Volume 8, Issue 12, pp. 68-76, 2024.
International Journal of Scientific Engineering and Science
ISSN (Online): 2456-7361

The Application of Dirichlet Process Mixture Models in Threshold Models

Ding Peng[1], Ying Han[2], Tengsong He[3]
School of Mathematical Sciences, Guizhou Normal University, Guiyang550001, China
Email address: 212067042@qq.com

***Abstract*—*Nonlinear models have always been a focal point of study in disciplines such as statistics, finance, and econometrics, with threshold models being a typical example of nonlinear models. This paper applies the Dirichlet process to threshold models to guarantee the flexibility of the approach. The threshold value, lag parameter, and the order of the autoregressive model can be directly estimated from the data. In this paper, the innovation of the threshold model is also considered. Instead of following a zero-mean normal distribution, it follows any distribution. In combination with the MCMC algorithm, and through numerical simulation and comparison with the Ordinary Least Squares method, it is demonstrated that the estimation in this paper is more effective.*

***Keywords*— *Bayesian theory; Dirichlet mixted model; threshold model; MCMC algorithm.*

I. INTRODUCTION

In nonlinear time series, the Threshold Autoregressive (TAR) model is considered an approximation of nonlinear autoregressive models. In fact, it is a piecewise linear model on the state space, being linear in each threshold value's domain. This paper will investigate the Self-Exciting Threshold Autoregressive (SETAR) model within the Threshold Autoregressive (TAR) framework. For further descriptions of this model, references can be made to Tong [9]. In the study of threshold autoregressive models, many scholars have conducted research on the estimation of various parameters. Tsay [12] used the method of least squares to estimate the parameters and developed a simple statistical measure to specify the threshold values. In the Bayesian estimation approach, Chen and Lee [10] conducted Bayesian estimation on the two-regime threshold model and utilized Gibbs sampling to obtain the expected marginal posterior densities for the threshold values and other parameters, thereby avoiding complex analysis and numerical multiple integrations. Furthermore, Chen [11] constructed a Bayesian framework for the generalized threshold autoregressive model, demonstrating that the MCMC algorithm could be successfully applied to parameter estimation. In the study of panel data threshold models. Zhang [14] compared the Bayesian estimation with the maximum likelihood estimation in the threshold autoregressive model. The results indicate that the regression parameters share the same distribution as the Maximum Likelihood Estimation (MLE), while the Bayesian estimation converges to a function of a compound Poisson process, which can be regarded as the time domain mean of the compound Poisson process. Pan [13] considered the multi-threshold autoregressive model, without the need to preset the number of thresholds, and developed a Bayesian random search selection method to identify the number and location of thresholds.

Due to the decisive role of the unknown parameter α in determining the number of clusters, Escobar & West [16] calculated the posterior distribution of parameter α to be a mixture of two gamma distributions, assuming that the prior of parameter α follows a gamma distribution. With continuous refinements by subsequent researchers, the Dirichlet process has found extensive applications in many fields. Liang Hong & Ryan Martin [17] developed a flexible nonparametric Bayesian model for modeling insurance losses to predict future claim amounts. Adesina [18] proposed a Bayesian Dirichlet process mixture prior for Generalized Linear Mixed Models (GLMMs) and applied it to fit both over dispersed and equidispersed count data. Zhang Yongxia [19] established a flexible semiparametric Bayesian hierarchical quantile regression model, employing a Dirichlet process prior for the estimation of the nonparametric part of the model, and conducted an empirical analysis based on actual insurance company data. The Dirichlet process mixture model has also been widely applied in time series analysis.[20][21][22][23].

This paper applies the Dirichlet process to the threshold autoregressive model, with the innovation lying in considering that the autoregressive model is not only subject to a single normal distribution but may follow other distributions or a mixture distribution. Therefore, the parameters of this distribution and the weights of the mixture distribution are calculated, and the posterior distribution of the parameters is derived. This method does not require pre-specification of the form and number of distributions, offering great flexibility.

