Quest Journals
Journal of Research in Applied Mathematics
Volume 11 ~ Issue 4 (April 2025) pp: 111-125
ISSN (Online): 2394-0743 ISSN (Print): 2394-0735
www.questjournals.org

Review Paper

A Segmented Autoregressive Model Based on the Horizontal and Vertical Axes

Ding Peng[1],Mingduan Liang[2], Ying Han[3],
[1,2,3]*School of Mathematical Sciences, Guizhou Normal University, Guiyang, China*
Corresponding Author: Ding Peng

ABSTRACT: *This paper aims to further discuss the segmentation and estimation issues of autoregressive (AR) models. The innovations in the AR model do not follow the default normal distribution but rather an arbitrary distribution or a mixture of several distributions. We use a mixture of normal distributions to fit the distribution of the innovations, and employ the Dirichlet process as the prior for the variance of the mixture normal distribution to obtain the posterior estimates of the parameters. By combining Gibbs sampling, we continuously update the parameter estimates of the model to achieve more accurate results.In this study, segmentation is performed along both the horizontal and vertical axes. For the horizontal axis, we use the Bayesian method to identify the threshold values and lag parameters. For the vertical axis, we employ the LASSO method to identify the locations of change points. Numerical simulations are conducted to demonstrate the feasibility of the two segmentation approaches. Moreover, the model estimation and analysis are carried out on real time series data.*

KEYWORDS: *Dirichlet Process Mixture Model; Threshold Autoregressive Model; Gibbs Sampling; LASSO*

R*eceived 20 Apr., 2025;* ***R****evised 28 Apr., 2025;* ***A****ccepted 30 Apr., 2025*
P*ublished with open access at* ***www.questjournas.org***

I. INTRODUCTION

The research in this paper focuses on the discussion of two different segmentation methods, namely the segmented autoregressive (AR) model based on the vertical axis and the segmented AR model based on the horizontal axis. The segmented AR model based on the vertical axis is the threshold autoregressive model (TAR). Since its introduction by Tong [7] in 1978, the TAR model has been extensively studied and is considered an approximation of nonlinear autoregressive models. Essentially, the TAR model is a piecewise linear model in the state space, where it is linear within each threshold value space. For more detailed descriptions of this model, refer to Tong [8].To estimate the parameters of the threshold autoregressive model, many scholars have conducted research. Tsay [9] used the least squares method to estimate the parameters and developed a simple statistic to specify the threshold values. In the Bayesian estimation approach, Chen and Lee [10] performed Bayesian estimation for the two-segment threshold model and used Gibbs sampling to obtain the expected marginal posterior densities of the threshold values and other parameters. This method avoids complex analysis and numerical multiple integrations. Moreover, Chen [11] constructed a Bayesian framework for the generalized threshold autoregressive model, demonstrating that the MCMC algorithm can be successfully applied to parameter estimation. For the self-exciting threshold autoregressive (SETAR) model with a single threshold variable, a considerable number of papers have proposed different information criteria and evaluated their performance [12][13][14][15]. Wang [16] studied the parameter estimation of panel data threshold models using the inverse LR method, focusing on small sample cases and providing corresponding confidence intervals. Zhang [17] compared the Bayesian estimation and maximum likelihood estimation of the threshold autoregressive model. The results showed that the regression parameters have the same distribution as the maximum likelihood estimates, while the Bayesian estimates converge to a function of a compound Poisson process, which can be regarded as the average of the compound Poisson process in the time domain. For inference on the multiple threshold autoregressive model, refer to the literature [18]-[23].

The study of segmented autoregressive models based on the vertical axis is essentially the study of change-point models. As research has progressed, it has become evident that linear time series models have certain limitations, especially when fitting stock data. Stock closing prices sometimes experience sudden increases or decreases. Page [24] first formally introduced the change-point problem in 1954, focusing on whether the

distribution parameters change once, i.e., whether there is a single change point in the time series data, and proposed the cumulative sum (CUSUM) detection method.Quandt [25] used the likelihood ratio test (LRT) to construct a test statistic for studying simple linear regression models with a single change point. For traditional methods on change-point research, refer to the monographs and review papers [26]-[34].After Tibshirani [35] proposed the LASSO method by adding a penalty term based on the L1 norm, many scholars applied the LASSO method to change-point detection, transforming the change-point detection problem into a variable selection problem. Zou [36] introduced a new sparse principal component analysis method—sparse principal component analysis, which obtains sparse loadings by imposing LASSO constraints on the regression coefficients. The analysis of the LASSO algorithm has yielded satisfactory results. Ciuperca [37] considered linear regression models with multiple change points and proposed model selection criteria for two cases. Comparisons showed that for sufficiently large sample sizes, the adaptive LASSO method selects the true model better than ordinary least squares (OLS), regardless of the location of the change points. Ciuperca [38] applied the adaptive LASSO method to detect change points in quantile regression models, obtaining the convergence rates of change-point estimates and regression parameter estimates in each stage. The sparsity of the adaptive LASSO quantile estimates of regression parameters is not affected by change-point estimation, and it was verified that the adaptive LASSO outperforms other variable selection methods at that time. For other types of LASSO methods for identifying change-point locations, refer to the literature [39]-[41].

The structure of this paper is as follows: Section 2 presents the matrix forms under the two segmentation methods and explains how the segmentation is performed; Section 3 derives the likelihood function of the time series data \(y \) and introduces the Dirichlet Process Mixture Model (DPMM) into the autoregressive model to compute the posteriors of the parameters; Section 4 conducts numerical simulations to demonstrate the feasibility of the proposed methods. For the threshold model, the proposed method is compared with the Ordinary Least Squares (OLS) method, and for the change-point model, it is compared with the pure LASSO and Maximum Likelihood Estimation (MLE) methods. The experimental results are analyzed, and it is concluded that the proposed methods outperform the other methods; Section 5 applies the two segmentation methods to real stock data, presenting the estimation results and forecasts for the next five steps; Section 6 concludes the paper by detailing the research findings.

II. IDENTIFICATION OF SEGMENTED AUTOREGRESSIVE

2.1 Identification of Vertical-Axis Segmented Autoregressive Models

Consider the Self-Exciting Threshold Autoregressive (SETAR) model:

$$y_t = \phi_{j0} + \sum_{i=1}^{q_j} \phi_{ji} y_{t-1} + \varepsilon_{jt}, \ \ r_{j-1} < y_{t-d} \leq \ r_j \tag{1}$$

$$j = 1,2,\cdots,m \qquad -\infty = r_0 < r_1 < \cdots < \ r_{m-1} < \ r_m = +\infty$$

where $\phi_{ji}$and r_j are real numbers, q_j, m, and d are positive integers. It is assumed that the innovation terms ε_t are independent and follow different mixed normal distributions, i.e., $\varepsilon_{jt} \sim N\left(0, \sigma_{je}^2\right), e = 1,2,\cdots,\mathrm{k}_j$, where e is a positive integer.

