axioms

MDPI

Article

Change-Point Estimation and Detection for Mixture of Linear Regression Models

Wenzhi Zhao [1], Tian Cheng [2] and Zhiming Xia [2,*]

[1] School of Science, Xi'an Polytechnic University, Xi'an 710048, China; zhaowenzhi@xpu.edu.cn
[2] School of Mathematics, Northwest University, Xi'an 710127, China; 201820540@nwu.edu.cn
* Correspondence: statxzm@nwu.edu.cn

Abstract: This paper studies the estimation and detection problems in the mixture of linear regression models with change point. An improved Expectation–Maximization (EM) algorithm is devised specifically for multi-classified mixture data with change points. Under appropriate conditions, the large-sample properties of the estimator are rigorously proven. This improved EM algorithm not only precisely locates the change points but also yields accurate parameter estimates for each class. Additionally, a detector grounded in the score function is proposed to identify the presence of change points in mixture data. The limiting distributions of the test statistics under both the null and alternative hypotheses are systematically derived. Extensive simulation experiments are conducted to assess the effectiveness of the proposed method, and comparative analyses with the conventional EM algorithm are performed. The results clearly demonstrate that the EM algorithm without considering change points exhibits poor performance in classifying data, often resulting in the misclassification or even omission of certain classes. In contrast, the estimation method introduced in this study showcases remarkable accuracy and robustness, with favorable empirical sizes and powers.

Keywords: change point; mixture of linear regression model; multi-classification; EM algorithm

MSC: 62F10; 62J05



Academic Editor: Ioan M. Stancu-Minasian

Received: 15 April 2025
Revised: 14 May 2025
Accepted: 19 May 2025
Published: 26 May 2025

Citation: Zhao, W.; Cheng, T.; Xia, Z. Change-Point Estimation and Detection for Mixture of Linear Regression Models. *Axioms* **2025**, *14*, 402. https://doi.org/10.3390/axioms14060402

Copyright: © 2025 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Mixture models have found extensive applications in econometrics and social sciences, and the associated theoretical frameworks have been thoroughly investigated. Among them, finite mixtures of linear regression models, a particularly valuable class of mixture models, have been widely employed across diverse fields, including econometrics, marketing, epidemiology, and biology. Simultaneously, the statistical inference and parameter estimation based on mixture models play a pivotal role in practical applications. On another front, the change point problem remains a focal point in statistical research. Conceptually, a change point denotes the specific location or instant at which a sudden alteration in a statistical property occurs. Change point phenomena are pervasive in both natural and social systems, and have enabled crucial advancements in numerous domains such as industrial control, financial economics, biomedicine, and signal processing. The detection and estimation of change points within mixture datasets are essential for discerning underlying patterns and making informed decisions, thereby holding significant practical implications.

In recent decades, remarkable progress has been achieved in the analysis of mixture models. Goldfeld and Quandt provided an outstanding introduction to mixtures of linear regression models [1] . They delved into the parameter estimation problem of such mixtures

and applied it to analyze the imbalance in the housing price market. Frühwirth and Schnatter summarized the Bayesian approach for mixtures of linear regression models in their paper [2]. Hurn and Justel proposed a generalized linear finite mixture model and overcame the label switching issue by normalizing the loss function [3]. Li, Chen, and their collaborators (Chen, Li, and Fu) concentrated on the hypothesis testing problem regarding the order of the mixture model and derived the limiting distribution of the test statistics [4,5]. Huang and Yao introduced a semiparametric mixture regression model, where the mixing proportion is a smooth function of covariates [6]. Huang, Li, and Wang developed a non-parametric finite mixture method for regression models and demonstrated the estimation procedure through an analysis of U.S. Housing Price Index (HPI) data [7].

The change point problem has attracted extensive research attention, with a vast body of literature accumulated over the years. It was first introduced by Page during his study of continuous inspection schemes, who proposed the Cumulative Sum (CUSUM) test method [8]. Subsequently, scholars like Sen and Srivastava, Hawkins, James et al., and Srivastava and Worsley investigated the detection of change points in the mean of normal random variable sequences [9–13]. Basseville and Nikiforov systematically expounded the theoretical foundation of change point detection and estimation algorithms [14]. Wang, Zou, and Yin delved into change point detection in multinomial data as the number of categories approaches infinity, rigorously proving the statistical properties of relevant test statistics [15]. Xia and Qiu proposed the JIC criterion to tackle the multi-change point problem in non-parametric regression models [16]. Bai employed the least squares method to estimate the mean change point of linear processes, deriving the consistency and convergence rate of the change point estimator [17]. Baranowski et al. utilized grid search algorithms based on binary segmentation to accurately determine the number and positions of change points [18]. Follain et al. developed a novel approach for estimating change points in partially observed, high-dimensional time series with simultaneous mean shifts in sparse coordinate subsets [19]. Gong et al. redefined the graph-based change point detection problem as a prediction task and proposed an innovative change point detection (CPD) method for dynamic graphs through a latent evolution model. Drabech et al. introduced a Markov Random Field (MRF) model to detect slope changes [20], while Ratnasingam et al. proposed an empirical likelihood-based non-parametric procedure for identifying structural changes in quantile regression models [21]. Notably, despite these abundant studies, the statistical analysis of change points within finite mixtures of linear models remains a relatively overlooked area [22].

This paper delves into the estimation and detection problems of the finite mixture of linear regression models with change points. Specifically, for multi-classified mixture data incorporating change points, we ingeniously propose an improved Expectation–Maximization (EM) algorithm. This algorithm is meticulously designed to accurately compute parameter estimators, such as the location of change points and regression coefficients. Through rigorous mathematical proofs, we firmly establish the consistency and asymptotic normality of these parameter estimators, providing a solid theoretical foundation. Moreover, to effectively detect the presence of change points in the mixture of linear regression models, we innovatively develop a detector based on the score function. By conducting in-depth analyses, we systematically derive the limiting distributions under both the null hypothesis and the alternative hypothesis, which significantly contributes to the theoretical framework of this study. To thoroughly verify the practical effectiveness of the proposed method, we carefully design and execute a series of simulation experiments.

The remaining sections of this paper are structured as follows. In Section 2, we introduce the estimation and detection methods, statistical models, along with their associated assumptions. In Section 3 is dedicated to the development of the estimation and detection

techniques for change points within the mixture of linear regression models. Additionally, we obtain the large sample properties of the estimators and test statistics. In Section 4, the simulation results are presented. Section 5 provides some concluding remarks and discussions. In Section 6, the limitations and future research directions are presented. The proofs of the theorems are included in the appendix.

2. Statistical Model and Assumptions

Assume that $\{(X_i, Y_i), i = 1, \cdots, n\}$ are independent random samples from the following data generating process:

$$Y_i \sim \begin{cases} \sum_{c=1}^{C} \pi_c^{*-} N\left(X_i^T \beta_c^{*-}, \sigma_c^{2*-}\right), & 1 \le i \le k^*, \\ \sum_{c=1}^{C} \pi_c^{*+} N\left(X_i^T \beta_c^{*+}, \sigma_c^{2*+}\right), & k^* + 1 \le i \le n, \quad k^* \in [1, n-1], \end{cases} \tag{1}$$

where k^* is a change point, X_i is a p-dimensional random covariate, C is the number of components, $\beta_c^{*-}, \pi_c^{*-}, \sigma_c^{*-}$, $\beta_c^{*+}, \pi_c^{*+}, \sigma_c^{*+}$ are mixture coefficients for the cth component, $1 \ge \pi_c^{*-} \ge 0$, $\pi_c^{*+} \ge 0$, $\sum_{c=1}^{C} \pi_c^{*-} = \sum_{c=1}^{C} \pi_c^{*+} = 1$. π_c^{*-} and π_c^{*+} are called mixing proportions or weights. $\beta_c^{*-}, \beta_c^{*+} \in R^p$, $\sigma_c^{2*-}, \sigma_c^{2*+} \in R^+$ and the parameters are unknown. Equivalently speaking, $\{Y_i, i = 1, \cdots, n\}$ follow a finite mixture of normals,

$$Y_i \sim \sum_{c=1}^{C} \pi_c^{*-} N(X_i^T \beta_c^{*-}, \sigma_c^{2*-}) \cdot I_{\{i \le k^*\}} + \sum_{c=1}^{C} \pi_c^{*+} N(X_i^T \beta_c^{*+}, \sigma_c^{2*+}) \cdot I_{\{i \ge k^*+1\}}.$$

Naturally, the log-likelihood function corresponding to the dataset $\{(X_i, Y_i), i = 1, \cdots, n\}$ can be expressed as

$$\begin{aligned} &l(k, \pi_c^-, \beta_c^-, \sigma_c^{2-}, \pi_c^+, \beta_c^+, \sigma_c^{2+}) = \\ &\sum_{i=1}^{n} \log \left[\sum_{c=1}^{C} \left\{ \pi_c^- \phi\left(Y_i | X_i^T \beta_c^-, \sigma_c^{2-}\right) \cdot I_{\{i \le k\}} + \pi_c^+ \phi\left(Y_i | X_i^T \beta_c^+, \sigma_c^{2+}\right) \cdot I_{\{i \ge k+1\}} \right\} \right] \end{aligned} \tag{2}$$

where $\phi(y|\mu, \sigma^2)$ denotes the normal density.