The structure of this paper is as follows: Section 2 introduces threshold models and the Dirichlet process, Section 3 estimates the parameters of the model, Section 4 uses the set threshold autoregressive model for data simulation, and Section 5 applies the model to estimate real-world data.

II. METHOD INTRODUCTION

2.1: Threshold Autoregressive Model

The Threshold Autoregressive (TAR) model, proposed by Tong [1] in 1978 and systematically outlined by Tong [3] in 1983, is a class of nonlinear time series models. However, due to the complexity of the modeling steps, these models were not easily applied to real-world problems until Ruey S. Tsay [5] (1989) introduced relatively simpler modeling and testing methods, which led to their widespread application. The Self-

International Journal of Scientific Engineering and Science
Volume 8, Issue 12, pp. 68-76, 2024. ISSN (Online): 2456-7361

Exciting Threshold Autoregressive (SETAR) model, which is strictly speaking, uses a piecewise linear model to describe nonlinear models, is a part of this family. The SETAR model is characterized by:

$$y_t = \begin{cases} \phi_{10} + \phi_{11} y_{t-1} + \cdots + \phi_{1p} y_{t-p} + \varepsilon_{1t}, & y_{t-d} \le r_1 \\ \phi_{20} + \phi_{21} y_{t-1} + \cdots + \phi_{2p} y_{t-p} + 2\varepsilon_t, & r_1 < y_{t-d} \le r_2 \\ \vdots & \\ \phi_{m0} + \phi_{m1} y_{t-1} + \cdots + \phi_{mp} y_{t-p} + \varepsilon_{mt}, & r_{m-1} < y_{t-d} \end{cases} \quad (1)$$

Where $\phi_{ji}, j=1,2,\cdots,m; i=1,2,\cdots,p$ is the real number, y_{t-d} is the threshold variable, d is the lag parameter, d,m,p are all positive integers, $r_j, j=1,2,\cdots,m-1$ is the threshold value, ε_{jt} is an assumed independent and identically distributed sequence, i.e., $\varepsilon_t \sim N(0,\sigma_\varepsilon^2)$. The above threshold model can also be represented by a single formula:

$$y_t = \phi_{j0} + \sum_{i=1}^{q_j} \phi_{ji} y_{t-1} + \varepsilon_t, \quad r_{j-1} < y_{t-d} \le \quad r_j, \qquad (2)$$
$$j=1,2,\cdots,m, \quad -\infty = r_0 < r_1 < \cdots < r_{m-1} < r_m = +\infty.$$

2.2: Dirichlet Process

2.2.1: Concept Introduction

The Dirichlet Process (DP) is a stochastic process defined in measure theory as follows: α is a positive real number, $(\Omega, \mathcal{B})$ is a measurable space, and G_0 is a probability distribution on the measurable space $(\Omega, \mathcal{B})$. For any finite partition $A_1, A_2, \cdots, A_n$ of the measurable space $(\Omega, \mathcal{B})$, if the discrete random measure G on the measurable space $(\Omega, \mathcal{B})$ satisfies the following condition:

$$(G(A_1), G(A_2), \cdots, G(A_n))$$
$$\sim DIR(\alpha G_0(A_1), \alpha G_0(A_2), \cdots, \alpha G_0(A_n)) \qquad (3)$$

Then G is said to follow a Dirichlet Process, denoted as $G \sim DP(\alpha, G_0)$, where α is called the precision parameter, and G_0 is referred to as the base distribution. The parameter α controls the degree of discreteness of the distribution G. The distribution G is typically a discrete distribution, which can be defined by the following formula:

$$G(\theta^j) = \sum_{i=1}^{\infty} \pi_i \delta_{\theta_i}(\theta^j) \qquad (4)$$