This paper divides the sample observations$\{y_t, t = 1,2,\cdots,N\}$ into multiple subsets (where N is the total sample size), denoted as $Y_1, Y_2, \cdots, Y_m$, with $Y_1 = \{y_t, y_{t-d} \leq r_1\}$, $Y_2 = \{y_t, r_1 < y_{t-d} \leq r_2\}$, and $Y_m = \{y_t, r_{m-1} < y_{t-d} \leq r_m\}$. The dimension of Y_j is $N_j \times (q_j + 1)$, and the sample sizes of $Y_1, Y_2, \cdots, Y_m$ are $N_1, N_2, \cdots, N_m$ respectively, with $N = \ N_1 + N_2 + \cdots + N_m$.Let $\phi_j = (\phi_{j0}, \phi_{j1}, \phi_{j2}, \cdots, \phi_{jq_j})'$. Each segment Y_j is further divided into smaller segments based on the different distributions of the variance, $Y_j = \{Y_{j,N_{je}}, e = 1,2,\cdots,k_j\}$, with each smaller segment $Y_{j,N_{je}}$ having a sample size of N_{je}, and $N_j = \sum_{e=1}^{k_j} N_{je}$, for $j = 1,2,\cdots m$. The equation (1) can be rewritten in matrix form as:

$$Y_{j,N_{je}} = Y_{j,N_{je}}^{*} \phi_j + E_{je} \ , \qquad r_{m-1} \leq y_{t-d} \leq r_j, \varepsilon_t \sim N\left(0, \sigma_{je}^2\right) \tag{2}$$

The specific expansion yields the following equation:

$$Y=\begin{cases}\begin{cases}Y_{1,N_{11}}=Y^*_{1,N_{11}}\phi_1+\varepsilon_{11} & , \quad y_{t-d}\le r_1,\varepsilon_t\sim N(0,\sigma^2_{11})\\ \vdots & \\ Y_{1,N_{1k_1}}=Y^*_{1,N_{1k_1}}\phi_1+\varepsilon_{1k_1} & , \quad y_{t-d}\le r_1,\varepsilon_t\sim N\left(0,\sigma^2_{1k_1}\right)\end{cases}\\ \vdots\\ \begin{cases}Y_{m,N_{m1}}=Y^*_{m,N_{m1}}\phi_m+\varepsilon_{m1} & , r_{m-1}<y_{t-d},\varepsilon_t\sim N(0,\sigma^2_{m1})\\ \vdots & \\ Y_{m,N_{mk_m}}=Y^*_{m,N_{mk_1}}\phi_m+\varepsilon_{mk_m}, r_{m-1}<y_{t-d},\varepsilon_t\sim N\left(0,\sigma^2_{mk_m}\right)\end{cases}\end{cases} \tag{3}$$

where N_{je} is the sample size of the e-th small segment in the j-th segment, $E_{je}=\{\varepsilon_{je}\}, r_0=-\infty$, $\Sigma^2=\{\sigma^2_{11},\cdots,\sigma^2_{1k_1},\cdots,\sigma^2_{m1},\cdots,\sigma^2_{mk_m}\}$, and its dimension is $\sum_{j=1}^{m}k_j\times 1$, Let $\Phi=\{\phi_1,\phi_2,\cdots,\phi_m\}'$, the sample observations be denoted as $Y=\{Y^*_1,\cdots,Y^*_m,Y_1,\cdots,Y_m\}$. The likelihood function for Y is given by:

$$P(Y|d,r,q,\Phi,\Sigma^2)\propto\prod_{j=1}^{m}\prod_{e=1}^{k_j}(\sigma^2_{je})^{-\frac{N_{je}}{2}}exp\left(-\frac{\left(Y_{j,N_{je}}-Y^*_{j,N_{je}}\phi_j\right)'\left(Y_{j,N_{je}}-Y^*_{j,N_{je}}\phi_j\right)}{2\sigma^2_{je}}\right) \tag{4}$$

Esmail Amiri [37] provided the joint posterior density of d,r and q conditional on the data:

$$f(d,r,q|\boldsymbol{Y})\propto\prod_{j=1}^{m}2^{\left(-\frac{v_j}{2}+1\right)}\pi^{\frac{v_j}{2}}\Gamma(\frac{v_j}{2})(\frac{v_js_j^2}{2})^{-\frac{v_j}{2}}|Y_i'Y_i|^{-\frac{1}{2}} \tag{5}$$

where $q=\{q_1,\cdots,q_m\}$, $r=\{r_1,\cdots,r_{m-1}\}$, $d=0,1,2,\cdots,I$, and I is a positive integer. $v_j=\{v_{je},e=1,2,\cdots,k_j\}$ and $s_j^2=\{s_{je}^2\}$ are given by: $s_{je}^2=\frac{\left(Y^*_{j,N_{je}}-Y_{j,N_{je}}\phi^*_j\right)'\left(Y^*_{j,N_{je}}-Y_{j,N_{je}}\phi^*_j\right)}{v_{je}}, v_{je}=N_{je}-q_j-1$.The MCMC method is used to estimate the values of d,r and q. The closed forms of the conditional posterior probability functions for each order or the delay given the other parameters are available. The conditional posterior for r is as follows:

$$f(r|\boldsymbol{Y},d,q)\propto\prod_{j=1}^{m}2^{\left(-\frac{v_j}{2}+1\right)}\pi^{\frac{v_j}{2}}\Gamma(\frac{v_j}{2})(\frac{v_js_j^2}{2})^{-\frac{v_j}{2}}\left|Y_j'Y_j\right|^{-\frac{1}{2}} \tag{6}$$

The conditional posterior for d is given by:

$$f(d|\boldsymbol{Y},r,q)\propto\frac{\prod_{j=1}^{m}\Gamma(\frac{v_j}{2})(\frac{v_js_j^2}{2})^{-\frac{v_j}{2}}|Y_i'Y_i|^{-\frac{1}{2}}}{\sum_{d=0}^{I}\prod_{j=1}^{m}\Gamma(\frac{v_j}{2})(\frac{v_js_j^2}{2})^{-\frac{v_j}{2}}\left|Y_i'Y_i\right|^{-\frac{1}{2}}} \tag{7}$$

where I represents the maximum delay parameter, which is a positive integer. The conditional posterior for q is given by:

$$f(q|\boldsymbol{Y},r,d)\propto\frac{\prod_{j=1}^{m}\Gamma(\frac{v_j}{2})(\frac{v_js_j^2}{2})^{-\frac{v_j}{2}}|Y_i'Y_i|^{-\frac{1}{2}}}{\sum_{q_1=0}^{n_{q_1}}\cdots\sum_{q_m=0}^{n_{q_m}}\prod_{j=1}^{m}\Gamma(\frac{v_j}{2})(\frac{v_js_j^2}{2})^{-\frac{v_j}{2}}\left|Y_i'Y_i\right|^{-\frac{1}{2}}} \tag{8}$$

where $n_{q_1},\cdots n_{q_m}$ are the maximum orders of the autoregression for each segment, which are less than the actual number of time series data points in each segment.

2.2 Estimation of Horizontal-Axis Segmented Autoregressive Models

The matrix form of the horizontal-axis segmented autoregressive model is:

$$y_t=\begin{cases}\begin{cases}Y_{1,N_{11}}=Y^*_{1,N_{11}}\phi_1+\varepsilon_{11} & ,t<ct_1,\varepsilon_t\sim N(0,\sigma^2_{11})\\ \vdots & \\ Y_{1,N_{1k_1}}=Y^*_{1,N_{1k_1}}\phi_1+\varepsilon_{1k_1} & ,t<ct_1,\varepsilon_t\sim N\left(0,\sigma^2_{1k_1}\right)\end{cases}\\ \vdots\\ \begin{cases}Y_{m,N_{m1}}=Y^*_{m,N_{m1}}\phi_m+\varepsilon_{m1} & ,ct_{m-1}\le t,\varepsilon_t\sim N(0,\sigma^2_{m1})\\ \vdots & \\ Y_{m,N_{mk_m}}=Y^*_{m,N_{mk_1}}\phi_m+\varepsilon_{mk_m},ct_{m-1}\le t,\varepsilon_t\sim N\left(0,\sigma^2_{m,k_m}\right)\end{cases}\end{cases} \tag{9}$$

where $\{ct_1, \cdots, ct_{m-1}\}$ is the set of change point locations, and the number of segments in the model is equal to the number of change point locations plus one. The data for each segment are $Y_1 = \{y_t, t < ct_1\}, \cdots, Y_m = \{y_t, ct_{m-1} \leq t\}$, and the other symbols are consistent with those in section 2.1.