For later use, we introduce some notations. Let $\theta^* = \left((\pi^*)^T, (\beta^*)^T, (\sigma^{2*})^T\right)^T$ denote the true parameter, where $\pi^* = (\pi_1^*, \cdots, \pi_{C-1}^*)^T$, $\beta^* = (\beta_1^{*T}, \cdots, \beta_C^{*T})^T$, $\sigma^{2^*} = (\sigma_1^{2^*}, \cdots, \sigma_C^{2^*})^T$ and $\beta_i^* \in R^P$, $i = 1, \cdots, C$. Let θ^{*-}, θ^{*+} represent the parameters on the left and right sides of the change point, respectively. Similarly, we can define θ, θ^-, θ^+.

Let $\hat{k}_n$ and $\hat{\theta}_n$ be the maximum likelihood estimator of k^* and θ, which can be obtained by maximizing (2). In Section 3.1, an improved EM algorithm is proposed to carry out the estimation. The following assumptions are imposed.

Assumption 1. *$\frac{k^*}{n} \to \tau^*$, $n \to \infty$ and $\tau^* \in (0, 1)$.*

The above assumption requires that k^* depends on the sample size n and is approximately proportional to n.

Assumption 2. *We suppose that $\theta \in \Theta$ and parameter space $\Theta \subset R^q$ is a compact set, $q = (p + 2) \cdot C - 1$.*

Assumption 2 restricts the parameter space to be compact, which means any open cover of Θ has a finite subcover.

Assumption 3. $\{(X_i, Y_i),\ i = 1, 2, \cdots, n\}$ *are independent.*

3. Parameter Inference

3.1. Estimation Procedure

In this subsection, we propose an effective EM algorithm to deal with the parameter estimation problem. We take $i \geq k+1$ as an example to illustrate the derivation of the iterative formula ($i < k$ can be obtained in the same way). At this time, the log-likelihood function corresponding to the data is

$$\sum_{i=k+1}^{n} \log \left[\sum_{c=1}^{C} \pi_c^+ \phi\{Y_i | X_i^T \boldsymbol{\beta}_c^+, \sigma_c^{2+}\} \right]. \tag{3}$$

In the EM framework, the mixture problem is described as an incomplete-data problem. We view the observed data $(X_i, Y_i, i = 1, 2, \cdots, n)$ as being incomplete, and then introduce unobserved *Bernoulli* random variables

$$z_{ic} = \begin{cases} 1, & if (X_i, Y_i) \text{ is in the cth group}, \\ 0, & otherwise, \end{cases}$$

$$i = 1, 2, \cdots, n, \quad c = 1, 2, \cdots, C.$$

and $\boldsymbol{z}_i = (z_{i1}, \cdots, z_{iC})^T$, which is the component label of (X_i, Y_i). Therefore, the complete data are $\{(X_i, Y_i, \boldsymbol{z}_i)\}_{i=1}^n$, and the complete log-likelihood function corresponding to (3) is

$$\sum_{i=k+1}^{n} \sum_{c=1}^{C} z_{ic} \Big[\log \pi_c^+ + \log \phi\{Y_i | X_i^T \boldsymbol{\beta}_c^+, \sigma_c^{2+}\} \Big].$$

Supposing in the lth cycle of the EM algorithm iteration, we have $\pi_c^{+(l)}, \boldsymbol{\beta}_c^{+(l)}, \sigma_c^{2+(l)}$; then, in the E-step of $(l+1)$th cycle, the expectation of the latent variable z_{ic} can be calculated by

$$r_{ic}^{(l+1)} = \frac{\pi_c^{+(l)} \phi\{Y_i | X_i^T \boldsymbol{\beta}_c^{+(l)}, \sigma_c^{2+(l)}\}}{\sum_{c=1}^{C} \pi_c^{+(l)} \phi\{Y_i | X_i^T \boldsymbol{\beta}_c^{+(l)}, \sigma_c^{2+(l)}\}}.$$

In the M-step of the $(l+1)$th cycle, we maximize

$$\sum_{i=k+1}^{n} \sum_{c=1}^{C} r_{ic}^{(l+1)} \Big[\log \pi_c^+ + \log \phi\{Y_i | X_i^T \boldsymbol{\beta}_c^+, \sigma_c^{2+}\} \Big]. \tag{4}$$

The maximization of Equation (4) is equivalent to separately maximize $\sum_{i=k+1}^{n} \sum_{c=1}^{C} r_{ic}^{(l+1)} \log \pi_c^+$ and $\sum_{i=k+1}^{n} \sum_{c=1}^{C} r_{ic}^{(l+1)} \log \phi\{Y_i | X_i^T \boldsymbol{\beta}_c^+, \sigma_c^{2+}\}$. Furthermore, the iterative formula can be easily derived as follows:

$$\pi_c^{+(l+1)} = \frac{\sum_{i=k+1}^{n} r_{ic}^{(l+1)}}{n},$$

$$\boldsymbol{\beta}_c^{+(l+1)} = \Big(\sum_{i=k+1}^{n} X_i r_{ic}^{(l+1)} X_i^T \Big)^{-1} \Big(\sum_{i=k+1}^{n} X_i r_{ic}^{(l+1)} Y_i \Big),$$

$$\sigma_c^{2+(l+1)} = \frac{\sum_{i=k+1}^{n} r_{ic}^{(l+1)} (Y_i - X_i^T \boldsymbol{\beta}_c^{+(l)})(Y_i - X_i^T \boldsymbol{\beta}_c^{+(l)})^T}{\sum_{i=k+1}^{n} r_{ic}^{(l+1)}}, \quad c = 1, 2, \cdots, C$$

Let the initial value be $\pi_c^{+(0)}, \boldsymbol{\beta}_c^{+(0)}, \sigma_c^{+(0)}, c = 1, \cdots, C$, repeat the above iteration algorithm, and when the iteration converges, we obtain the estimator of mixture parameters, denoted as $\pi_c^+(k), \boldsymbol{\beta}_c^+(k), \sigma_c^{2+}(k), c = 1, \cdots, C$. At this time, the log-likelihood function corresponding to the data $\{(X_i, Y_i)\}_{i=k+1}^n$ is

$$l^+(k, \boldsymbol{\pi}_c^+, \boldsymbol{\beta}_c^+, \sigma_c^{2+}) = \sum_{i=k+1}^{n} \log \left[\sum_{c=1}^{C} \pi_c^+(k) \phi\{Y_i | X_i^T \boldsymbol{\beta}_c^+(k), \sigma_c^{2+}(k)\} \right].$$

Whenever a value of k is determined, the mixture data are divided into two parts: $\{(X_i, Y_i)\}_{i=1}^k$ and $\{(X_i, Y_i)\}_{i=k+1}^n$. Let the initial value be $\pi_c^{\pm(0)}, \boldsymbol{\beta}_c^{\pm(0)}, \sigma_c^{\pm(0)}$, $c = 1, \cdots, C$, and we can obtain the estimations of change point and mixture parameter according to Algorithm 1:

$$(\hat{k}, \hat{\boldsymbol{\theta}^-}, \hat{\boldsymbol{\theta}^+}) = \underset{k, \boldsymbol{\pi}_c^{\pm}, \boldsymbol{\beta}_c^{\pm}, \sigma_c^{\pm}}{\arg\max}\ l(k, \boldsymbol{\pi}_c^-, \boldsymbol{\beta}_c^-, \sigma_c^{2-}, \boldsymbol{\pi}_c^+, \boldsymbol{\beta}_c^+, \sigma_c^{2+}),$$

where $\hat{\boldsymbol{\theta}}^{\pm} = ((\hat{\boldsymbol{\pi}}^{\pm})^T, (\hat{\boldsymbol{\beta}}^{\pm})^T, (\hat{\boldsymbol{\sigma}}^{2\pm})^T)^T$, $\hat{\boldsymbol{\pi}}^{\pm} = (\hat{\pi}_1^{\pm}, \cdots, \hat{\pi}_C^{\pm})^T$ and $\hat{\boldsymbol{\beta}}^{\pm} = (\hat{\beta}_1^{\pm}, \cdots, \hat{\beta}_C^{\pm})^T$, $\hat{\boldsymbol{\sigma}}^{2\pm} = (\hat{\sigma}_1^{2\pm}, \cdots, \hat{\sigma}_C^{2\pm})^T$, $l(k, \boldsymbol{\pi}_c^-, \boldsymbol{\beta}_c^-, \sigma_c^{2-}, \boldsymbol{\pi}_c^+, \boldsymbol{\beta}_c^+, \sigma_c^{2+}) = l^+(k, \boldsymbol{\pi}_c^+, \boldsymbol{\beta}_c^+, \sigma_c^{2+}) + l^-(k, \boldsymbol{\pi}_c^-, \boldsymbol{\beta}_c^-, \sigma_c^{2-})$.

