979) reported some simulation results on the asymptotic null distribution of \( F_{\text{max}} \). Beckman and Cook (1979) and Worsley (1983) derived simple bound of the null exceedance probability of \( F_{\text{max}} \) using Bonferroni's inequality. Instead of using 2 degrees of freedom in the numerator of \( F_k \), Hinkley (1971) used 3 and reported strong empirical evidence that the distribution of \( F_k \) is \( F \) distribution with degrees of freedom \( (3, n-4) \), hence the distribution of the likelihood ratio statistic under \( H_0 \) is asymptotically \( \chi^2 \) under i.i.d. errors. This result has been widely adopted by applied disciplinary. Recently, Lund and Reeves (2002) experimented the degrees of freedom 3 in Hinkley's report and found that using degrees of freedom 2 would be more suitable for undocumented time under the two phase model with both step-type (intercept) and trend-type (slope) changes. They conjectured that a reasonable approximating limiting distribution of \( F_{\text{max}} \) (as \( n \to \infty \)) would involve the Gumbel extreme value law:

\[
P[F_{\text{max}} \leq x] = \exp\{-[\alpha - (a_n)/b_n]x\},
\]

for appropriately selected scaling constants \( a_n \) and \( b_n \). This is still an open problem for readers. Wang (2003) in commenting Lund and Reeves's article pointed out that for the presence of only step-type change, it would be even more suitable to use \( (1, n-3) \) degree of freedom for \( F_k \). Both Lund and Reeves (2002), and Wang (2003) gave percentile tables for \( F_{\text{max}} \). The null hypothesis of a single regression can be rejected at the \( 100\alpha\% \) level if \( F_{\text{max}} \) is in the upper \( 100\alpha\% \) of simulated values.

### 2.2 Informational Approach

The other criteria for hypothesis testing is to use informational approach. The most classical criterion is Akaike Information Criterion (AIC) initially proposed by Akaike (1973). That is to minimize

\[
\text{AIC}(p) = -2 \log L(\hat{\theta}) + 2p
\]

with respect to \( p \), where \( L(\hat{\theta}) \) is the maximum likelihood function for the model with \( p \) parameters.

Schwarz (1978) proposed a modification of AIC, called Schwarz Information Criterion (SIC), taking the sample size into consideration as follow:

\[
\text{SIC}(n, p) = -2 \log L(\hat{\theta}) + p \log n.
\]
For normal error, under $H_0$ and using SIC, Chen and Gupta (2001) obtain

$$\text{SIC}(n) = n \log L_0(\hat{b}_0, \hat{b}_1, \hat{\sigma}^2) + 3 \log n$$

$$= n \log \left[ \sum_{i=1}^{n} (y_i - \hat{b}_0 - \hat{b}_1 x_i)^2 \right] + n(1 + \log 2\pi) + (3 - n) \log n,$$

where $(\hat{b}_0, \hat{b}_1, \hat{\sigma}^2)$ is the maximum likelihood estimator of the regression parameter $(b_0, b_1, \sigma^2)$ under the null hypothesis $H_0$.

Under $H_1$ with a fixed change point $k$, for $k = 2, \ldots, n - 2$, one derives

$$\text{SIC}(k) = -\log L_1(\hat{b}_{10}, \hat{b}_{11}, \hat{b}_{20}, \hat{b}_{21}, \hat{\sigma}^2) + 5 \log n$$

$$= n \log \left[ \sum_{i=1}^{k} (y_i - \hat{b}_{10} - \hat{b}_{11} x_i)^2 + \sum_{i=k+1}^{n} (y_i - \hat{b}_{20} - \hat{b}_{21} x_i)^2 \right]$$

$$+ n(1 + \log 2\pi) + (5 - n) \log n,$$

where $L_1(\hat{b}_{10}, \hat{b}_{11}, \hat{b}_{20}, \hat{b}_{21}, \hat{\sigma}^2)$ is the maximum likelihood function under $H_1$. Therefore the decision rule is to

- select the model with no change if SIC($n$) $\leq$ SIC($k$) for all $2 \leq k \leq n - 2$;
- select the model with a change at $k$ if

$$\text{SIC}(k) = \min \{\text{SIC}(k) : 2 \leq k \leq n - 2\} < \text{SIC}(n).$$

In the next two sections, we present some asymptotic limiting theory for continuous and discontinuous models.