III. PARAMETER ESTIMATION

The conditional likelihood function for Y is:

$$P(Y|ct_1, \cdots, ct_{m-1}, \Phi, \Sigma^2) \propto \prod_{j=1}^{m} \prod_{e=1}^{k_j} \left(\sigma_{je}^2\right)^{-\frac{N_{je}}{2}} exp\left(-\frac{\left(Y_{j,N_{je}} - Y_{j,N_{je}}^* \phi_j\right)' \left(Y_{j,N_{je}} - Y_{j,N_{je}}^* \phi_j\right)}{2\sigma_{je}^2}\right) \quad (10)$$

It should be noted that, whether it is the segmented autoregressive model based on the horizontal axis or the vertical axis as mentioned earlier, the estimation methods for the parameters of the segmented autoregressive models are the same. Therefore, the following posterior estimation methods for these parameters are shared.

This paper introduces the Dirichlet Process Mixture Model (DPMM) into the autoregressive model. Since the posterior distribution of the Dirichlet Process Mixture Model does not have an analytical solution, the unknown model is partitioned into blocks, and the joint posterior distribution of the parameters is obtained using Bayes' theorem:

$$\pi(\Sigma^2, \Phi, \alpha, Z|Y) = f(Y|\Sigma^2, \Phi, \alpha, Z)\pi(\Phi|\Sigma^2, \alpha, Z)\pi(\Sigma^2|\alpha, Z)\pi(Z|\alpha)\pi(\alpha) \quad (11)$$

where $Z = \{z_{y_t}, t = 1,2, \cdots, N\}$ is the set of categorical labels. Given the other parameters and latent variables, the Markov chain is constructed by iteratively sampling from the posterior distribution of each block.

It should be noted that the parameter estimation in this section is based on the following premise: the autoregressive model has already been segmented as described earlier. For the threshold model based on the vertical axis, the parameters d, r and q are known through Equations (6)–(8). For the change-point model based on the horizontal axis, the number and locations of the change points have already been identified through LASSO.

We now begin to estimate ϕ_i. When calculating the posterior distribution of ϕ_i, to facilitate the derivation using conjugate distributions, this paper chooses the prior distribution of ϕ_i to be a multivariate normal distribution:

$$p(\phi_j|d, r) \sim N_{q_j+1}(u_j, \sqrt{v_j}) \quad (12)$$

Combining Equation (4), the conditional posterior distribution of ϕ_j is a multivariate normal distribution:

$$\phi_j|d, r, q \sim N\left(\phi_j^*, \sigma_{\phi_j}^2\right) \quad (13)$$

where, $\phi_j^* = \sigma_{\phi_j}^2 \left(\frac{(Y_{j,N_{je}})' Y_{j,N_{je}}^*}{\sigma_{je}^2} + \frac{u_j}{v_j}\right)$, $\sigma_{\phi_j}^2 = \left(\frac{(Y_{j,N_{je}}^*)' Y_{j,N_{je}}^*}{\sigma_{je}^2} + v_j^{-1}\right)^{-1}$.

Assuming V_t is the variance carried by the t-th data point, we introduce an indicator variable $Z = (z_{y_1}, z_{y_2}, \cdots, z_{y_N})$ such that when $V_t = \sigma_{je}^2, e = 1,2, \cdots, k_j$, then $z_{y_t} = e$. Under the Dirichlet process prior, the distribution of z_{y_t} is as follows:

$$z_{y_t} \sim \sum_{e=1}^{\infty} \pi_e \delta_{\sigma_{je}^2}(dV) \quad (14)$$

where δ is an indicator function that is 1 only when $V_t = \sigma_{je}^2$ and 0 otherwise. The weight π is defined through the stick-breaking process as $\pi_i = p_i \prod_{ii=1}^{i-1}(1 - p_{ii})$, where $p|\alpha \sim Beta(1, \alpha)$. σ_{je}^2 is the variance of the e-th normal distribution in the j-th segment of the segmented autoregressive model, and it is one of the distinct values in V. By introducing z_{y_t} into the DPMM model, we obtain:

$$\varepsilon_t \sim N\left(0, \sigma_{z_{y_t}}^2\right)$$

$$z_{y_t} \sim \sum_{e=1}^{\infty} \pi_e \delta_e$$

$$\sigma_{z_{y_t}}^2 \sim G_0$$

$$G_0 \sim Inv - gamma(v_1, c_1) \quad (15)$$

where $\sigma^2_{z_{y_t}}$ represents the variance of the normal distribution that the t-th data point follows. There is a premise that the innovations are independent. Define $\boldsymbol{\varepsilon} = (\varepsilon_{11}, \cdots, \varepsilon_{1N_1}, \cdots, \varepsilon_{m1}, \cdots, \varepsilon_{mN_m})'$, and $\varepsilon_{ji'} = y_{ji'} - \phi_{j0} - \sum_{i=1}^{q_j} \phi_{ji} y_{j,i'-i}$, for $i' = q_j + 1, q_j + 2, \cdots, N_j$. The conditional posterior distribution of V_t is obtained as:

$$V_t|\{V_{t'}, t' \neq t\}, \varepsilon_t, \alpha \sim \frac{\alpha}{n+\alpha-1} g(\varepsilon_t) G(dV|\varepsilon_t) + \frac{1}{n+\alpha-1} \sum_{t' \neq t} f_N(\varepsilon_t|V_{t'}) \delta_{V_{t'}}(dV) \tag{16}$$

From the form of the inverse gamma distribution's density function, we can deduce:

$$G(dV|\varepsilon_t) \sim Inv-gamma\left(v_1 + \frac{1}{2}, \frac{2c_1 + \varepsilon_t^2}{2}\right) \tag{17}$$

The marginal likelihood function is:

$$g(\varepsilon_t) = \int f_N(\varepsilon_t|0, V) G_0(V) dV = \frac{1}{\sqrt{2\pi}} \times \frac{(v_1)^{c_1}}{\Gamma(v_1)} \times \frac{\Gamma\left(v_1 + \frac{1}{2}\right)}{\left(\frac{2c_1 + \varepsilon_t^2}{2}\right)^{v_1 + \frac{1}{2}}} \tag{18}$$

It is important to note that $G(dV|\varepsilon_t)$ is the core of the inverse gamma distribution, while $g(\varepsilon_t)$ is the marginal likelihood function obtained by integrating out the inverse gamma distribution. To sample $\Sigma^2$from the posterior distribution $\pi(\Sigma^2|\alpha, Z)$, we need to sample Z from the posterior distribution $\pi(Z|\alpha)$. Here, $\Sigma^2 = \{\sigma^2_{11}, \cdots, \sigma^2_{1k_1}, \cdots, \sigma^2_{m1}, \cdots, \sigma^2_{mk_m}\}$ is the set composed of distinct V_t,for $t = 1,2, \cdots N$, and $Z = \{z_{y_1}, z_{y_2}, \cdots, z_{y_N}\}$ is the vector composed of the assignment variables z_{y_t}. When $V_t = \sigma^2_{je}$, $z_{y_t} = e$. Therefore, sampling Σ^2 requires two steps:

1. Sample the cluster labels z_{y_t} first:

$$z_{y_t}|\sigma^2_{je}, \varepsilon_t, \alpha \sim \begin{cases} \dfrac{\alpha g(\varepsilon_t) \delta_{k_j+1}(dz_{y_t})}{\sum_{e=1}^{k_j} f_N(\varepsilon_t|\sigma^2_{je}) \delta_e(dz_{y_t}) + \alpha g(\varepsilon_t) \delta_{k_j+1}(dz_{y_t})} \\ \dfrac{\sum_{e=1}^{k_j} f_N(\varepsilon_t|\sigma^2_{je}) \delta_e(dz_{y_t})}{\sum_{e=1}^{k_j} f_N(\varepsilon_t|\sigma^2_{je}) \delta_e(dz_{y_t}) + \alpha g(\varepsilon_t) \delta_{k_j+1}(dz_{y_t})} \end{cases} \tag{19}$$

The above is obtained after normalization. The cluster label z_{y_t} will be sampled from $\{1,2, \cdots, k_j, k_j + 1\}$ with the sampling probability given by Equation (19). If the sampled value is from $\{1,2, \cdots, k_j\}$, then the variance of that category will be updated. If the sampled value is $k_j + 1$, then k_j is updated to$k_j + 1$, increasing the number of mixture normal distributions, and the variance of the new distribution will be randomly sampled from the inverse gamma distribution.