Algorithm 1 EM algorithm considering change point

Intput: $\{X_i, Y_i\}_{i=1}^n$, the number of mixture component C, the maximum iterations T.
Output: $\hat{k}, \hat{\boldsymbol{\theta}}^-, \hat{\boldsymbol{\theta}}^+$

1: **for** $k = 1, 2, \cdots, n-1$ **do**
2: For any given k, the mixture data are divided into two parts: $\{(X_i, Y_i)\}_{i=1}^k$ and $\{(X_i, Y_i)\}_{i=k+1}^n$. Using EM algorithm separately for two parts of data. For $\{(X_i, Y_i)\}_{i=1}^k$, let the initial value be $\pi_c^{-(0)}, \boldsymbol{\beta}_c^{-(0)}, \sigma_c^{-(0)}, c = 1, \cdots, C$.
3: **for** $t = 1, 2, \cdots, T$ **do**
4: E-step: calculate $r_{ic}^{(l+1)}(k)$,

$$r_{ic}^{(l+1)}(k) = \frac{\pi_c^{-(l)} \phi\{Y_i | X_i^T \boldsymbol{\beta}_c^{-(l)}, \sigma_c^{2-(l)}\}}{\sum_{c=1}^{C} \pi_c^{-(l)} \phi\{Y_i | X_i^T \boldsymbol{\beta}_c^{-(l)}, \sigma_c^{2-(l)}\}},\ i = 1, \cdots, k, c = 1, \cdots, C$$

5: M-step: calculate

$$\pi_c^{-(l+1)}(k) = \frac{\sum_{i=1}^n r_{ic}^{(l+1)}}{n}, \quad \boldsymbol{\beta}_c^{-(l+1)}(k) = (\sum_{i=1}^{n} X_i r_{ic}^{(l+1)} X_i^T)^{-1} (\sum_{i=1}^{n} X_i r_{ic}^{(l+1)} Y_i)$$

$$\sigma_c^{2-(l+1)}(k) = \frac{\sum_{i=1}^n r_{ic}^{(l+1)} (Y_i - X_i^T \boldsymbol{\beta}_c^{-(l)})(Y_i - X_i^T \boldsymbol{\beta}_c^{-(l)})^T}{\sum_{i=1}^n r_{ic}^{(l+1)}}$$

6: **end for**
7: When the iteration reaches convergence, we obtain the estimator $\pi_c^-(k), \boldsymbol{\beta}_c^-(k), \sigma_c^{2-}(k), c = 1, \cdots, C$ and the log-likelihood function value $l^-(k)$
8: For $\{(X_i, Y_i)\}_{i=k+1}^n$, repeat steps 3–7 and obtain $\pi_c^+(k), \boldsymbol{\beta}_c^+(k), \sigma_c^{2+}(k)$, $c = 1, \cdots, C$ and $l^+(k)$
9: **end for**
10: $\hat{k} = \underset{1 \le k \le n-1}{\arg\max}(l^-(k) + l^+(k)), \quad \hat{\boldsymbol{\theta}}^- = \hat{\boldsymbol{\theta}^-}(\hat{k}), \quad \hat{\boldsymbol{\theta}}^+ = \hat{\boldsymbol{\theta}^+}(\hat{k})$,

3.2. Hypothesis Test

Define the following notations for later use:

$$\eta(Y|\boldsymbol{\theta}) = \sum_{c=1}^{C} \pi_c \phi\{X, Y|\beta_c, \sigma_c^2\}, \quad \ell(\boldsymbol{\theta}, Y) = \log \eta(Y|\boldsymbol{\theta}),$$

$$q_1\{\boldsymbol{\theta}, Y\} = \frac{\partial \ell\{\boldsymbol{\theta}, Y\}}{\partial \boldsymbol{\theta}}, \quad q_2\{\boldsymbol{\theta}, Y\} = \frac{\partial^2 \ell\{\boldsymbol{\theta}, Y\}}{\partial \boldsymbol{\theta} \partial \boldsymbol{\theta}^T},$$

$$\mathcal{I}(\boldsymbol{\theta}) = -E[q_2\{\boldsymbol{\theta}, Y\}], \quad \mathcal{B}(\boldsymbol{\theta}) = Var[q_1\{\boldsymbol{\theta}, Y\}] = E[q_1\{\boldsymbol{\theta}, Y\} q_1^T\{\boldsymbol{\theta}, Y\}].$$

In the following, we are interested in testing the null hypothesis:

$$H_0: \ \boldsymbol{\theta}_i = \boldsymbol{\theta}^* \quad (i = 1, \cdots, n)$$

against the local alternative hypothesis:

$$H_A: \boldsymbol{\theta}_i = \boldsymbol{\theta}^* + \frac{1}{\sqrt{n}} \cdot \boldsymbol{g}(\frac{i}{n}),$$

where $\boldsymbol{g} = (g_1, g_2, \cdots, g_q)^T$ is a function of bounded variation on [0,1] which describes the patter of departure from stability of the parameter $\boldsymbol{\theta}^*$.

Firstly, we give the following lemmas and then we build an empirical process and infer its limiting process under the H_0 and H_1 hypothesis, respectively, as described by Theorem 1.

Theorem 1. *Under H_0 and Lemma A3 in the Appendix A, given by*

$$\Delta_n(t, \hat{\boldsymbol{\theta}}_n) = \frac{1}{\sqrt{n}} \sum_{i=1}^{[nt]} q_1(\hat{\boldsymbol{\theta}}_n, Y_i), \quad t \in (0, 1), \tag{5}$$

then, we have

(1). under H_0,

$$\hat{\mathcal{B}}_n^{-1/2} \Delta_n(t, \hat{\boldsymbol{\theta}}_n) \xrightarrow{d} \boldsymbol{W}^0(t), n \to \infty,$$

where $\boldsymbol{W}^0(t)$ is a q-dimensional standard Brownian bridge with $\boldsymbol{W}^0(t) = \boldsymbol{W}(t) - t\boldsymbol{W}(1)$ and $\hat{\mathcal{B}}_n$ some consistent covariance matrix estimation, such as $\hat{\mathcal{B}}_n = \frac{1}{n} \sum_{i=1}^{n} q_1(\hat{\boldsymbol{\theta}}_n, Y_i) q_1(\hat{\boldsymbol{\theta}}_n, Y_i)^T$.

(2). Under H_A,

$$\hat{\mathcal{B}}_n^{-1/2} \Delta_n(t, \hat{\boldsymbol{\theta}}_n) \xrightarrow{d} W^0(t) + \mathcal{B}(\boldsymbol{\theta}^*)^{1/2} \boldsymbol{G}^0(t), n \to \infty,$$

where $\boldsymbol{G}^0(t) = \boldsymbol{G}(t) - t\boldsymbol{G}(1)$ with $\boldsymbol{G}(t) = \int_0^t \boldsymbol{g}(y) dy$.

Next, we construct test statistics according to the empirical process (5) stated above. The key idea is that the empirical process reflects some symptoms of structural changes, and its performance under H_0 and H_A is significantly different; then, the test statistics can be constructed as follows:

$$T_n = \sup_{1 \le k \le n} \left\| \hat{\mathcal{B}}_n^{-1/2} \Delta_n(k/n, \hat{\boldsymbol{\theta}}_n) \right\|^2.$$

Further from Theorem 1 and continuous mapping theorem, the following corollary can be obtained:

Corollary 1. *Given the empirical process* (5) *in Theorem* 1*, we have*
(1). under H_0*,*

$$T_n \xrightarrow{d} \sup_{t\in[0,1]} \left\| \boldsymbol{W}^0(t) \right\|^2 = \sup_{t\in[0,1]} \left\{ \sum_{i=1}^{q} W_i^0(t)^2 \right\}, \quad n \to \infty. \tag{6}$$

(2). Under H_A*,*

$$T_n \xrightarrow{d} \sup_{t\in[0,1]} \left\| \boldsymbol{W}^0(t) + \mathcal{B}(\boldsymbol{\theta}^*)^{1/2} \boldsymbol{G}^0(t) \right\|^2, \quad n \to \infty$$

The above results show that the limiting distributions of the test statistics are different under the null and alternative hypothesis; thus, we can judge the existence of the change point according to it.