## 3 Asymptotic Theorems: Continuous Case

In the model (1.1), we impose the continuity condition:

$$b_{10} + b_{11}\tau = b_{20} + b_{21}\tau. \quad (3.1)$$

Denote $\Delta_0 = b_{20} - b_{10}$ and $\Delta_1 = b_{21} - b_{11}$. Then the change point $\tau = -\Delta_0/\Delta_1$ if $\Delta_1 \neq 0$. Otherwise the change point is non-identifiable. Thus within continuous model, we assume $\Delta_1 \neq 0$. Seber and Wild (1989, section 9.3) summarized the results before 1989. In this paper, we review and summarize results developed after 1989. Bhattacharya (1990, 1994) derived that the limiting distribution of the MLE for fixed design case.

More specifically, in the fixed design case, let $G_n(x) = n^{-1} \sum_{i=1}^{n} f(x_i \leq x)$ and suppose that there is a function $G$ with $G'(\tau) = p \in (0,1)$ and $g(\tau) = \frac{d}{d\tau}G(\tau) > 0$ such that as $n \to \infty$, $G_n(\tau) \to G(\tau)$ and $G_n(x) - G_n(\tau) = [G(x) - G(\tau)] : [1 + o(1)]$ uniformly in a neighborhood of $\tau$. Suppose further that as $n \to \infty$,

$$\text{np}^{-1} \sum_{i=1}^{np} \left[ \frac{x_i}{x_i^2} \right] \to \left[ \frac{1}{\mu_1} \frac{\mu_1}{\mu_1^2 + \sigma_1^2} \right] = A'_1 A_1, \quad (3.2)$$

and

$$\text{np}^{-1} \sum_{i=\text{np}+1}^{n} \left[ \frac{x_i}{x_i^2} \right] \to \left[ \frac{1}{\mu_2} \frac{\mu_2}{\mu_2^2 + \sigma_2^2} \right] = A'_2 A_2, \quad (3.3)$$

where $\sigma_1^2$ and $\sigma_2^2$ are positive and that $\max_{x \leq \tau} G(x) = o(n^{1/2})$. Furthermore, we assume that $x_{np} \leq \tau < x_{np+1}$ and $x_{np+1} - \tau \to 0$ as $n \to \infty$. So we can treat $x_{np} = \tau$.

Let $\delta_1 = (u_0, u_1, v_0, v_1) = (u', v')$. Consider any point in the parameter space

$$(\theta_1, k) \equiv (\theta_1 + n^{-1/2} \delta_1, \text{np} + j_1(\delta_1))$$

such that the lines $(b_{10} + n^{-1/2} u_0) + (b_{11} + n^{-1/2} u_1) x$ and $(b_{20} + n^{-1/2} v_0) + (b_{21} + n^{-1/2} v_1) x$ intersect at $x_{np+j_1(\delta_1)}$. Furthermore, the change point parameter $s(\delta_1) \equiv s_n(\delta_1)$ and the integer $j(\delta_1) \equiv j_n(\delta_1)$ satisfy

$$x_{np+j(\delta_1)} \leq s(\delta_1) < x_{np+j(\delta_1)+1} \quad \text{and} \quad x_{np+j(\delta_1)} = s(\delta_1).$$

Under the normal error assumption and $\delta'_1 = (\theta_{10}, \theta_{11}, \theta_{20}, \theta_{21})$, the log-likelihood ratio process is defined as:

$$R_n(\delta_1) = l_n(\theta_1, k) - l_n(\theta_1, np), \quad (3.4)$$

where

$$l_n(\theta_1, k) = \sigma^{-2} \left[ \sum_{i=1}^{k} (Y_i - \theta_{10} - \theta_{11} x_i)^2 + \sum_{i=k+1}^{n} (Y_i - \theta_{20} - \theta_{21} x_i)^2 \right]$$

**Theorem 3.1.** For the fixed design continuous model of (1.1) with homogeneous Gaussian error terms, if conditions (3.2) and (3.3) hold, then

$$R_n(\delta_1) \Rightarrow (\|W_1\|^2 + \|W_2\|^2)$$

$$+ \| p^{1/2} \sigma^{-1} A_1 u - W_1 \|^2 + \| (1 - p)^{1/2} \sigma^{-1} A_2 v - W_2 \|^2,$$

where $W_1$ and $W_2$ are independent bivariate normal vectors with mean vector $0$ and covariance matrix $I$. 