2. After knowing the distribution of z_{y_t} for the *j*-th segment, start updating the variance in each class:

$$\pi(\sigma^2_{je}|Z, \varepsilon, k_j) \propto (\sigma^2_{je})^{-\left[\frac{N_{je}+2v_1}{2}+1\right]} exp\left(-\frac{N_{je}\varepsilon_t^2 + 2c_1}{2\sigma^2_{je}}\right) \sim Invgamma\left(\frac{N_{je} + 2v_1}{2}, \frac{N_{je}\varepsilon_t^2 + 2c_1}{2}\right) \tag{20}$$

where $j = 1,2, \cdots, m$, e $= 1,2, \cdots,$ k$_j$ and $\varepsilon = \{\varepsilon_1, \varepsilon_2, \cdots, \varepsilon_{N_j}\}$. If the cluster label z_{y_t} sampled in the first step is a new category, then $N_{je} = 1$. Therefore, the variance of the normal distribution that the data in each segment follows can be estimated using Equation (20).

The posterior estimation of the parameter α is given by Escobar, who provided a method based on Gibbs sampling. The posterior estimation of α is a mixture of two gamma distributions:

$$\alpha|k, \eta \sim \pi_\eta G(c + k, d - ln\eta) + (1 - \pi_\eta) G(c + k - 1, d - ln\eta) \tag{21}$$

In Equation (21), η and π_η are still unknown. π_η is defined by the following equation:$\frac{\pi_\eta}{1-\pi_\eta} = \frac{c+k-1}{n(d-ln\eta)}$,which can be rearranged to:

$$\pi_\eta = \frac{c + k - 1}{n(d - ln\eta) + c + k - 1} \tag{22}$$

Next, we calculate the posterior distribution of η, $P(\eta|k, \alpha) \propto \eta^\alpha (1 - \eta)^{n-1}$, which is:

$$\eta|k, \alpha \sim Beta(\alpha + 1, n) \tag{23}$$

IV. NUMERICAL SIMULATIONS

4.1 Simulation Based on the Vertical-Axis Segmented Autoregressive Model

In this section, this paper illustrates the simulation study of the aforementioned model and compares it with conventional methods to demonstrate the superiority of the proposed model. The following TAR model is considered, which has one threshold value, and each segment is a first-order autoregressive model with the variance of the innovation term in each segment coming from a mixture of two variances:

$$Y_t = \begin{cases} 7.8 + Y_1^* 02 + \varepsilon_1 & , y_{t-1} \le 17, \varepsilon_1 \sim \pi_{11} N(0,10) + \pi_{12} N(0,50) \\ 14 + 0.2 Y_2^* + \varepsilon_2 & , 17 < y_{t-1}, \varepsilon_2 \sim \pi_{21} N(0,10) + \pi_{22} N(0,1) \end{cases} \quad (24)$$

where $(\pi_{11}, \pi_{12}, \pi_{21}, \pi_{22})$ = (0.410, 0.590, 0.435, 0.565).

This paper considers that the variances of the innovation terms in each segment are independent and come from a mixture distribution, estimated using the DPMM method. The estimation results are compared with the case where the innovations are assumed to follow a single normal distribution, with the benchmark method being OLS. The estimated threshold values are shown in the following table:

Table 4-1: Estimation of Threshold Values, Lag Parameters, and Autoregressive Model Orders.

Parameter	r	d	q_1	q_2
True Value	17	1	1	1
DPMM	17.009	1	1	1
OLS	17.845	1	1	1

From Table 4-1, it can be seen that the threshold values estimated by the DPMM method are slightly better, which may be related to the estimation of the coefficients and variances.

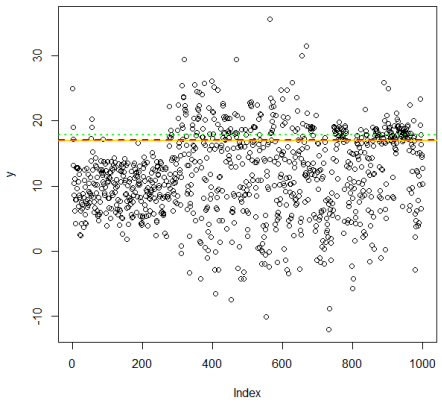

Figure 4-1: The positional relationship between the true values and the threshold values estimated by the two methods.

In Figure 4-1, the gold line represents the true value of the threshold r, while the red and green lines above it represent the threshold values estimated by the DPMM and OLS methods, respectively. The red line is almost completely overlapping with the gold line.

Table 4-2: Estimation of the model's constant term and coefficients by DPMM and OLS.

Parameter	ϕ_{10}	ϕ_{11}	ϕ_{20}	ϕ_{21}
True Value	7.800	0.200	14.000	0.200
DPMM	7.826	0.205	13.901	0.206
OLS	7.574	0.292	12.700	0.262

In Table 4-2, the second row represents the true values of the parameters, while the third row shows the average estimates obtained using the DPMM method. ϕ_{10} and ϕ_{11} denote the constant term and the coefficient of the first-order autoregressive term in the first segment, respectively. ϕ_{20} and ϕ_{21} denote the constant term and the coefficient of the first-order autoregressive term in the second segment, respectively. As can be seen from Table 4-2, the estimates obtained using the DPMM method are closer to the true values of the parameters, indicating that the proposed method is more effective.

The constant terms and coefficients of each segment of the threshold model estimated by DPMM are plotted for the last 500 iterations as follows:

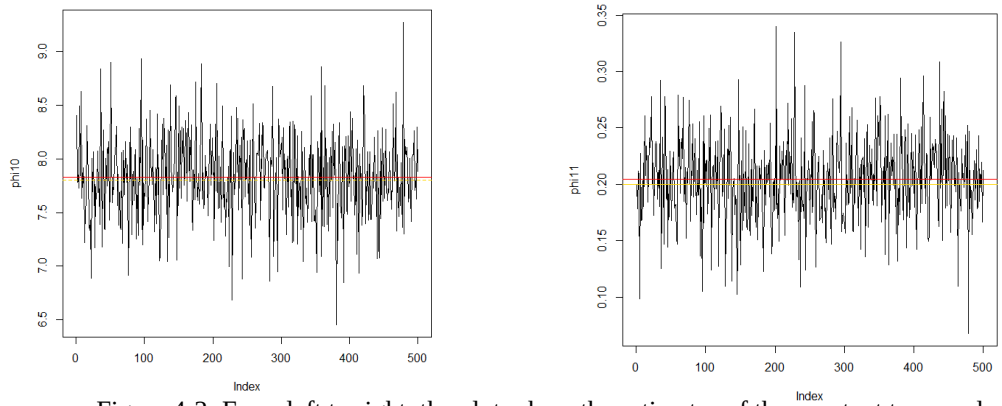

Figure 4-2: From left to right, the plots show the estimates of the constant term and coefficient for the first segment of the autoregressive model.