3.3. Consistency

In order to prove the consistency of the estimator, we first construct a function $\mathcal{L}_n$ as follows:

$$\mathcal{L}_n(\tau, \boldsymbol{\theta}) = \frac{1}{n}\sum_{i=1}^{n} \tilde{f}(X_i, Y_i|t, \boldsymbol{\theta}) = \frac{1}{n}\sum_{i=1}^{[n\tau]} \tilde{f}(X_i, Y_i|t, \boldsymbol{\theta}) + \frac{1}{n}\sum_{i=[n\tau]+1}^{n} \tilde{f}(X_i, Y_i|t, \boldsymbol{\theta}), \tag{7}$$

where $[n\tau] = k$, $[a]$ means the integer part of a and

$$\tilde{f}(X_i, Y_i|t, \boldsymbol{\theta}) = \log\left[\sum_{c=1}^{C} \left\{ \pi_c^- \phi(Y_i|X_i^T\beta_c^-, \sigma_c^{2-}) I_{\{i\le[nt]\}} + \pi_c^+ \phi(Y_i|X_i^T\beta_c^+, \sigma_c^{2+}) I_{\{i\ge[nt]+1\}} \right\} \right].$$

Without loss of generality, we only consider the case $k \le k^*$ for brevity; then, by *Kolmogorov*'s Law of Large Numbers and *Jessen* inequality, the Formula (7) can be rewritten as

$$\begin{aligned}
\mathcal{L}_n(\tau, \boldsymbol{\theta}) &= \frac{1}{n}\sum_{i=1}^{[n\tau]} f(X_i, Y_i|\tau, \boldsymbol{\theta}^-) + \frac{1}{n}\sum_{i=[n\tau]+1}^{[n\tau^*]} f(X_i, Y_i|\tau, \boldsymbol{\theta}^-) + \frac{1}{n}\sum_{i=[n\tau^*]+1}^{n} f(X_i, Y_i|\tau, \boldsymbol{\theta}^+) \\
&= \frac{[n\tau]}{n}\cdot\frac{1}{[n\tau]}\sum_{i=1}^{[n\tau]} f(X_i, Y_i|\tau, \boldsymbol{\theta}^-) + \frac{[n\tau^*]-[n\tau]}{n}\cdot\frac{1}{[n\tau^*]-[n\tau]}\sum_{i=[n\tau]+1}^{[n\tau^*]} f(X_i, Y_i|\tau, \boldsymbol{\theta}^-) \\
&\quad + \frac{n-[n\tau^*]}{n}\cdot\frac{1}{n-[n\tau^*]}\sum_{i=[n\tau^*]+1}^{n} f(X_i, Y_i|\tau, \boldsymbol{\theta}^+) \\
&\xrightarrow{p} \tau\cdot E_{(\tau^*,\boldsymbol{\theta}^{*-})} f(X, Y|\tau, \boldsymbol{\theta}^-) + (\tau^*-\tau)\cdot E_{(\tau^*,\boldsymbol{\theta}^{*-})} f(X, Y|\tau, \boldsymbol{\theta}^-) \\
&\quad + (1-\tau^*)\cdot E_{(\tau^*,\boldsymbol{\theta}^{*+})} f(X, Y|\tau, \boldsymbol{\theta}^+) \\
&\triangleq \mathcal{L}(\tau, \boldsymbol{\theta}|\tau^*, \boldsymbol{\theta}^*)
\end{aligned}$$

where

$$f(X_i, Y_i|\tau, \boldsymbol{\theta}^-) = \tilde{f}(X_i, Y_i|\tau, \boldsymbol{\theta})|_{\boldsymbol{\theta}=\boldsymbol{\theta}^-}, \;\; f(X_i, Y_i|\tau, \boldsymbol{\theta}^+) = \tilde{f}(X_i, Y_i|\tau, \boldsymbol{\theta})|_{\boldsymbol{\theta}=\boldsymbol{\theta}^+}.$$

Now, we state our second theoretical result, which is the consistency of the obtained estimator.

Theorem 2. *Let* $\hat{\tau} = \frac{\hat{k}_n}{n}$ *and under the conditions of the Lemmas A1 and A2 in Appendix A; then, we have*

$$(\hat{\tau}, \hat{\boldsymbol{\theta}}_n) \xrightarrow{p} (\tau^*, \boldsymbol{\theta}^*).$$

4. Simulation

To assess the performance of our proposed method, we shall consider two simulation experiments as follows.

4.1. Estimation Experiment

In Experiment 1, we focus on the scenario where the vector $\boldsymbol{\beta}$ is two-dimensional and the value of C is set to 2. Additionally, when dealing with the parameter estimation problem of mixed models that incorporate change points, the conventional approach is to overlook the presence of these change points. Instead, the Expectation–Maximization (EM) algorithm is directly employed to estimate the unknown parameters, as if the data were generated from a mixed model consisting of $2C$ components. Consequently, we conduct a comparison between the estimation results obtained by the EM method that does not take change points into account (referred to as "NCPEM") and those derived from our improved EM method ("CPEM"). We utilize the data generation mechanism of Model (1) to generate the data. Subsequently, we construct a two-component mixed model with a change point and proceed to specify the parameters as follows:

$$(\boldsymbol{\beta}_1^{*-})^T = (1,2),\ \sigma_1^{2*-} = 0.1,\ \pi_1^{*-} = 0.4,\ (\boldsymbol{\beta}_2^{*-})^T = (1,3),\ \sigma_2^{2*-} = 0.2,\ \pi_2^{*-} = 0.6.$$
$$(\boldsymbol{\beta}_1^{*+})^T = (3,4),\ \sigma_1^{2*+} = 0.2,\ \pi_1^{*+} = 0.7,\ (\boldsymbol{\beta}_2^{*+})^T = (3,5),\ \sigma_2^{2*+} = 0.2,\ \pi_2^{*+} = 0.3.$$

$\boldsymbol{X}_i^T = (1, X_i)$ and $X_i \sim U(0,1)$ for $i = 1, 2, \cdots, n$. Let the total sample size n from $\{500, 1000, 2000\}$ and the change point k^* be $\{250, 500, 1000\}$, respectively. The change point estimators and the corresponding standard errors are $\hat{k} = 250.45(0.730), 500.18(0.458), 1000.5(0.759)$ over 100 simulations. It can be seen that the change point estimator is close to the true value. The mixture parameter estimators and the corresponding standard errors are presented in Tables 1–3.

Table 1. Estimated value and standard error of $\boldsymbol{\beta}^*$.

	$\boldsymbol{\beta}^*$	$(\boldsymbol{\beta}_1^{*-})^T = (1,2)$	$(\boldsymbol{\beta}_2^{*-})^T = (1,3)$	$(\boldsymbol{\beta}_1^{*+})^T = (3,4)$	$(\boldsymbol{\beta}_2^{*+})^T = (3,5)$
$n = 500$	CPEM	(0.994, 2.004)	(0.975, 3.023)	(2.993, 4.004)	(2.999, 4.999)
		(0.045, 0.029)	(0.150, 0.129)	(0.064, 0.041)	(0.085, 0.057)
	NCPEM	(2.235, 1.911)	(0.988, 3.048)	(3.086, 3.874)	(3.022, 4.941)
		(9.906, 1.624)	(7.933, 1.290)	(9.285, 1.609)	(0.511, 0.278)
$n = 1000$	CPEM	(0.982, 2.054)	(1.006, 2.993)	(2.992, 4.005)	(2.996, 5.001)
		(0.042, 0.017)	(0.104, 0.077)	(0.039, 0.024)	(0.058, 0.038)
	NCPEM	(0.452, 2.220)	(0.884, 3.044)	(1.550, 4.192)	(3.187, 4.939)
		(8.490, 1.664)	(3.710, 0.678)	(8.211, 1.367)	(2.540, 0.484)
$n = 2000$	CPEM	(1.000, 1.999)	(1.014, 2.995)	(2.930, 4.017)	(3.001, 4.997)
		(0.022, 0.014)	(0.069, 0.046)	(0.030, 0.019)	(0.041, 0.035)
	NCPEM	(1.225, 2.054)	(1.318, 2.975)	(2.880, 3.921)	(3.009, 4.975)
		(4.212, 0.624)	(2.383, 0.505)	(0.557, 0.342)	(0.182, 0.107)

Table 2. Estimated value and standard error of σ^{2*}.