See Bhattacharya (1990) for details of proof. These weak convergence results indicate that \( n^{1/2}(\hat{\theta} - \theta) \) and \( n^{1/2}(\hat{\theta}_1 - \theta_1) \), \( n^{1/2}(\hat{\theta}_2 - \theta_2) \) are asymptotically independently distributed as \( \sigma A^{-1} W_1 \) and \( \sqrt{n} A^{-1} W_2 \), respectively. Consequently,

\[
 n^{1/2}(\hat{\tau} - \tau) = n^{1/2} \left( \frac{b_{10} - \bar{b}_{10} - b_{10} - b_{10}}{b_{21} - b_{11}} \right) \sim N(0, \alpha^2),
\]

where the asymptotic variance \( \alpha^2 \) obtained by the delta method is equal to

\[
\alpha^2 = (b_{21} - b_{11})^{-2} \left\{ \left( \text{Var}[b_{10}] + \text{Var}[^{\bar{b}_{10}}] \right) \\
+ 2\tau \left( \text{Cov}[b_{10}, \bar{b}_{11}] + \text{Cov}[\bar{b}_{20}, \bar{b}_{21}] \right) \\
+ \tau^2 \left( \text{Var}[\bar{b}_{11}] + \text{Var}[\bar{b}_{21}] \right) \right\},
\]

Denote all \( x_i \)'s less than or equal to \( \tau \) as \( x_{11}, \ldots, x_{i_{m1}} \) and those bigger than \( \tau \) as \( x_{21}, \ldots, x_{i_{n2}} \) with the sample means \( \bar{x}_j = \frac{1}{n_j} \sum_{i=1}^{n_j} x_{ji} \).

**Corollary 3.2.** For the fixed design continuous model of (1.1) with homogeneous Gaussian error terms, then for \( j = 1, 2 \),

\[
\hat{b}_{j0} \sim N \left( b_{j0}, \frac{\sigma^2 \sum x_{j1}^2}{n_j \sum (x_{j1} - \bar{x}_j)^2} \right),
\]

\[
\hat{b}_{j1} \sim N \left( b_{j1}, \frac{\sigma^2}{\sum (x_{j1} - \bar{x}_j)^2} \right),
\]

and

\[
\text{Cov}[\hat{b}_{j0}, \hat{b}_{j1}] = -\frac{\sigma^2 \bar{x}_j}{\sum (x_{j1} - \bar{x}_j)^2}.
\]

**Proof:** See Seber and Wild (1989, p. 450) for details.

In the random design case, let \( G \) be the distribution function of \( X \) having a continuous Lebesgue density \( g \) at \( \tau \). For \( s \in \mathbb{R} \), denote \( \tilde{G}(s) = 1 - G(s) \), and \( \mu_j(s) = EX^j I(X \leq s) \), \( j = 1, 2 \). Then,

\[
E \left\{ (np)^{-1} \sum_{i=1}^{np} \left[ \frac{1}{X_i} X_i^2 \right] I(X_i \leq \tau) \right\} \rightarrow \begin{bmatrix} G(\tau) \\ \mu_1(\tau) \end{bmatrix}, \\
\begin{bmatrix} \mu_1(\tau) & \mu_2(\tau) \end{bmatrix} = \Sigma_{1\tau},
\]

and

\[
E \left\{ (n - np)^{-1} \sum_{i=\text{uppen}}^{n} E \left[ \frac{1}{X_i} X_i^2 \right] I(X_i > \tau) \right\} \rightarrow \begin{bmatrix} \tilde{G}(\tau) \\ \tilde{\mu}_1(\tau) \end{bmatrix}, \\
\begin{bmatrix} \tilde{\mu}_1(\tau) & \tilde{\mu}_2(\tau) \end{bmatrix} = \Sigma_{2\tau},
\]

\[
\text{Denote} \Sigma_{\tau} = \text{diag}(\Sigma_{1\tau}, \Sigma_{2\tau}) = \begin{bmatrix} G(\tau) & \mu_1(\tau) \\ \mu_1(\tau) & \mu_2(\tau) \\ 0 & 0 \\ 0 & 0 \end{bmatrix}.
\]

**Theorem 3.3.** In the continuous random design case, assume the finite second moment of \( X \) and homogeneous normal error with mean zero and finite variance \( \sigma^2 \). Then

\[
n^{1/2}(\hat{\theta}_1 - \theta_1) \sim \sigma \Sigma^{-1/2} \begin{bmatrix} W_1 \\ W_2 \end{bmatrix},
\]

where \( W_1 \) and \( W_2 \) are the same as in Theorem 3.1. Hence the change point estimator is the same as in the fixed design case.

Koul (2000) introduced a marked empirical process in fitting two phase linear regression models and obtained asymptotically distributed free tests in testing \( M_0 : m(x) = m(x, \theta) \). For \( x \in \mathbb{R}, v \in \Theta \), a marked empirical process is defined as

\[
V_n(x, v) = n^{-1} \sum_{i=1}^{n} [Y_i - m(X_i, v)] I(X_i \leq x),
\]

where the residuals \( Y_i - m(X_i, v) \) are the marks. For \( x \in \mathbb{R} \), denote \( a(x)' = (-1, -x, 1, x) \) with \( a_1' = (-1, -\tau, 0, 0) \) and \( a_2' = (0, 0, 1, \tau) \). Let \( F \) be the distribution function of error and \( \nu_p(1) \) means uniformly convergent to zero in probability.