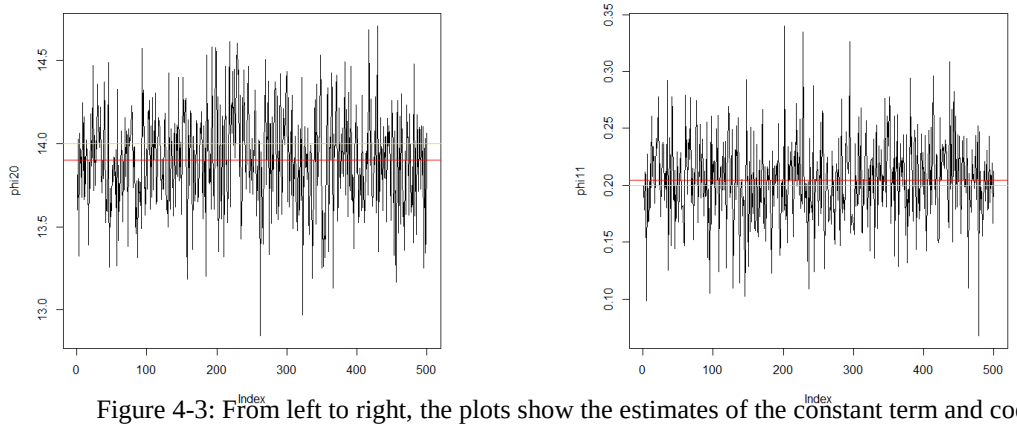

Figure 4-3: From left to right, the plots show the estimates of the constant term and coefficient for the second segment of the autoregressive model.

Figures 4-2 and 4-3 display the estimates of the constant terms and coefficients for the first and second segments of the autoregressive model, respectively. The gold lines represent the true values, while the red lines indicate the mean estimates obtained using the DPMM method. In Figure 4-2, the left plot shows that the true value of the constant term for the first segment is 7.8, and the mean estimate from the DPMM method is 7.826. The right plot indicates that the true value of the coefficient for the first segment is 0.20, with a mean estimate of 0.206 from the DPMM method. In Figure 4-3, the left plot shows that the true value of the constant term for the second segment is 14, and the mean estimate from the DPMM method is 13.901. The right plot indicates that the true value of the coefficient for the second segment is 0.20, with a mean estimate of 0.206 from the DPMM method.

Table 4-3: This table presents the estimation of the variances of the innovation's mixture normal distributions using the Dirichlet Process Method.

Parameter	σ_{11}^2	σ_{12}^2	σ_{21}^2	σ_{22}^2
True Value	10	50	10	1
DPMM	13.682	56.176	10.032	0.636
OLS	34.803	34.803	7.380	7.380

In Table 4-3, the second row represents the true values of the variances, the third row shows the mean estimates of the variances obtained using the DPMM method, and the fourth row shows the mean estimates of the variances obtained using the OLS method. The OLS method assumes that the innovation terms in each segment follow the same distribution, hence it provides identical estimates for σ_{11}^2 and σ_{12}^2, as well as for σ_{21}^2 and σ_{22}^2.For the DPMM method, we choose to consider the two clusters with the highest number of assignments as the final clustering result, discarding the other clusters with very few assignments. Therefore, the DPMM method estimates the innovation terms as a mixture of two normal distributions, which matches the original design of the number of mixture components.In the data setup, the first segment's innovation terms follow a mixture distribution of the

form: $\varepsilon_1 \sim 0.3894166N(0,\sqrt{10}) + 0.6105834N(0,\sqrt{50})$ .The DPMM method estimates the mixture normal distribution as: $\varepsilon_1 \sim 0.4097425N(0,\sqrt{13.682}) + 0.5902575N(0,\sqrt{56.176})$ .For the second segment, the innovation terms follow a mixture distribution of the form:$\varepsilon_2 \sim 0.5648855N(0,\sqrt{10}) + 0.4351145N(0,\sqrt{1})$.The DPMM method estimates the mixture form as:$\varepsilon_2 \sim 0.6224674N(0,\sqrt{10.032}) + 0.3775326\ (0,\sqrt{0.636})$.

Table 4-4: Number of Clusters and Their Frequencies Over 500 Iterations

Number of Clusters.	1	2	3	4	5	6	7	8	9	10
Frequency of the First Segment.	0	191	162	94	31	14	6	2	0	0
Frequency of the Second Segment.	0	316	140	34	9	1	0	0	0	0

It is worth noting that when using the DPMM method to cluster the innovations in this paper, the number of clusters often exceeds two. In the first segment of the autoregressive model, the DPMM identifies up to 8 clusters, with weights for one instance being (0.356368564 , 0.616310160 , 0.002710027 , 0.004065041 , 0.005420054 , 0.001355014 , 0.001355014 , 0.001355014). In fact, apart from the first two categories having a larger number of samples, the sample sizes for the third category and beyond are very small. The sample sizes for each category are (263 , 463 , 2 , 3 , 4 , 1 , 1 , 1). Therefore, this paper ignores these categories and assigns them to the second category, considering that the innovations in the first segment follow a mixture of two normal distributions. Similarly, in the second segment of the autoregressive model, the DPMM identifies up to 6 clusters, which only appear once. The weights for this instance are (0.609195402 , 0.360153257 , 0.015325670 , 0.007662835 , 0.003831418 , 0.003831418). The sample sizes for each category are (159 , 94 , 4 , 2 , 1 , 1). Thus, the innovations in the second segment are also considered to follow a mixture of two normal distributions.

The parameter α does not have a true value; its magnitude merely restricts the number of clusters. Generally, under the same conditions, the smaller the value of α, the fewer the number of clusters, and vice versa. In this paper, the average value of the clustering parameter for the first segment is $\alpha_1 = 0.379508$, and for the second segment, it is $\alpha_2 = 0.1256694$.

4.2 Simulation Based on the Horizontal-Axis Segmented Autoregressive Model

In this subsection, the DPMM method used in this paper is compared with the LASSO regression method and the MLE method. First, we consider a two-segment autoregressive model with the specific parameters as follows:

$$y = \begin{cases} 0.6 \times y_{t-1} + 0.2y_{t-2} + \varepsilon_1, n \leq 400 \\ 0.8 \times y_{t-1} - 0.1y_{t-2} + \varepsilon_2, n > 400 \end{cases} \tag{25}$$

where $\varepsilon_1 \sim t(df = 10)$ with a default mean of 0 and a default standard deviation of 1, and $\varepsilon_2 \sim 0.5N(0,0.5) + 0.5N(0,2)$. The model has a sample size of $n = 1000$, and the true location of the change point is at the 401st time point.

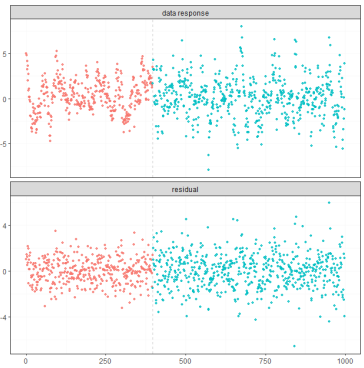

Figure 4-4: Change point location identified by the LASSO method.