	σ^{2*}	$\sigma_1^{2*-} = 0.1$	$\sigma_2^{2*-} = 0.2$	$\sigma_1^{2*+} = 0.1$	$\sigma_2^{2*+} = 0.2$
n = 500	CPEM	0.099 (0.006)	0.199 (0.012)	0.100 (0.008)	0.197 (0.011)
	NCPEM	0.193 (0.466)	0.346 (0.626)	0.171 (0.436)	0.441 (0.786)
n = 1000	CPEM	0.098 (0.011)	0.200 (0.005)	0.099 (0.006)	0.199 (0.008)
	NCPEM	0.143 (0.287)	0.375 (0.686)	0.147 (0.338)	0.376 (0.661)
n = 2000	CPEM	0.100 (0.003)	0.203 (0.015)	0.099 (0.011)	0.204 (0.005)
	NCPEM	0.153 (0.339)	0.333 (0.616)	0.096 (0.022)	0.376 (0.687)

Table 3. Estimated value and standard error of π^*.

	π^*	$\pi_1^{*-} = 0.6$	$\pi_2^{*-} = 0.4$	$\pi_1^{*+} = 0.3$	$\pi_2^{*+} = 0.7$
n = 500	CPEM	0.601 (0.031)	0.399 (0.031)	0.301 (0.030)	0.699 (0.030)
	NCPEM	0.182 (0.053)	0.310 (0.050)	0.189 (0.045)	0.320 (0.062)
n = 1000	CPEM	0.594 (0.063)	0.406 (0.063)	0.295 (0.019)	0.705 (0.019)
	NCPEM	0.189 (0.060)	0.310 (0.044)	0.193 (0.039)	0.308 (0.056)
n = 2000	CPEM	0.599 (0.014)	0.401 (0.013)	0.298 (0.033)	0.702 (0.033)
	NCPEM	0.194 (0.035)	0.311 (0.058)	0.186 (0.049)	0.310 (0.045)

As is evident from Table 1, the estimation results obtained by the Change Point Expectation–Maximization (CPEM) method are much closer to the true parameter values compared to those derived from the Non-Change Point Expectation–Maximization (NCPEM) method. Moreover, the CPEM method yields a significantly smaller standard error. Based on the estimations of σ^{2*} and π^* presented in Tables 2 and 3, it is clear that the estimation results of the NCPEM method deviate substantially from the true parameter values. Additionally, as the sample size increases, the majority of the estimation results tend to converge more closely to the true parameter values, accompanied by a smaller standard error. These findings strongly indicate that our improved CPEM method exhibits superior accuracy and enhanced robustness.

4.2. Detection Experiment

In experiment 2, we explore the performance of the detection procedure. Let $p = 1$ and consider the following mixture model:

$$Y_i \sim \begin{cases} \sum_{c=1}^{2} \pi_c^{*-} N\left(\mu_c^{*}-, \sigma_c^{2*-}\right), & 1 \le i \le k^*, \\ \sum_{c=1}^{2} \pi_c^{*+} N\left(\mu_c^{*+}, \sigma_c^{2*+}\right), & k^*+1 \le i \le n, \quad k^* \in [1, n], \end{cases} \tag{8}$$

where

$$\mu_1^{*-} = 1,\ \sigma_1^{2*-} = 0.1,\ \pi_1^{*-} = 0.6,\ \mu_2^{*-} = 2,\ \sigma_2^{2*-} = 0.2,\ \pi_2^{*-} = 0.4,$$
$$\mu_1^{*+} = 3,\ \sigma_1^{2*+} = 0.2,\ \pi_1^{*+} = 0.3,\ \mu_2^{*+} = 4,\ \sigma_2^{2*+} = 0.3,\ \pi_2^{*+} = 0.7.$$

The limiting distribution function in (6) is well known and tabulated by Kiefer [23]. In model (8), $q = 5$, at the significance level of 0.05, the critical value is 2.0005. We set the values of n to be 500, 1000, 2000, k to be $n/4$, $n/2$, and $3n/4$ to represent the case where the change point is located in the front, middle, and back end of the data. The empirical sizes and powers are shown in Tables 4 and 5 under 100 simulations.

Table 4. Empirical sizes.

n	500	1000	2000
size	0.07	0.06	0.05

Table 5. Empirical powers.

n \ k^*	$\frac{n}{4}$	$\frac{n}{2}$	$\frac{3n}{4}$
500	0.60	0.83	0.59
1000	0.96	0.88	0.67
2000	1	0.98	0.94

Based on the analysis of the data in the tables, the performance of the method proposed in this paper under large-sample scenarios can be summarized as follows:

(1) Empirical Level Performance: As clearly shown in Table 4, with the continuous increase in the sample size n, the empirical level gradually approaches the significance level of 0.05. This indicates that, under large sample conditions, the actual test level of this method is in good agreement with the preset theoretical level.

(2) Empirical Power Performance: According to the data results in Table 5, there is a significant positive correlation between the increase in sample size and the improvement in empirical power. That is, the larger the sample size, the closer the value of empirical power is to 1, suggesting that the method can more effectively detect true change point situations in large samples.

(3) Influence of Change Point Location: The position of the change point in the data sequence has an impact on the detection effect. When the change point is located in the middle or front part of the data, the detection effect of this method is significantly better than when the change point is located at the end of the data.

(4) Comprehensive Performance Advantages: In large sample scenarios, the method proposed in this paper demonstrates good empirical size and power.

5. Concluding Remarks

This paper proposes an improved EM method to solve the parameter estimation problem with a mixture of the linear regression model and change point. This method is more accurate and robust against usual EM algorithms. Furthermore, a detection method based on score function is obtained to test the change point in mixture data. Simulations demonstrate the effectiveness of the proposed methodology.

6. Limitations and Future Research Directions

The method presented in this article also has certain limitations and challenges. Firstly, there is the issue of robustness in complex noisy scenarios. The method proposed in this article performs well in dealing with normal noise, but the detection accuracy remains to be studied under complex noise interference. The algorithm mentioned in the article still needs further research in dealing with high-dimensional data point detection and estimation problems. Finally, the computational complexity of the method proposed in this article is relatively high, and it is worth studying whether there is a simpler and faster processing method. The above issues are also possible research directions for the future.

Author Contributions: Methodology, Z.X.; Software, W.Z.; Writing—original draft, T.C.; Writing—review & editing, W.Z. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the National Natural Science Foundation of China (Grant No. 12171391) and the Natural Science Basic Research Program of Shaanxi (Grant No. 2024JC-YBQN-0039), Shaanxi Fundamental Science Research Project for Mathematics and Physics (Grant No. 23JSY043).

Data Availability Statement: Data are contained within the article.

Conflicts of Interest: The authors declare no conflicts of interest.

Appendix A. Appendix for Proof

Lemma A1. *Under the Assumption 3 and Cramér-Rao regular conditions, we have*

$$E[q_1(\boldsymbol{\theta}^*, Y)] = 0,$$

$$E[q_2(\boldsymbol{\theta}^*, Y)] = -E[q_1(\boldsymbol{\theta}^*, Y) q_1^T(\boldsymbol{\theta}^*, Y)].$$

Proof. See McLachlan and Peel [24]. □

Lemma A2. *For the process given by*

$$\Delta_n(t, \boldsymbol{\theta}^*) = \frac{1}{\sqrt{n}} \sum_{i=1}^{[nt]} q_1(\boldsymbol{\theta}^*, Y_i), \quad t \in (0,1). \tag{A1}$$

and under the conditions of Lemma A1 and under H_0 the following functional central limit theorem holds:

$$\Delta_n(\cdot, \boldsymbol{\theta}^*) \xrightarrow{d} Z(\cdot),$$

where $Z(\cdot)$ is the Gaussian process whose mean is $E[Z(t)] = 0$ and variance function is $Var[Z(t)] = \mathcal{B}(\boldsymbol{\theta}^*)$.

Furthermore, if $\mathcal{B}(\boldsymbol{\theta}^*)$ is reversible, then we have

$$\mathcal{B}(\boldsymbol{\theta}^*)^{-1/2} \Delta_n(\cdot, \boldsymbol{\theta}^*) \xrightarrow{d} W(\cdot),$$

where $W(\cdot)$ is standard Brownian motion.