**Theorem 3.4.** In the random design case of model (1.1), suppose \( EX^2 < \infty, g \) is continuous at \( \tau \) and \( F \) has uniformly continuous Lebesgue density \( f > 0 \), a.e. Then

\[
n^{1/2} \begin{bmatrix} V_n(x, \theta) - V_n(x, \theta) \\ V_n(x, \theta) \end{bmatrix} = \begin{bmatrix} \theta'_1 a(\tau) g(\tau) \\ \theta_1 a(\tau) g(\tau) \end{bmatrix} \left\{ \sqrt{n} \begin{bmatrix} I(\tau < x \leq \hat{\tau}) + \sqrt{n} (\hat{\tau} - \tau) I(x > \hat{\tau}) \end{bmatrix} I(\hat{\tau} > \tau) \\
- \sqrt{n} (\hat{\tau} - \tau) I(\tau < \hat{\tau} \leq x) - \sqrt{n} (\hat{\tau} - \tau) I(x > \hat{\tau}) \end{bmatrix} I(\hat{\tau} \leq \tau) \right\}
\]

\[
- \sqrt{n} (\hat{\theta}_1 - \theta_1) T_{\tau}(x) + u_p(1).
\]

Consequently, in the continuous random design case, i.e. \( \theta'_1 a(\tau) = 0 \), we have

\[
n^{1/2} \begin{bmatrix} V_n(x, \hat{\theta}) - V_n(x, \theta) \\ V_n(x, \hat{\theta}) \end{bmatrix} = -n^{1/2}(\hat{\theta}_1 - \theta_1) [T_{\tau}(x) + u_p(1)],
\]
where
\[
\Gamma_\tau(x) = E\left[\begin{bmatrix} I(X \leq x \land \tau) \\ XI(X \leq x \land \tau) \\ I(X > \tau, X \leq x) \\ XI(X > \tau, X \leq x) \end{bmatrix}\right] \quad \text{and} \quad \Gamma_\tau = \Gamma_\tau(\infty).
\]

Combining Theorems 3.3 and 3.4, we obtain
\[
\sqrt{n} V_n(x, \hat{\theta}) = \left\{ \frac{1}{\sqrt{n}} \sum c_t I(X_t \leq x) - \sigma \left[ \frac{W_1}{\sum \Gamma_\tau(x)} \right]^{-1/2} \right\} + u_n(1).
\]

Let \( Z_n(x) \) be the leading process in the right hand side of the above approximation. Then a simple calculation shows that
\[
EZ_n(x)Z_n(y) = \sigma^2 \left[ G(x \land y) + \Gamma_\tau(x)\Sigma^{-1}\Gamma_\tau(y) \right], \quad x, y \in \mathbb{R}.
\]

Similarly to Koul (2000), one can construct a transformation \( T_n \) of the marked empirical process \( V_n(\cdot, \hat{\theta}) \) such that under null hypothesis \( M_0 : m(x) = m(x, \theta), \forall x \in \mathbb{R} \) and for some \( \theta \in \mathbb{R}^2 \), the transformed process converges weakly to \( \sigma^{-1} B(G) \) before the change point, where \( B \) is a Brownian motion on [0, \infty]. Hence, any test of \( M_0 \) based on a continuous function of the transformed process will be asymptotically distribution free test \( M_0 \).

4 Asymptotic Theorems: Discontinuous Case

Let \( \nu_n \) be a sequence of positive numbers and \( c_j \) be real numbers for \( j = 1, 2 \). Let
\[
d_n \equiv (b_{20} - b_{10}) + (b_{21} - b_{11}) \tau \equiv (c_0 + \tau c_1)/\nu_n \neq 0 \quad (4.1)
\]
be the jump size at the change point \( \tau \) of the model (1.1). The model (1.1) is discontinuous since \( d_n \neq 0 \) for all \( n \geq 1 \). Within the discontinuous model, if \( \nu_n \equiv 1 \) for all \( n \geq 1 \), then the model has a fixed jump size \( d \equiv d_n \) for all \( n \geq 1 \) indicating abrupt change. If \( \nu_n = o(n^{1/2}) \) and \( \nu_n \to \infty \), then the model is indicating a gradually structural change.