In the formula, δ_{θ_i} is the indicator function, which takes the value of 1 if and only if $\theta_i = \theta^j$, and 0 otherwise. θ^j corresponds to the parameter of the region, and π_i represents the weight, which can be obtained from the following formula:

$$p_i | \alpha \sim Beta(1, \alpha), \quad \pi_i = p_i \prod_{ii=1}^{i-1} (1 - p_{ii}) \qquad (5)$$

$\{p_i\}_{i=1}^{\infty}$ is a sequence of independent and identically distributed random variables, and the π_i produced by the aforementioned process can be denoted as $\boldsymbol{\pi} = (\pi_i)_{i=1}^{\infty} \sim GEM(\alpha)$, representing a random probability measure (GEM stands for Griffiths, Engen, and McCloskey).

Blackwell (1973) derived an important formula, which subsequent Dirichlet process mixture models introduced into threshold models have been based on for the derivation of the posterior distribution of parameters:

$$p\left(y_{n+1} = j \,\middle|\, \overline{z_{-n}}, \alpha\right) = \sum_{i=1}^{n} \frac{1}{n+\alpha} \delta(y_i, j) + \frac{\alpha}{n+\alpha} G \qquad (6)$$

2.2.1: Dirichlet Process Mixture Model

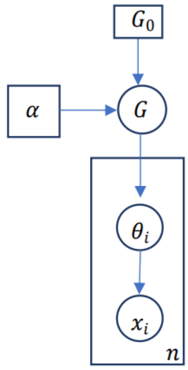

Fig. 2.2.1. The measure model of DPMM.

In the Dirichlet mixture model, the notation n indicates that the operations within the box are repeated n times. Figure 2-1 illustrates the Dirichlet mixture model as follows:

$$G | \alpha_0, G_0 \sim DP(\alpha_0, G_0)$$
$$\boldsymbol{\theta} | G \sim G$$
$$x_i | \theta_i \sim F(\theta_i) \qquad (7)$$

Where $\boldsymbol{\theta} = (\theta_1, \theta_2, \cdots, \theta_n)$. One of the most important applications of the Dirichlet process is as the prior for the parameters in a Dirichlet Process Mixture Model (DPMM) [25].

III. Bayesian Parameter Estimation.

Consider the SETAR model of formula (2):

$$y_t = \phi_{j0} + \sum_{i=1}^{q_j} \phi_{ji} y_{t-1} + \varepsilon_{jt}, \quad r_{j-1} < y_{t-d} \le \quad r_j$$
$$j=1,2,\cdots,m, \; -\infty = r_0 < r_1 < \cdots < \quad r_{m-1} < \quad r_m = +\infty \qquad (8)$$

In which, ϕ_{ji} and r_j are real numbers, q_j, m and d are positive integers. It is assumed that the innovation term ε_t is independent and comes from different mixtures of normal distributions, i.e., $\varepsilon_{jt} \sim N(0, \sigma_{je}^2), e=1,2,\cdots,\mathrm{k}_j$, where e is a positive integer.

This paper divides the observed values of the sample $\{y_t, t=1,2,\cdots,N\}$ into multiple subsets (where N is the

International Journal of Scientific Engineering and Science
Volume 8, Issue 12, pp. 68-76, 2024. ISSN (Online): 2456-7361

total sample size), denoted as $Y_1, Y_2, \cdots, Y_m$, with $Y_1 = \{y_t, y_{t-d} \le r_1\}, Y_2 = \{y_t, r_1 < y_{t-d} \le r_2\}$ and $Y_m = \{y_t, r_{m-1} < y_{t-d} \le r_m\}$. The sample sizes of $Y_1, Y_2, \cdots, Y_m$ are $N_1, N_2, \cdots, N_m$ respectively, with $= N_1 + N_2 + \cdots + N_m$, and $\phi_j = (\phi_{j0}, \phi_{j1}, \phi_{j2}, \cdots, \phi_{jq_j})'$ Each segment Y_j is further divided into smaller segments according to the different distributions that the variances follow, denoted as $Y_j^* = \{Y_{j,N_{je}}^*\}$ with the sample size of each small segment $Y_{j,N_{je}}^*$ being 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:

$$Y_{j,N_{je}}^* = Y_{j,N_{je}}\phi_j + E_{je},\ r_{m-1} < y_{t-d} \le r_j, \varepsilon_t \sim N(0, \sigma_{je}^2) \quad (9)$$