Proof. The proof follows by direct application of Donsker's theorem (Billingsley [25]). □

Usually in applications, the parameters $\boldsymbol{\theta}^*$ under the null hypothesis are not known but have to be estimated. Let $\hat{\boldsymbol{\theta}}_n$ is the maximum likelihood estimator of $\boldsymbol{\theta}^*$; therefore,

$$\sum_{i=1}^{n} q_1(\hat{\boldsymbol{\theta}}_n, Y_i) = 0.$$

Taylor expansion of $\Gamma_n(\boldsymbol{\theta}) = \frac{1}{n} \sum_{i=1}^{n} q_1(\boldsymbol{\theta}, Y_i)$ at $\boldsymbol{\theta}^*$ when $\boldsymbol{\theta} = \hat{\boldsymbol{\theta}}_n$; then,

$$0 = \Gamma_n(\hat{\boldsymbol{\theta}}_n) = \Gamma_n(\boldsymbol{\theta}^*) + \Gamma_n'(\boldsymbol{\theta}^*)(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}^*) + R_n.$$

Under suitable regularity conditions

$$-\Gamma_n'(\boldsymbol{\theta}^*) = \frac{1}{n} \sum_{i=1}^{n} -q_2(\boldsymbol{\theta}^*, Y_i) \xrightarrow{p} \mathcal{I}(\boldsymbol{\theta}^*),$$

$$\sqrt{n}\Gamma_n(\boldsymbol{\theta}^*) \xrightarrow{d} \mathcal{N}(0, \mathcal{B}(\boldsymbol{\theta}^*)),$$

$$\sqrt{n} R_n \xrightarrow{p} 0.$$

Therefore, the following holds:

$$\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}^*) \xrightarrow{d} \mathcal{N}(0, \mathcal{V}(\boldsymbol{\theta}^*)),$$

where $\mathcal{V}(\boldsymbol{\theta}^*) = \mathcal{I}(\boldsymbol{\theta}^*)^{-1}\mathcal{B}(\boldsymbol{\theta}^*)\{\mathcal{I}(\boldsymbol{\theta}^*)^{-1}\}^T$. Equivalently, we can write

$$\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}^*) \doteq \mathcal{I}(\boldsymbol{\theta}^*)^{-1}\Delta_n(1, \boldsymbol{\theta}^*).$$

Lemma A3. *Under H_0, the following holds:*

$$\Delta_n(\cdot, \hat{\boldsymbol{\theta}}_n) \xrightarrow{d} Z^0(\cdot),$$

where $Z^0(t) = Z(t) - tZ(1)$.

Proof. Taylor expansion of $\Delta_n(t, \hat{\boldsymbol{\theta}}_n)$ at $\boldsymbol{\theta}^*$,

$$\begin{aligned}\Delta_n(t, \hat{\boldsymbol{\theta}}_n) &\doteq \frac{1}{\sqrt{n}}\sum_{i=1}^{[nt]} q_1(\boldsymbol{\theta}^*, Y_i) + \frac{1}{n}\sum_{i=1}^{[nt]} q_2(\boldsymbol{\theta}^*, Y_i)\cdot\sqrt{n}(\hat{\boldsymbol{\theta}}_n - \boldsymbol{\theta}^*)\\ &\doteq \Delta_n(t, \boldsymbol{\theta}^*) - \frac{[nt]}{n}\mathcal{I}(\boldsymbol{\theta}^*)\cdot\mathcal{I}(\boldsymbol{\theta}^*)^{-1}\Delta_n(1, \boldsymbol{\theta}^*)\\ &\xrightarrow{d} Z(t) - tZ(1).\end{aligned}$$

□

Next, we will provide the proof process for Theorem 1.

Proof. (1) On the basis of the three lemmas stated above and if $\mathcal{B}(\boldsymbol{\theta}^*)$ is reversible, we can obtain the first conclusion in Theorem 1:

$$\hat{\mathcal{B}}_n^{-1/2}\Delta_n(\cdot, \hat{\boldsymbol{\theta}}_n) \xrightarrow{d} W^0(\cdot),$$

where $W^0(\tau) = W(\tau) - \tau W(1)$ is a standard Brownian bridge, $\hat{\mathcal{B}}_n$ is the variance estimation, such as $\hat{\mathcal{B}}_n = \frac{1}{n}\sum_{i=1}^{n} q_1(\hat{\boldsymbol{\theta}}_n, Y_i)q_1(\hat{\boldsymbol{\theta}}_n, Y_i)^T$.

(2) Next, we prove the second conclusion. Y_i has the probability density function under H_A hypothesis:

$$\begin{aligned}\eta(Y|\boldsymbol{\theta}_i) &\doteq \eta(Y|\boldsymbol{\theta}^*) + \frac{\partial\eta(Y|\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}\\ &= \eta(Y|\boldsymbol{\theta}^*) + \frac{\partial\log\eta(Y|\boldsymbol{\theta})}{\partial\boldsymbol{\theta}}\cdot\eta(Y|\boldsymbol{\theta})|_{\boldsymbol{\theta}=\boldsymbol{\theta}_0}\cdot\frac{1}{\sqrt{n}}\cdot I_{\{i\geq[n\tau]+1\}}\\ &= \eta(Y|\boldsymbol{\theta}^*)\{1 + q_1(\boldsymbol{\theta}^*, Y)^T\cdot\frac{1}{\sqrt{n}}\cdot I_{\{i\geq[n\tau]+1\}}\},\end{aligned}$$

which can be derived from the Taylor expansion of η. Therefore, under the alternative hypothesis, the process (A1) no longer has zero mean in general but

$$\begin{aligned}E[q_1(\boldsymbol{\theta}_i, Y)] &\doteq \int q_1(\boldsymbol{\theta}^*\eta(y|\boldsymbol{\theta}^*)dy + \int q_1(\boldsymbol{\theta}^*, y)q_1(\boldsymbol{\theta}^*, y)^T\eta(y|\boldsymbol{\theta}^*)\frac{1}{\sqrt{n}}\cdot I_{\{i\geq[nt]+1\}}dy\\ &= 0 + \frac{1}{\sqrt{n}}\mathcal{B}(\boldsymbol{\theta}^*)\cdot I_{\{i\geq[nt]+1\}}.\end{aligned}$$

Similar to the Lemmas A2 and A3, we can deduce

$$\Delta_n(\cdot, \boldsymbol{\theta}^*) \xrightarrow{d} W^A(\cdot),$$

where $W^A(t) = W(t) + \mathcal{B}(\boldsymbol{\theta}^*)G(t)$, $(G(t) = \int_0^t \cdot I_{\{\frac{i}{n} \geq t + \frac{1}{n}\}} d(\frac{i}{n}) = t(1-t) + o(1))$.

Finally, the following limiting process can be derived, which is our second conclusion:

$$\begin{aligned}\hat{\mathcal{B}}_n^{-1/2}\Delta_n(\tau, \hat{\boldsymbol{\theta}}_n) &\doteq \hat{\mathcal{B}}_n^{-1/2}\{\Delta_n(\tau, \boldsymbol{\theta}^*) - \tau\Delta_n(1, \boldsymbol{\theta}^*)\} \\ &\doteq W^0(\tau) + \hat{\mathcal{B}}(\hat{\boldsymbol{\theta}}_n)^{1/2}G^0(\tau) \\ &\xrightarrow{d} W^0(t) + \mathcal{B}(\boldsymbol{\theta}^*)^{1/2}\boldsymbol{G}^0(t).\end{aligned}$$

□

Next, we prove Theorem 2 and first give two useful lemmas.

Lemma A4. *Under the conditions of model (8) and Assumptions 1 and 2, $\mathcal{L}_n$ uniformly converges to $\mathcal{L}$ with probability.*

Proof. According to the structure of the function $\mathcal{L}$, we can show that $\mathcal{L}_n$ converges to $\mathcal{L}$ point by point. Next, we prove uniform convergence.

Due to the fact that the parameter space Θ is a compact set, $\forall i, \delta_i > 0, \Theta \subseteq \bigcup_{i=1}^{m} U((\tau_i, \boldsymbol{\theta}_i), \delta_i), m \in R$, where $U((\tau_i, \boldsymbol{\theta}_i), \delta_i)$ is the neighborhood with center $(\tau_i, \boldsymbol{\theta}_i)$ and radius δ_i. According to the continuity of $\mathcal{L}_n$, $\forall \epsilon > 0$,

$$|(\tau_i, \boldsymbol{\theta}_i) - (\tau, \boldsymbol{\theta})| < \delta_i \Rightarrow |\mathcal{L}_n(\tau_i, \boldsymbol{\theta}_i) - \mathcal{L}_n(\tau, \boldsymbol{\theta})| < \epsilon, \forall (\tau, \boldsymbol{\theta}) \in U((\tau_i, \boldsymbol{\theta}_i), \delta_i).$$