Among discontinuous fixed jump models, Quandt (1958, 1960) studied the mean change problem over observation time. Chow (1960) presented two tests based on prediction intervals and F-test for the hypothesis of no change in the simple linear regression versus abrupt change with application in econometrics. Bhattacharya and Johnson (1968) investigated the non-parametric tests for step-type change (location in one sample problem) at an unknown time point. Hinkley (1970) derived the asymptotic distribution of the maximum likelihood estimator (MLE) by random-walk considerations when these regression phases are parallel (intercept change, step-type change).

Hawkins (1980) reported that the inferential theory depends strongly on whether or not continuity at the change-point is assumed. Furthermore, the likelihood ratio test for the presence of two phase regression tends to infinity. A special case is one in which a change only happens in intercept (step-type). Worsley (1983) specialized his improved Bonferroni approximation to the likelihood ratio test to this case. Null quantiles for the likelihood ratio test were calculated from simulations by Maronna and Yohai (1978). Csörgő and Horváth (1988), van de Geer (1988) and Bhattacharya (1990) reported that the asymptotic result for the contiguous case is similar to the continuous case. Bai and Perron (1998) investigated the asymptotics of the least squares estimators and the corresponding tests in multi-phase contiguous random design linear regression models when the errors satisfy some martingale type assumptions.

Koul and Qian (KQ, 2002) established the consistency and the limiting distribution of the MLE in the fixed jump size two-phase random design regression models for a class of error densities that excludes the double exponential and such non-smooth densities. Koul, Qian and Surgailis (2003) extended KQ’s results to M-estimation and gave out much details for proving the limiting distribution of the change point. Fiteni (2004) obtained the asymptotic distribution of the break location estimator introduced by Yohai and Zamar (1988) for a shift of fixed magnitude as well as for a shift with its magnitude converging to zero (gradual change or contiguous model) when the sample size increases under possibly contaminated distribution for both the regressors and the error term.

For data collected sequentially over time, sequential control-chart approaches to detect general changes in multiple regression parameters are discussed by Brown et al. (1975). CUSUM methods for detecting changes in intercept only are discussed by Schweder (1976).

Seber and Wild (1989, p.438) claimed that “we cannot hope to estimate the changepoint.” Bhattacharya (1994) reported that in the fixed design case, the estimators of the regression parameters behave in the same way in discontinuous case as in continuous case, but \( \hat{\beta} \) is asymptotically independent of the estimated regression parameters for contiguous case only. Koul and Qian (2002) filled up the gap for fixed jump case. The following theorem states the consistent results for both discontinuous cases.

**Theorem 4.1.** Assume the error term in model (1.1) is Gaussian with mean zero and finite homogeneous variance and \( EX^2 < \infty \). If the model...
is discontinuous with (i) $\nu_n^2/n = o(1)$ and $\nu_n \to \infty$ for contiguous case 
(ii) $\nu_n = 1$ for fixed jump case, then the maximum likelihood estimator of $\theta$ satisfies

$$||\hat{\theta}_1 - \theta_1|| = O_p(1), \quad ||n\nu_n^{-2}(\hat{\tau} - \tau)|| = O_p(1).$$

Proof. See Bhattacharya (1990) for contiguous case, Koul and Qian (2002) for fixed jump case.

Note that the assumption of $\nu_n \to \infty$ obviously excludes the fixed jump size discontinuous case. In fact, within the discontinuous model, the two types of discontinuity behave distinctly different in terms of limiting distribution of the change point estimator.

First we include results for contiguous case. Let the jump size $\nu_n \to \infty$ and $\nu_n = o(n^{1/2})$. Denote $k = np + \nu_n^2 t$, $u' = (u_0, u_1)$ and $v' = (v_0, v_1)$. Then the likelihood ratio process is

$$R_n(u, v, t) = l_n(\theta_1, k) - l_n(\theta_1, np)$$

with five free parameters. Bhattacharya (1990, 1994) concluded that

**Theorem 4.2.** (Contiguous case) In the contiguous case of model (1.1), let all assumptions on the design variable in the continuous case hold. Then,

$$R_n(u, v, t) \leq -\left(\|W_1\|^2 + \|W_2\|^2\right)$$

$$+ \left(\|p^{1/2}\sigma^{-1}A_1u - W_1\|^2 + \|(1 - p)^{1/2}\sigma^{-1}A_2v - W_2\|^2ight)$$

$$+ \sigma^{-1}|c_0 + c_1\tau| \cdot \left[ B(t) + \frac{1}{2}\sigma^{-1}|c_0 + c_1\tau| |t| \right],$$

where $W_1$, $W_2$, $A_1$, $A_2$ are as in Theorem 3.1 and $B(\cdot)$ is a standard two-sided Brownian motion which is independent of $W_1$ and $W_2$.