Figure 5-6 illustrates the identification of the change point by the LASSO method, which divides the time series model into two segments, with the change point located at the 396th time point. The vertical line in the figure indicates the position of this change point, which is very close to the set change point location. In Figure

5-6, the horizontal axis represents the time series of data points, and the vertical axis represents the data values. The green part indicates good estimation results. First, the LASSO method is used to identify the change point location. Subsequently, the DPMM, LASSO, and MLE methods are applied to estimate the parameters for each segment of the data, yielding the results shown in the following table:

Table 4-5: Estimation of Parameters for the Autoregressive Model

Parameter	b_{11}	b_{12}	b_{21}	b_{22}
True Value	0.6	0.2	0.8	-0.1
DPMM	0.5973	0.2378	0.8015	-0.1005
LASSO	0.5950	0.2538	0.8096	-0.1116
MLE	0.6044	0.2416	0.8057	-0.1048

Table 4-5 presents the estimates of the coefficients for the two-segment autoregressive model obtained using the three methods. The initial coefficients were sampled from a standard normal distribution. From this table, it can be seen that the DPMM method provides the most accurate estimates of the autoregressive coefficients for both segments, especially for the second segment, while the MLE method is the next best, and the LASSO regression is the weakest among the three methods.The innovations are an important part of the autoregressive model, representing all the new information at a given time that cannot be explained by past sequence values. The innovations follow an unknown distribution, which may be non-normal or a mixture distribution. This paper applies the Dirichlet process to the prior of the innovations, using a mixture of normal distributions to fit the unknown distribution, resulting in Table 4-6:

Table 4-6: Parameter Estimation of the Distribution Followed by the Innovations

Parameter	t-distribution	Mixture of Normal Distributions	
True Value	t(df=10)	N(0,2)	N(0,0.5)
DPMM	N(0, 1.0677)+N(0,2.2338)	N(0, 2.1702)	N(0,0.5789)
LASSO	N(0,1.4248)	N(0,1.6093)	N(0,1.6093)
MLE	N(0, 1.1018)	N(0,1.5287)	N(0, 1.5287)

In Table 4-6, it should be noted that: The innovations in the first segment of the autoregressive model follow a t-distribution with 10 degrees of freedom. The DPMM method uses a mixture of two normal distributions to approximate this t-distribution, with specific standard deviations and weights as follows: 0.6302392 N(0, 1.0677) + 0.3697608 N(0, 2.2338).For the innovations in the second segment of the autoregressive model, which follow a mixture of two normal distributions with equal weights of 0.5 each, the DPMM method correctly identifies the number of mixture components, with specific standard deviations and weights as follows: 0.4972397 N(0, 2.1702) + 0.5027603 N(0, 0.5789).In contrast, the LASSO and MLE methods are unable to identify the number of components in the mixture. When the innovations do not follow a single normal distribution, the estimated variance values are inaccurate.

Similar to Section 5.1, the Dirichlet process often identifies more clusters than the actual number of categories. In the first segment of the autoregressive model, the DPMM identifies up to 14 clusters at its maximum, with weights as follows: (0.63023919, 0.21786768, 0.07667684, 0.03166412, 0.02159288, 0.01172519, 0.00424936, 0.00303308, 0.00107888, 0.00110433, 0.00051908, 0.00015267, 0.00008142, 0.00001527). In fact, apart from the first two categories, the sample sizes of the remaining categories are very small, with only a few or a dozen samples. Therefore, this paper ignores these categories and assigns them to the second category, considering that the innovations in the first segment follow a mixture of two normal distributions. In the second segment of the autoregressive model, the DPMM uses up to 7 mixture normal distributions at its maximum, which only occurs once, with the weights of the mixture normal distributions being (0.462809917, 0.507438017, 0.004958678, 0.018181818, 0.001652893, 0.001652893, 0.003305785). The Dirichlet process often identifies more categories than the actual number, but many of these categories have very low weights. Similarly, it is considered that only two categories are generated here. The LASSO and MLE methods assume that the innovations follow a single normal distribution.

This paper processes the innovations to better explain the time series data and also to predict the time series model. Therefore, this paper conducts predictions for the first and second segments of the autoregressive model, forecasting the data for the next five time points. The results are as follows:

Table 4-7: Forecasted Values for the Next Five Time Points in the First Segment of the Autoregressive Model.

	First Segment of the Autoregressive Model					MSE
True Value	3.839809	1.883247	4.481181	3.410880	3.354103	NULL
DPMM	3.256816	3.977992	4.588088	2.272528	2.389873	1.392970
LASSO	-0.337823	-0.002889	-0.418313	-1.410194	0.151177	15.703330
MLE	1.164346	1.067213	0.993561	0.925571	0.866678	6.470308

In the five-step-ahead time series prediction for the first segment of the autoregressive model, the DPMM method provides an acceptable estimate for the first time point and a very close estimate for the third time point. However, the errors for the other three time points are greater than 1, which can be attributed to the randomness in the estimation process. Despite this, the Mean Squared Error (MSE) value is relatively small. The LASSO regression has the largest MSE value, while the MLE method has a smaller MSE value. A smaller MSE indicates a more accurate estimation.

Table 4-8: Forecasted Values for the Next Five Time Points in the Second Segment of the Autoregressive Model

	The second segment of the autoregressive model.					MSE
True Value	1.817131	1.213866	1.896725	2.018733	2.114893	NULL
DPMM	2.0109773	1.7938544	0.7503575	0.9040297	2.10309257	0.586165
LASSO	3.500133	2.112954	2.979287	5.274592	6.007251	6.112773
MLE	1.3726238	0.9267491	0.6462279	0.4669116	0.3518173	1.472070

In the prediction of the second segment of the autoregressive model, the DPMM method provides an estimate close to the true value for the first time series point and an acceptable estimate for the second time series point. Due to the characteristics of random sampling, the third and fourth time series points deviate from the true values, while the fifth time series point is again close to the true value. The LASSO method has the largest MSE, followed by MLE, with DPMM having the smallest MSE.

Additionally, the clustering effect of the Dirichlet process is influenced by the hyperparameter α. The closer the value of α is to 0, the fewer the number of clusters. The estimated values of α for the two segments of the autoregressive model are $\alpha_1 = 0.495951$ and $\alpha_2 = 0.426070$, respectively. In practice, the calculated values of α often tend to be close to 0, and there is no so-called "true value" for α.

V. EMPIRICAL ANALYSIS

In this section, the two segmentation methods are applied to the same real dataset—the closing prices of Salt Lake Corporation, to better demonstrate the practicality of the DPMM method in real data. The dataset consists of 2,492 time series data points, collected hourly from January 2, 2020, to November 10, 2023. The data were sourced from East Money Finance.

5.1 Empirical Analysis of the Vertical-Axis Segmented Autoregressive Model

Using the original data, an Augmented Dickey-Fuller (ADF) test was conducted. The Dickey-Fuller statistic value was -2.668, with a lag order of 13, indicating that the model considered data from the previous 13 time points to estimate the current point's value. The p-value was 0.2955, which is greater than the commonly used significance levels (such as 0.05 or 0.01), so the time series is non-stationary.

After differencing the data, the threshold value was calculated as -0.17, with a lag parameter of 1, and the order of the two-segment threshold autoregressive model was 2. Conducting the ADF test again, the ADF statistic value for the first segment of the autoregressive model was -10.2526, which is much smaller than the critical value at the 1% significance level. The ADF statistic value for the second segment of the threshold autoregressive model was -22.3589, which is also much smaller than the critical value at the 1% significance level. This indicates that the segmented time series does not have a unit root.

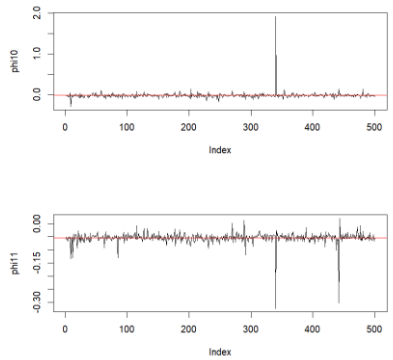

Figure 5-1: This figure shows the estimates for the first segment of the autoregressive model, with the coefficients for the first-order and second-order autoregressive terms displayed from top to bottom.

In this subsection, the number of iterations is set to 1000, with the first 500 iterations discarded as burn-in. The results from the last 500 iterations are used for analysis. The estimates of the constant term and coefficients for the first segment are plotted in Figure 6-2, where the two red lines represent the mean estimates from the last 500 iterations, with values of -0.0218882 and -0.0543517, respectively.