Consequently,

$$\begin{aligned}|\mathcal{L}_n(\tau, \boldsymbol{\theta}) - \mathcal{L}(\tau, \boldsymbol{\theta})| &= |\mathcal{L}_n(\tau, \boldsymbol{\theta}) - \mathcal{L}_n(\tau_i, \boldsymbol{\theta}_i) + \mathcal{L}_n(\tau_i, \boldsymbol{\theta}_i) - \mathcal{L}(\tau_i, \boldsymbol{\theta}_i) + \mathcal{L}(\tau_i, \boldsymbol{\theta}_i) - \mathcal{L}(\tau, \boldsymbol{\theta})| \\ &\leq |\mathcal{L}_n(\tau, \boldsymbol{\theta}) - \mathcal{L}_n(\tau_i, \boldsymbol{\theta}_i)| + |\mathcal{L}_n(\tau_i, \boldsymbol{\theta}_i) - \mathcal{L}(\tau_i, \boldsymbol{\theta}_i)| + |\mathcal{L}(\tau_i, \boldsymbol{\theta}_i) - \mathcal{L}(\tau, \boldsymbol{\theta})| \\ &< 3\epsilon.\end{aligned}$$

□

Lemma A5. *Under the condition of model (8) and Assumptions 1 and 3, the parameters are identifiable.*

Proof. From the *Jessen* inequality,

$$\begin{aligned}&\mathcal{L}(\tau, \boldsymbol{\theta}|\tau^*, \boldsymbol{\theta}^*) - \mathcal{L}(\tau^*, \boldsymbol{\theta}^*|\tau^*, \boldsymbol{\theta}^*) \\ &= \tau \cdot E_{(\tau^*, \boldsymbol{\theta}^{*-})} f(X, Y|\tau, \boldsymbol{\theta}^-) + (\tau^* - \tau) \cdot E_{(\tau^*, \boldsymbol{\theta}^{*-})} f(X, Y|\tau, \boldsymbol{\theta}^+) \\ &\quad + (1 - \tau^*) \cdot E_{(\tau^*, \boldsymbol{\theta}^{*+})} f(X, Y|\tau, \boldsymbol{\theta}^+) - (\tau \cdot E_{(\tau^*, \boldsymbol{\theta}^{*-})} f(X, Y|\tau^*, \boldsymbol{\theta}^{*-}) \\ &\quad + (\tau^* - \tau) \cdot E_{(\tau^*, \boldsymbol{\theta}^{*-})} f(X, Y|\tau^*, \boldsymbol{\theta}^{*-}) + (1 - \tau^*) \cdot E_{(\tau^*, \boldsymbol{\theta}^{*+})} f(X, Y|\tau^*, \boldsymbol{\theta}^{*+})) \\ &= \tau \cdot [E_{(\tau^*, \boldsymbol{\theta}^{*-})} f(X, Y|\tau, \boldsymbol{\theta}^-) - E_{(\tau^*, \boldsymbol{\theta}^{*-})} f(X, Y|\tau^*, \boldsymbol{\theta}^{*-})] + (\tau^* - \tau)[E_{(\tau^*, \boldsymbol{\theta}^{*-})} f(X, Y|\tau, \boldsymbol{\theta}^+) \\ &\quad - E_{(\tau^*, \boldsymbol{\theta}^{*-})} f(X, Y|\tau^*, \boldsymbol{\theta}^{*-})] + (1 - \tau^*) \cdot [E_{(\tau^*, \boldsymbol{\theta}^{*+})} f(X, Y|\tau, \boldsymbol{\theta}^+) - E_{(\tau^*, \boldsymbol{\theta}^{*+})} f(X, Y|\tau^*, \boldsymbol{\theta}^{*+})] \\ &= \tau \cdot E_{(\tau^*, \boldsymbol{\theta}^{*-})}[\log g_1(X, Y|\boldsymbol{\theta}^-, \boldsymbol{\theta}^{*-})] + (\tau^* - \tau) \cdot E_{(\tau^*, \boldsymbol{\theta}^{*-})}[\log g_2(X, Y|\boldsymbol{\theta}^+, \boldsymbol{\theta}^{*-})] \\ &\quad + (1 - \tau^*) \cdot E_{(\tau^*, \boldsymbol{\theta}^{*+})}[\log g_3(X, Y|\boldsymbol{\theta}^+, \boldsymbol{\theta}^{*+})] \\ &\leq \tau \cdot \log\{E_{(\tau^*, \boldsymbol{\theta}^{*-})}[g_1(X, Y|\boldsymbol{\theta}^-, \boldsymbol{\theta}^{*-})]\} + (\tau^* - \tau) \cdot \log\{E_{(\tau^*, \boldsymbol{\theta}^{*-})}[g_2(X, Y|\boldsymbol{\theta}^+, \boldsymbol{\theta}^{*-})]\} \\ &\quad + (1 - \tau^*) \cdot \log\{E_{(\tau^*, \boldsymbol{\theta}^{*+})}[g_3(X, Y|\boldsymbol{\theta}^+, \boldsymbol{\theta}^{*+})]\} \\ &= 0,\end{aligned}$$

where

$$g_1(X,Y|\boldsymbol{\theta}^-,\boldsymbol{\theta}^{*-}) = \frac{\sum_{c=1}^{C} \frac{\pi_c}{\sqrt{2\pi\sigma_c^{2-}}} \exp\{-\frac{(Y-X^T\beta_c^-)^2}{2\sigma_c^{2-}}\}}{\sum_{c=1}^{C} \frac{\pi_c^*}{\sqrt{2\pi\sigma_c^{2*-}}} \exp\{-\frac{(Y-X^T\beta_c^{*-})^2}{2\sigma_c^{2*-}}\}},$$

$$g_2(X,Y|\boldsymbol{\theta}^+,\boldsymbol{\theta}^{*-}) = \frac{\sum_{c=1}^{C} \frac{\pi_c}{\sqrt{2\pi\sigma_c^{2+}}} \exp\{-\frac{(Y-X^T\beta_c^+)^2}{2\sigma_c^{2+}}\}}{\sum_{c=1}^{C} \frac{\pi_c^*}{\sqrt{2\pi\sigma_c^{2*-}}} \exp\{-\frac{(Y-X^T\beta_c^{*-})^2}{2\sigma_c^{2*-}}\}},$$

$$g_3(X,Y|\boldsymbol{\theta}^+,\boldsymbol{\theta}^{*+}) = \frac{\sum_{c=1}^{C} \frac{\pi_c}{\sqrt{2\pi\sigma_c^{2+}}} \exp\{-\frac{(Y-X^T\beta_c^+)^2}{2\sigma_c^{2+}}\}}{\sum_{c=1}^{C} \frac{\pi_c^*}{\sqrt{2\pi\sigma_c^{2*+}}} \exp\{-\frac{(Y-X^T\beta_c^{*+})^2}{2\sigma_c^{2*+}}\}}.$$

Therefore, we have $\mathcal{L}(\tau,\boldsymbol{\theta}|\tau^*,\boldsymbol{\theta}^*) \leq \mathcal{L}(\tau^*,\boldsymbol{\theta}^*|\tau^*,\boldsymbol{\theta}^*)$, and the equal sign holds only when $(\tau,\boldsymbol{\theta}) = (\tau^*,\boldsymbol{\theta}^*)$; that is, $\mathcal{L}$ maximizes at $(\tau^*,\boldsymbol{\theta}^*)$. Next, we prove that the maximum point is unique.

According to *Jessen* inequality, the three terms on the left side of the inequality are all less than or equal to 0. Therefore, the left side is equal to 0 if and only if the three terms are equal to 0. Since $\tau,\tau^* \in (0,1)$, both $g_1(X,Y|\boldsymbol{\theta}^-,\boldsymbol{\theta}^{*-})$ and $g_3(X,Y|\boldsymbol{\theta}^+,\boldsymbol{\theta}^{*+})$ are equal to 1 if and only if $(\pi_c,\beta_c^-,\sigma_c^{2-}) = (\pi_c^*,\beta_c^{*-},\sigma_c^{2*-})$ and $(\pi_c,\beta_c^+,\sigma_c^{2+}) = (\pi_c^*,\beta_c^{*+},\sigma_c^{2*+})$; meanwhile, because of $\beta_c^+ \neq \beta_c^{*-},\sigma_c^{2+} \neq \sigma_c^{2*-}$, $g_2(X,Y|\boldsymbol{\theta}^+,\boldsymbol{\theta}^{*-})$ is equal to 1 if and only if $\tau = \tau^*$. Therefore, $\mathcal{L}(\tau,\boldsymbol{\theta}|\tau^*,\boldsymbol{\theta}^*)$ obtains the unique maximum at $(\tau^*,\boldsymbol{\theta}^*)$. □

The proof of Theorem 2 is as follows.