Furthermore, the estimated regression parameters are asymptotically normal as in the continuous case and

$$\nu_n^{-2}(\hat{\tau} - \tau) \leq T^*/g(\tau)$$

where $T^* = \text{arg min}_{t} |B(t) + \frac{1}{2}\sigma^{-1}|c_0 + c_1\tau| |t|$, a Brownian motion with a linear drift over $(-\infty, \infty)$ and $g(\tau)$ is the density (limiting density in fixed design case) of the design points at $\tau$.

In the fixed jump size,

$$d = d_n = c_0 + \tau c_1 \neq 0.$$ (4.2)

Koul and Qian (2002) proposed likelihood ratio test under more general error density function $f$. Under i.i.d. errors, the part of the likelihood function of the data from (1.1) under $\theta$ that is relevant for making inference about $\theta$ is given by

$$L_n(\theta) := \prod_{i=1}^{n} f(Y_i - m(X_i, \theta)).$$

The maximization algorithm still take two steps. Recall, $\theta' = (\theta'_1, \tau)$ and $\theta'' = (\theta''_1, \tau')$. First, for each fixed $s$, obtain the maximizers $\bar{\theta}_1(s)$ of $L_n(\theta_1, s)$ with respect to $\theta_1$ over $K$. Notice that $\bar{\theta}_1(s)$ is constant in $s$ over any interval of two consecutive ordered $X_i's$ and that the profile likelihood function $L_n(\bar{\theta}_1(s), s)$ has only finite number of possible jumps with the jump points located at the ordered $X_i's$.

At the second stage, we compute maximizer $\bar{\tau}$ of $L_n(\bar{\theta}_1(\bar{\tau}), s)$ with respect to $s$ over the sample percentiles of $\{X_i, 1 \leq i \leq n\}$. Because of (4.2), for each $\bar{\tau}$ fixed, the likelihood as a function of $s$ changes only when the change-point parameter $s$ passes over the sample points of the independent variable $X$. This maximizer may be taken to be the left end point of the interval over which it is obtained. Then the estimator $\hat{\theta} = (\hat{\theta}_1(\hat{\tau}), \hat{\tau}) \equiv (\hat{\theta}_1, \hat{\tau})$ is the MLE of the underlying parameter $\theta$.

The general density function $f$ of error is required to satisfy the following conditions:

(a) $f$ is absolutely continuous and positive everywhere on $\mathbb{R}$ with $f$ denoting its a.e. derivative and $\varphi = f/f$.

(b) The function $\varphi$ is Lip(1) and $I(f) := \int \varphi^2(x) f(x) dx < \infty$.

(c) $\varphi$ is differentiable and its derivative $\varphi$ is Lip(1).

Furthermore, the design variable $X$ is assumed to have a bounded and positive Lebesgue density $g$ on $\mathbb{R}$ with $E|X|^3 < \infty$.

For $t \in \mathbb{R}$, let $P(t) = P_1(-t)I(t \leq 0) + P_2(t)I(t > 0)$. The two random walks $P_1(t)$ and $P_2(t)$ associated with this process are two independent compound Poisson processes on $[0, \infty)$, both with rate $g(\tau)$ and $P_1(0) = P_2(0) = 0$, a.s.; the distribution of jumps being given by the conditional distribution of $\ln[f(\epsilon + c_0 + c_1 X)/f(\epsilon)]$, given $X = \tau^+$ and the conditional distribution of $\ln[f(\epsilon - c_0 - c_1 X)/f(\epsilon)]$ given $X = \tau^-$, respectively. Furthermore, $P_1(t)$ and $P_2(t)$ tend to $-\infty$, a.s., as $|t| \to \infty$ which implies the existence of the unique random interval $[M_-, M_+]$ on which the process $\{P(t) : t \in \mathbb{R}\}$ attains its global maximum, a.s.

**Remark:** If the error term is Gaussian, then all conditions (a)-(c) imposed on the error density function are satisfied.