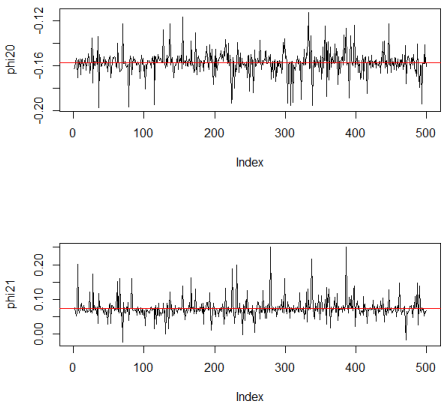

Figure 5-2: This figure displays the estimates for the second segment of the autoregressive model, with the coefficients for the first-order and second-order autoregressive terms shown from top to bottom.

The estimates of the constant term and coefficients for the second segment are plotted in Figure 5-2, where the two red lines represent the mean estimates from the last 500 iterations, with values of -0.1570886 and 0.0736061, respectively. Figures 5-1 and 5-2 show that the estimates fluctuate around the red lines, indicating that the estimation results are stable. This suggests that the chosen number of iterations is appropriate.

Table 5-1: Coefficient Estimates for the Two-Segment Threshold Autoregressive Model.

Parameter	ϕ_{10}	ϕ_{11}	ϕ_{20}	ϕ_{21}
DPMM	-0.022	-0.054	-0.157	0.074

The results of the calculations, rounded to three decimal places, are filled into Table 6-1. The initial values of the coefficients are random numbers drawn from a standard normal distribution.

Table 5-2: Variance Estimates of the Mixture of Normal Distributions for the Innovations in the Two-Segment Threshold Autoregressive Model.

	Mixture of Normal Distributions.	MSE
The first segment of the autoregressive model.	N(0,0.111)+N(4.531)	2.555005
The second segment of the autoregressive model.	N(0, 0.018)	0.1428687

In Table 6-2, for the first segment of the autoregressive model, the maximum number of clusters identified was 12, with weights as follows: (0.848906561, 0.047713718, 0.015904573, 0.039761431, 0.019880716, 0.009940358, 0.001988072, 0.003976143, 0.001988072, 0.001988072, 0.005964215, 0.001988072). If we consider only this single result, only one category would appear. However, taking the average of 500 results, it is ultimately divided into two categories with weights of (0.8679881, 0.1320119). For the second segment of the autoregressive model, the final result has only one category because the weight of the largest category is 0.998, and the other categories can be ignored.

In this section, a model is built for the data of Salt Lake Corporation, resulting in a two-segment threshold model with one threshold value. The clustering parameter for the first segment is $\alpha_1 = 0.5823436$, and for the second segment, it is $\alpha_2 = 0.9355127$. The final estimated threshold model is as follows:

$$Y_t = \begin{cases} -0.022Y_{t-1} - 0.054Y_{t-2} + \varepsilon_1 \ , Y_{t-1} \leq -0.17 \\ -0.157Y_{t-1} + 0.0742Y_{t-2} + \varepsilon_2, -0.17 < Y_{t-1} \end{cases} \tag{26}$$

where $\varepsilon_1 \sim 0.1320119N\left(0, \sqrt{0.111}\right) + 0.8679881N\left(0, \sqrt{4.531}\right)$, and $\varepsilon_2 \sim N\left(0, \sqrt{0.998}\right)$. The model proposed in this paper can be applied to various fields, such as stock markets, bank credit risk, healthcare, and insurance claims.

5.2 Empirical Analysis of the Horizontal-Axis Segmented Autoregressive Model

In this subsection, the conclusions obtained from the numerical simulations are applied. First, LASSO regression is used to identify the change points, and then the DPMM method is employed to estimate the model parameters. The data used in this subsection are the same as the original data used in Section 6.1, but not all data are differenced in this subsection. The specific discussion is as follows.

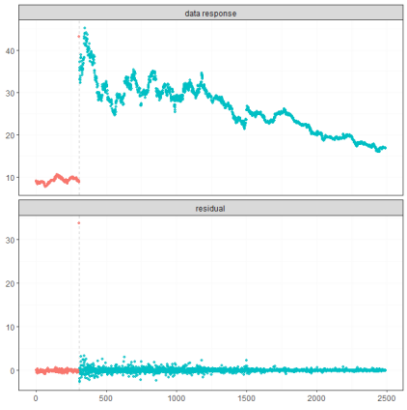

Figure 5-3: Change Point Identification Using LASSO Regression.

Using LASSO regression, Figure 6-4 is obtained, which clearly identifies the change point location at 307. Subsequently, this paper conducts the ADF test on the two segments of time series data separately. The test results are as follows: for the first segment, Dickey-Fuller = -1.4952, p-value = 0.7891; for the second segment, Dickey-Fuller = -3.7367, p-value = 0.02215. Therefore, the p-value of the second segment is less than 0.05, indicating no unit root, while the p-value of the first segment is greater than 0.05, so the null hypothesis is rejected, and it is concluded that the first segment of time series data is not stationary. Thus, this paper performs differencing

on the first segment of data, resulting in Dickey-Fuller = -6.4397, and the p-value is far less than 0.01 at this time, indicating that the differenced time series data is stationary.

Table 5-3: Estimated Coefficients for the Two-Segment Autoregressive Model

	b11	b21	b22
DPMM	0.003613054	0.8513004	0.1473344

In Table 6-3, according to the Akaike Information Criterion (AIC), the optimal orders of the two-segment autoregressive models are determined to be first-order and second-order, respectively. The estimated coefficients for the first segment of the autoregressive model are small because the original data itself has small differences, and the differenced data is even smaller. The time series values in the second segment of the autoregressive model are highly correlated with the first-order lagged values. In fact, in most autoregressive models, time points closer to the t-th moment are more useful for the model.

Table 5-4: Estimates of the Unknown Distributions Followed by the Innovations in the Two-Segment Autoregressive Model

	Mixture of Normal Distributions.	MSE
The first segment of the autoregressive model.	N(0,0.1981)	0.0146692
The second segment of the autoregressive model.	N(0, 0.2327)+ N(0,0.8000)	0.0096657

From Table 5-4, it can be seen that the unknown distribution followed by the innovations in the first segment of the model can be approximated by a single normal distribution, while the unknown distribution followed by the innovations in the second segment requires a mixture of two normal distributions for approximation. When the number of clusters is at its maximum, 10 categories are generated, with weights as follows: (0.5380384968, 0.2873510541, 0.1448212649, 0.0210815765, 0.0018331806, 0.0018331806, 0.0009165903, 0.0009165903, 0.0009165903, 0.0022914757). If we consider this clustering result alone, there should be three normal distributions. However, in the numerous iterations of this experiment, the weight of the third category rarely exceeds 5%. Therefore, this paper takes the mean of the weights from 500 iterations, and the weight of the third category is very small and can be ignored. Ultimately, the weights of the two normal distributions obtained in this paper are: 0.5614537 N(0, 0.2327) + 0.4385463 N(0, 0.8000).

Additionally, since it can be seen from Figure 6-4 that the second segment of the autoregressive model fits well, the estimation of parameters for the second segment should be better than that for the first segment. The clustering parameters generated by the two segments of the autoregressive model are $\alpha_1 = 0.2079229$ and $\alpha_2 = 0.6201116$, respectively.

In summary, the time series model estimated for Salt Lake Corporation is as follows:

$$y = \begin{cases} 0.0036 \times y_{t-1} + \varepsilon_1, n \leq 306 \\ 0.4255 \times y_{t-1} + 0.0738 y_{t-2} + \varepsilon_2, n > 306 \end{cases} \quad (27)$$

where the innovations of the autoregressive model are $\varepsilon_1 \sim \mathrm{N}(0,0.1981)$ and $\varepsilon_2 \sim 0.5614537\mathrm{N}(0, 0.2327) + 0.4385463\mathrm{N}(0,0.8000)$.