Proof. For any $\epsilon > 0$, let

$$\begin{aligned} A &= \{\mathcal{L}_n(\tau^*,\boldsymbol{\theta}^*) > \mathcal{L}(\tau^*,\boldsymbol{\theta}^*) - \epsilon/3\},\\ B &= \{\mathcal{L}_n(\hat{\tau},\hat{\boldsymbol{\theta}}_n) > \mathcal{L}_n(\tau^*,\boldsymbol{\theta}^*) - \epsilon/3\},\\ C &= \{\mathcal{L}(\hat{\tau},\hat{\boldsymbol{\theta}}_n) > \mathcal{L}_n(\hat{\tau},\hat{\boldsymbol{\theta}}_n) - \epsilon/3\}. \end{aligned}$$

From Lemmas A4 and A5, $\mathcal{L}_n$ uniformly converges to $\mathcal{L}$; furthermore, $(\hat{\tau},\hat{\boldsymbol{\theta}}_n)$ and $(\tau^*,\boldsymbol{\theta}^*)$ are the maximum points of $\mathcal{L}_n$ and $\mathcal{L}$, respectively; therefore, when $n \to \infty$, $P(A), P(B), P(C)$ all tend to 1. So when the event $A \cap B \cap C$ occurs,

$$\begin{aligned} \mathcal{L}(\tau^*,\boldsymbol{\theta}^*) - \mathcal{L}(\hat{\tau},\hat{\boldsymbol{\theta}}_n) &= (\mathcal{L}(\tau^*,\boldsymbol{\theta}^*) - \mathcal{L}_n(\tau^*,\boldsymbol{\theta}^*)) + (\mathcal{L}_n(\tau^*,\boldsymbol{\theta}^*) - \mathcal{L}_n(\hat{\tau},\hat{\boldsymbol{\theta}}_n))\\ &\quad + (\mathcal{L}_n(\hat{\tau},\hat{\boldsymbol{\theta}}_n) - \mathcal{L}(\hat{\tau},\hat{\boldsymbol{\theta}}_n))\\ &< \frac{\epsilon}{3} + \frac{\epsilon}{3} + \frac{\epsilon}{3} = \epsilon. \end{aligned}$$

Furthermore, the parameter space Θ is a compact set and $\mathcal{L}$ is continuous, so for a sufficiently small δ, let $N = U((\tau^*,\boldsymbol{\theta}^*),\delta)$, we have

$$\sup_{(\tau,\boldsymbol{\theta}) \in \Theta \cap N^c} \mathcal{L}(\tau,\boldsymbol{\theta}) \triangleq \mathcal{L}(\tau_0,\boldsymbol{\theta}_0) < \mathcal{L}(\tau^*,\boldsymbol{\theta}^*).$$

We might as well take $\epsilon = \mathcal{L}(\tau^*,\boldsymbol{\theta}^*) - \mathcal{L}(\tau_0,\boldsymbol{\theta}_0) > 0$, then $-\epsilon < \mathcal{L}(\tau^*,\boldsymbol{\theta}^*) - \mathcal{L}(\hat{\tau},\hat{\boldsymbol{\theta}}_n) < \epsilon$ holds with probability 1.

Therefore, the following holds:

$$P(|\mathcal{L}(\tau^*,\boldsymbol{\theta}^*) - \mathcal{L}(\hat{\tau},\hat{\boldsymbol{\theta}}_n)| < \epsilon) \to 1.$$

In other words, when $(\hat{\tau}, \hat{\boldsymbol{\theta}}_n) \in N^c$ and $|(\tau^*, \boldsymbol{\theta}^*) - (\hat{\tau}, \hat{\boldsymbol{\theta}}_n)| \geq \delta$, then $|\mathcal{L}(\tau^*, \boldsymbol{\theta}^*) - \mathcal{L}(\hat{\tau}, \hat{\boldsymbol{\theta}}_n)| \geq \epsilon$. It can be shown that

$$P(|(\tau^*, \boldsymbol{\theta}^*) - (\hat{\tau}, \hat{\boldsymbol{\theta}}_n)| \geq \delta) \leq P(|\mathcal{L}(\tau^*, \boldsymbol{\theta}^*) - \mathcal{L}(\hat{\tau}, \hat{\boldsymbol{\theta}}_n)| \geq \epsilon) \to 0.$$

i.e., $P(|(\tau^*, \boldsymbol{\theta}^*) - (\hat{\tau}, \hat{\boldsymbol{\theta}}_n)| \geq \delta) \to 0$. Therefore, we obtain the desired result:

$$(\hat{\tau}, \hat{\boldsymbol{\theta}}_n) \xrightarrow{p} (\tau^*, \boldsymbol{\theta}^*).$$

□

References

1. Goldfeld, S.M.; Quandt, R.E. A Markov model for switching regressions. *J. Econom.* **1973**, *1*, 3–15. [CrossRef]
2. Frühwirth-Schnatter, S. *Finite Mixture and Markov Switching Models*; Springer: New York, NY, USA, 2006; pp. 241–275.
3. Hurn, M.; Justel, A.; Robert, C.P. Estimating Mixtures of Regressions. *J. Comput. Graph. Stat.* **2003**, *12*, 55–79. [CrossRef]
4. Li, P.; Chen, J. Testing the Order of a Finite Mixture. *J. Am. Stat. Assoc.* **2010**, *105*, 1084–1092. [CrossRef]
5. Chen, J.; Li, P.; Fu, Y. Inference on the Order of a Normal Mixture. *J. Am. Stat. Assoc.* **2012**, *107*, 1096–1105. [CrossRef]
6. Huang, M.; Yao, W. Mixture of Regression Models With Varying Mixing Proportions: A Semiparametric Approach. *J. Am. Stat. Assoc.* **2012**, *107*, 711–724. [CrossRef]
7. Huang, M.; Li, R.; Wang, S. Nonparametric Mixture of Regression Models. *J. Am. Stat. Assoc.* **2013**, *108*, 929–941. [CrossRef]
8. Page, E.S. Continuous inspection schemes. *Biometrika* **1954**, *41*, 100–115. [CrossRef]
9. Sen, A.; Srivastava, M.S. On tests for detecting change in mean. *Ann. Stat.* **1975**, *3*, 98–108. [CrossRef]
10. Sen, A.; Srivastava, M.S. Some one-sided tests for change in level. *Technometrics* **1975**, *17*, 61–64. [CrossRef]
11. Hawkins, D.M. Testing a sequence of observations for a shift in location. *J. Am. Stat. Assoc.* **1977**, *72*, 180–186. [CrossRef]
12. James, B.; James, K.; Siegmund, D. Tests for a change-point. *Biometrika* **1987**, *74*, 71–83. [CrossRef]
13. Srivastava, M.S.; Worsley, K.J. Likelihood ratio tests for a change in the multivariate normal mean. *J. Am. Stat. Assoc.* **1986**, *81*, 199–204. [CrossRef]
14. Bassevile, M.; Nikiforov, I. *Detection of Abrupt Changes: Theory and Applications*; Prentice Hall: Hoboken, NJ, USA, 1993.
15. Wang, G.; Zou, C.; Yin, G. Change-point detection in multinomial data with a large number of categories. *Ann. Stat.* **2018**, 46, 2020–2044. [CrossRef]
16. Xia, Z.; Qiu, P. Jump information criterion for statistical inference in estimating discontinuous curves. *Biometrika* **2015**, *102*, 397–408. [CrossRef]
17. Bai, J. Least squares estimation of a shift in linear processes. *J. Time Ser. Anal.* **1994**, *15*, 453–472. [CrossRef]
18. Baranowski, R.; Chen, Y.; Fryzlewicz, P. Narrowest-over-threshold detection of multiple change points and change-point-like features. *J. R. Stat. Soc. Ser. Stat. Methodol.* **2019**, *81*, 649–672. [CrossRef]
19. Follain, B.; Wang, T.; Samworth, R.J. High-dimensional changepoint estimation with heterogeneous missingness. *J. R. Stat. Soc. Ser. Stat. Methodol.* **2022**, *84*, 1023–1055. [CrossRef]
20. Drabech, Z.; Douimi, M.; Zemmouri, E. A Markov random field model for change points detection. *J. Comput. Sci.* **2024**, *83*, 102429. [CrossRef]
21. Ratnasingam, S.; Gamage, R. Empirical likelihood change point detection in quantile regression models. *Comput. Stat.* **2025**, *40*, 999–1020. [CrossRef]
22. Gong, Y.; Dong, X.; Zhang, J.; Chen, M. Latent evolution model for change point detection in time-varying networks. *Inf. Sci.* **2023**, *646*, 119376. [CrossRef]
23. Kiefer, J. K-sample analogues of the Kolmogorov-Smirnov and Cramér-Von Mises tests. *Ann. Math. Stat.* **1959**, *30*, 420–447. [CrossRef]
24. McLachlan, G.; Peel, D. *Finite Mixture Models*; Wiley: New York, NY, USA, 2000.
25. Billingsley, P. *Convergence of Probability Measures*, 2nd ed.; Wiley: New York, NY, USA, 1999.

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.