We are now ready to state the asymptotical properties of the estimators which includes the root-$n$ asymptotic normality for the regression parameters and $n$- consistency of the estimator $\hat{\tau}$ under general error density setting.
Theorem 4.3. (Fixed jump case) Suppose that conditions (a)-(c), (4.2) hold and \( g \) is continuous and positive at \( \tau \) and \( E|X|^3 < \infty \). Then, \( n(\hat{\tau} - \tau) \) converges weakly to \( M_- \) and \( n^{1/2}(\hat{\theta}_1 - \theta_1) \) converges weakly to a \( N(0, \Gamma^{-1}) \) r.v., where \( \Gamma = I(f)Em(X)m'(X) \) and

\[
\hat{m}_s(x) = (I(x \leq s), xI(x \leq s), I(x > s), xI(x > s))', \quad s \in \mathbb{R}, \ x \in \mathbb{R},
\]

Notice that

\[
Em(X)m'(X) = \begin{bmatrix}
G(\tau) & \mu_1(\tau) & 0 & 0 \\
\mu_1(\tau) & \mu_2(\tau) & 0 & 0 \\
0 & 0 & G(\tau) & \mu_1(\tau) \\
0 & 0 & \mu_2(\tau) & G(\tau)
\end{bmatrix} = \Sigma_r.
\]

Consequently, we conclude that \( \hat{\theta}_{11} \) and \( \hat{\theta}_{12} \) are asymptotically independent, which is the similar results as in continuous and contiguous cases.

Koul, Qian and Surgailis (KQS, 2003) extended the above results further to \( M \)-estimator and obtained some results. Detail proofs can be found in KQS (2003). Ciperca (2004) derived same results of KQ (2002) when extending the model to nonlinear model.

5 Computational Aspects

Almost all of computational algorithms in the literatures focus on continuous model. Hudson (1966) developed two-stage algorithm using so-called Type-I minimum and Type-II minimum for two phase regression in which the regression function, but not its derivative, is continuous at the change point (see Seber and Wild, p. 456-457). This algorithm was refined by Hinkley (1969, 1971) for two phase linear regression model, and further extended to three phase by Williams (1970). Lerman (1980) noticed the difficulty of solving the problem analytically using Hudson’s algorithm and proposed grid search method. Draper and Smith (1981, Section 5.4) suggested a grid search method based on the smallest residual sum of squares criterion to estimate the join point (Hudson’s idea), which is the same algorithm as in section 2.1.

Galant and Fuller (1973) proposed a different approach and advocated re-parameterizing the regression into a single nonlinear function using indicator functions and recommended using the Gauss-Newton procedure to help estimate the joint point directly. Hawkins (1976) proposed dynamic programming and hierarchical procedure.

Bacon and Watts (BW, 1971) use transition smoothing function \( \tanh \), also called \( sgn \) formulation. Tishler and Zang’s (TZ, 1981) proposed so-called bent-cable quadratic smoothing around the change point for an arbitrary number of regimes and variables with fast convergence, also named as max-min formulation.

Rewrite the continuous two phase linear model into

\[
Y_i = \begin{cases}
\{ b_0 + b_1 x_i + \varepsilon_i, & x_i \leq \tau, \\
b_0 + b_1 x_i + b_2 (x_i - \tau) + \varepsilon_i, & x_i \leq \tau.
\end{cases}
\]

Bacon and Watt (1971)’s transition model is defined as

\[
Y_i = b_0 + b_1 x_i + \frac{1}{2}(b_1 + b_2)(x_i - \tau) + \frac{1}{2}(b_2 - b_1) \tanh \left( \frac{x_i - \tau}{\gamma} \right) + \varepsilon_i
\]

where \( \gamma \) is the smoothing parameter and the transition function

\[
\tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}
\]

satisfies

\[
\lim_{\gamma \to 0} \frac{x - \tau}{\gamma} = \text{sgn}(x - \tau).
\]

and

\[
\text{sgn}(x) = \begin{cases}
-1, & x < 0, \\
0, & x = 0, \\
1, & x > 0.
\end{cases}
\]

Tishler and Zang discussed the piecewise linear model:

\[
Y_i = \max\{b'_1 X_{1i}, b'_2 X_{2i}, \ldots, b'_p X_{pi}\} + \varepsilon_i
\]

where \( b'_k = (b_{k1}, \ldots, b_{kp}) \) for \( k = 1, \ldots, p \). When \( p = 2 \), it reduces to the two phase linear model which can also be expressed as the following equivalent form:

\[
Y_i = b'_1 x_{i1} + q(b'_2 x_{i2} - b'_1 x_{i1}) + \varepsilon_i
\]

where

\[
q(x) = \max(0, x), \ x \in \mathbb{R}.
\]

Tishler and Zang approximate \( q(x) \) by \( \tilde{q}(x) \), which smoothes out the kink and uses a quadratic smoothing function, defined as:

\[
\tilde{q}(x) = \begin{cases}
0, & \text{if } x \leq -\beta, \\
\frac{(x+\beta)^2}{4\beta^2}, & \text{if } -\beta < x \leq \beta, \\
x, & \text{if } x > \beta.
\end{cases}
\]

The smoothed model is called TZ’s bent-cable model. Figure 5.1 shows the approximation of \( q(x) \) by \( \tilde{q}(x) \) for small \( \beta = .5 \).