VI. CONCLUSION

This paper investigates segmented autoregressive models, discussing parameter estimation for both the vertical-axis and horizontal-axis segmented autoregressive models. For the vertical-axis segmented autoregressive model, the values of d, r and q are estimated first. For the horizontal-axis segmented autoregressive model, the number and locations of change points are identified initially. Subsequently, the Dirichlet Process Mixture Model (DPMM) is introduced into the autoregressive model. For innovations following an unknown distribution, a mixture of normal distributions is used to approximate this distribution, with the variances of the mixture normal distributions sampled from the distribution G generated by the Dirichlet process. The posterior estimates of the parameters of the autoregressive model are derived.

Through numerical simulations, the effectiveness of the proposed method is demonstrated by comparing it with OLS, MLE, and LASSO methods. The Gibbs sampling algorithm is used in this paper to continuously update the model parameters. The initial values of the model's constant, coefficients, and variances are random numbers sampled from distributions. The variances are correlated with the autoregressive model's constant and coefficients. Good variance fitting results enhance the estimation accuracy of other parameters in the autoregressive model. The numerical simulation process shows that the DPMM method significantly outperforms other methods in estimating model parameters. The case study also demonstrates the applicability of the proposed method to real data, proving that DPMM has certain practical value.

When clustering with the DPMM method, the number of categories is generally greater than the actual number of categories. In such cases, it is necessary to examine the data in each category. This paper chooses to discard categories with very few data points in the clustering and assign their data to the last category. If the innovations follow a mixture of normal distributions, the DPMM method can identify the number of mixture normal distributions, with weights close to the actual values, and the estimated variances of the normal distributions are close to the true variances. For innovations following non-normal distributions (such as the t-distribution), the DPMM method can also provide good fitting results using a mixture of normal distributions. Moreover, it does not require pre-specification of the unknown distribution, making the model more flexible and increasing the credibility of the results.

REFERENCES

[1] Tong H. Ona Threshold Model[J]. InPatter n Recognition and Signal Processing (CH Chen, ed.), PP: l01-141. Sijthof and Noordhoi f, Amsterdam, 1978.
[2] Tong H. Non-linear time series: a dynamical system approach[M]. Oxford university press, 1990.
[3] Tsay R S. Testing and modeling threshold autoregressive processes[J]. Journal of the American statistical association, 1989, 84(405): 231-240.
[4] Chen C W S, Lee J C. Bayesian inference of threshold autoregressive models[J]. Journal of time series analysis, 1995, 16(5): 483-492.
[5] Chen C W S. A Bayesian analysis of generalized threshold autoregressive models[J]. Statistics & probability letters, 1998, 40(1): 15-22.
[6] Wong, C. S. and Li, W. K. (1998). A note on the corrected Akaike information criterion for threshold autoregressive models, Journal of Time Series Analysis 19(1): 113–124.
[7] De Gooijer, J. G. (2001). Cross-validation criteria for SETAR model selection, Journal of Time Series Analysis 22(3): 267–281.
[8] Peña, Daniel, and Julio Rodriguez. "Detecting nonlinearity in time series by model selection criteria." International Journal of Forecasting 21.4 (2005): 731-748.
[9] McQuarrie, A., Shumway, R. and Tsai, C.-L. (1997). The model selection criterion AICu, Statistics & Probability Letters 34(3): 285–292.
[10] Wang, Q. (2015). Fixed-Effect Panel Threshold Model using Stata. The Stata Journal: Promoting Communications on Statistics and Stata, 15(1), 121–134.
[11] Zhang R, Liu Q, Shi J. Estimation of generalized threshold autoregressive models[J]. Communications in Statistics-Theory and Methods, 2023, 52(18): 6456-6474.
[12] Pan J, Xia Q, Liu J. Bayesian analysis of multiple thresholds autoregressive model[J]. Computational Statistics, 2017, 32: 219-237.
[13] Yang Y, Ling S, Self-weighted LAD-based inference for heavy-tailed threshold autoregressive models[J]. Journal of Econometrics, 2017, 197(2):368-381.
[14] Li D , Tong H .Nested sub-sample search algorithm for estimation of threshold models[J].London School of Economics and Political Science, LSE Library, 2016
[15] Strikholm, Birgit, and Timo Teräsvirta. "A sequential procedure for determining the number of regimes in a threshold autoregressive model." The Econometrics Journal 9.3 (2006): 472-491.
[16] Gonzalo J, Pitarakis J. Estimation and model selection based inference insingle and multiple threshold models[J]. Journal of Econometrics, 2002,110(2): 319-352.
[17] Hansen B. Testing for linearity[J]. Journal of Economic Surveys, 1999, 13(5):551 -576.
[18] Page, E. S. "Continuous Inspection Schemes." Biometrika 41, no. 1/2 (1954): 100–115
[19] Quandt R E. The estimation of the parameters of a linear regression system obeying two separate regimes[J]. Journal of the American Statistical Association,1958, Vol.53, No.284:873-880.
[20] Chen X R. Introduction of the statistical analysis of change points. Journal of Applied Statistics and Managment, 1991, 10(1): 55–58. (in Chinese)
[21] Wang L M. Research advances in statistical analysis of change point. Statistical Research, 2003, 20(1): 50–51. (in Chinese)
[22] Brodsky E, Darkhovsky B S. Nonparametric Methods in Change Point Problems. Berlin: Springer Science & Business Media, 2013.
[23] Aue A , Horvath L . Structural breaks in time series. Journal of Time Series Analysis, 2013, 34(1): 1–16.
[24] Barry D, Hartigan J A. A Bayesian analysis for change point problems. Journal of the American Statistical Association, 1993, 88(421): 309–319.
[25] Bai J. Likelihood ratio tests for multiple structural changes. Journal of Econometrics, 1999, 91: 299–323.
[26] Zou C L, Yin G S, Feng L, et al. Nonparametric maximum likelihood approach to multiple change-point problems. The Annals of Statistics, 2014, 42(3): 970–1002.
[27] Zhou Y H. Ni Z X. Zhu P F. Study of change points in Shanghai Composite Index based on least absolute deviation criterion. Chinese Journal of Management Science, 2015, 23(10): 38–45. (in Chinese)
[28] Chen Z, Xu Q, Li H. Inference for multiple change points in heavy-tailed time series via rank likelihood ratio scan statistics. Economics Letters, 2019, 179(7): 53–56.
[29] R.Tibshirani. Regression shrinkage and selection via the lasso. Journal of the Royal Statistical Society B, 58:267–288, 1996.

[30] Zou, H., Hastie, T., Tibshirani, R. Sparse principal component analysis[J]. Journal of Computational and Graphical Statistics, 2006, 15(2): 265-286.
[31] Ciuperca G. Model selection by LASSO methods in a change-point model[J]. Statistical Papers,2012, Vol.55, No.2: 349-374.
[32] Ciuperca G. Adaptive Lasso model selection in a multiphase quantile regression[J]. Statistics,2016, Vol.50, No.5:1100-1131.
[33] Zhang B, Jun G and Li-Feng L. Multiple change-points estimation in linear regression models via sparse group Lasso[J]. IEEE Transaction on Signal Processing. 2015, Vol.63, No.9:2209-2224.
[35] Qian J and L. Su. Shrinkage estimation of common breaks in panel models via adaptive group fused lasso[J]. Journal of Econometrics, 2016, Vol.191, No.1:86-109.
[36] Li, Aomei. "Multiple Change-Point Estimation in Linear Regression Models Based on Adaptive Group Lasso Algorithm." [D]. Xiamen: Xiamen University, 2018.
[37] Amiri E. Bayesian Automatic Parameter Estimation of Threshold Autoregressive (TAR) Models using Markov Chain Monte Carlo (MCMC)[C]//Compstat: Proceedings in Computational Statistics. Physica-Verlag HD, 2002: 189-194.