We analyzed numerous real data sets using three fore-mentioned algorithms (BW, TZ, and Grid search) and concluded that the smoothing
6 Conclusions and Final Remarks

In this paper, we first motive readers with real life examples using multiple-phase linear regression models. Then, we address hypothesis testing issues through likelihood ratio and informational based approaches. Further, we review the asymptotic theorems for the two types of models: Continuous (fixed design and random design) and discontinuous (fixed design and random design) models. At last, we summarize the computational aspects for piecewise regression models. All results surveyed in this paper are focused on parametric approaches after 1989. Nonparametric, Bayesian and CUSUM approaches are also used in literatures.


Among Bayesian approaches, Bacon and Watts (1971) utilized Bayesian approach to estimate the joint point of two straight lines and introduced a transition function to represent the model at the intersection of the two lines by a smooth curve. El-Sayyad (1975) gave a Bayesian analysis of the change-point problem and applied it to the cases of normal and exponential distribution. Ferreira (1975) gave a Bayesian analysis of switching regression model with known number of regimes and showed that the mean square error (MSE) of the Bayesian estimates are uniformly smaller than those of the maximum likelihood (ML) estimates. Holbert and Bromelin (1977) derived Bayesian inferences related to shifting sequences and two-phase regression. Piegorsch (1987) considered Bayesian decision procedures with a discretized normal prior density as the prior distribution for the change point. Kezim and Abdelli (2004) conducted a Bayesian analysis of a structural change in the parameters of a time series.

Among the literatures surveyed, most of the researchers focus on the continuous model. On the other hand, these researchers advise caution...
in the use of the continuity assumption and illustrated, by examples, that it can lead to results different from those when no constraint is imposed. F-test using the extra sum of squares for testing two-phase versus simple linear regression was initially proposed by Hinkley, but the degrees of freedom used is a big mystery.

Furthermore, one notices that the types of models, in terms of fixed jump size versus contiguous or continuous, can result in huge difference for asymptotical theory as indicated by Hawkins (1980), Seber & Wild (1989), Koul and Qian (2002), Koul et al. (2003). To be more specific: we summarize as below:

- In the continuous and contiguous model, the asymptotic distribution of the maximum likelihood estimated change point between the two-phase is related to a linear drifted Brownian process for both fixed and random design cases.
- In discontinuous fixed jump size model, the limiting distribution of the MLE and M-estimation for the change point is related to a compound Poisson process, which shows the distinct difference between the types of discontinuous models: fixed jump size and contiguous cases.

Though in this paper, we only focus on homogeneous variance model, nonhomogeneous error variance is a realistic problem. Kim and Siegmund (KS, 1989) and Kim (K, 1993) studied the likelihood ratio test for a change point in simple linear regression with homogeneous and nonhomogeneous errors, respectively. Brodeau (1999) derived asymptotic results and tests for the choice of approximative models in nonlinear two-phase regression models, under heteroscedastic assumption. Diniz and Brochi (2005) compared the robustness of the tests proposed by Quandt (Q-test), KS-test and K-test through simulation, and concluded that KS-test is superior to Q-test for both models considered while the K-test is more powerful than the other two tests for nonhomogeneous models with known variances.

Recent development indicates that testing the existence of change point is still a hot topic. Juškovič (1998) developed a new test procedure for detecting gradual (contiguous) changes in simple regression models and studied its limit behavior under the null hypothesis 'no change'. Hušková and her collaborators in their series of papers (1998, 2001, 2002, 2003) proposed a new class of test procedures for testing structural change and studied the limit behaviors both under the null hypothesis (no change) and under some alternatives.


Besides the hot topic in change point detection, recent research activities show that the structural change problem in regression models has been extended to more general models such as time series (Qian, 1998; Maekawa, Hiz and Tee, 2004 and references therein), measurement error models (Gbur and Dahm, 1985; Staudermayer and Spiegelman, 2002), generalized (bi)linear models (Guegan and Pham, 1989; Gabriel, 1998; Pastor and Guallar, 1998; Antoch, Gregoire and Jaruszková, 2004) and analyzing longitudinal data (Piepho and Ogutu, 2003).

References


[26] Davies, R. B., Hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrika*, 64 (1977), 247–254.

[27] Davies, R. B., Hypothesis testing when a nuisance parameter is present only under the alternative. *Biometrika*, 74 (1987) 33–43.