Among them, the dimension of Y_j is $N_j \times (q_j + 1)$, $E_{je} = \{\varepsilon_{je}\}, r_0 = -\infty$, N_{je} is the sample size of the model for the e-th small segment of the j-th segment, and $\mathbf{\Sigma}^2 = \{\varepsilon_{11}, \cdots, \varepsilon_{1k_1}, \cdots, \varepsilon_{m1}, \cdots \varepsilon_{mk_2}\}$, its dimension is $\sum_{j=1}^{m} k_j \times 1$, and $\mathbf{\Phi} = \{\phi_1, \phi_2, \cdots, \phi_m\}'$.

Next, given the sample observations $\boldsymbol{Y} = \{Y_1^*, \cdots, Y_m^*, Y_1, \cdots, Y_m\}, \boldsymbol{Y}^* = \{Y_1^*, \cdots, Y_m^*\}$, and $Y_j = \{Y_{j,N_{j1}}, \cdots, Y_{j,N_{jk_{j}}}, j = 1, 2, \cdots, m\}$, this paper will calculate the Bayesian estimates of the various parameters.

3.1: Estimation of $\mathbf{\Phi}$

Assuming that $r = \{r_i, i = 1, 2, \cdots m-1\}$ and $q = \{q_{ij}, i = 1, 2, \cdots, m, j = 1, 2, \cdots, k_1, \cdots, k_m\}$, given that d, r and q are known, the estimation of ϕ_i begins in the following. Suppose that the prior for each segment's ϕ_j follows a multivariate normal distribution:

$$p(\phi_j \mid d, r) \sim N_{q_j+1}\left(u_j, \sqrt{v_j}\right) \quad (10)$$

The likelihood function for $\boldsymbol{Y}$ is:

$$P(\boldsymbol{Y} \mid d, r, q, \mathbf{\Phi}, \mathbf{\Sigma}^2) = \prod_{j=1}^{m}\prod_{e=1}^{k_j}(2\pi\sigma_{je}^2)^{-\frac{N_{je}}{2}} \exp\left(-\frac{(Y_{j,N_{je}}^* - Y_{j,N_{je}}\phi_j)'(Y_{j,N_{je}}^* - Y_{j,N_{je}}\phi_j)}{2\sigma_{je}^2}\right) \quad (11)$$

Thus, the conditional posterior density of $\mathbf{\Phi}$ is:

$$P(\mathbf{\Phi} \mid d, r, q, \boldsymbol{Y}) \propto p(\mathbf{\Phi} \mid d, r, q) \times P(\boldsymbol{Y} \mid d, r, q, \mathbf{\Phi}, \mathbf{\Sigma}^2)$$

$$\propto \prod_{j=1}^{m}\prod_{e=1}^{k_j} \exp\left(-\frac{(\phi_j - u_j)'(\phi_j - u_j)}{2v_j}\right) \exp\left(-\frac{(Y_{j,N_{je}}^* - Y_{j,N_{je}}\phi_j)'(Y_{j,N_{je}}^* - Y_{j,N_{je}}\phi_j)}{2\sigma_{je}^2}\right)$$

$$\propto \prod_{j=1}^{m}\prod_{e=1}^{k_j} \exp\left(-\frac{(\phi_j - u_j)'(\phi_j - u_j)\sigma_{je}^2 + v_j(Y_{j,N_{je}}^* - Y_{j,N_{je}}\phi_j)'(Y_{j,N_{je}}^* - Y_{j,N_{je}}\phi_j)}{2v_j\sigma_{je}^2}\right) \quad (12)$$

Simplifying equation (12), the posterior density of $\mathbf{\Phi}$ can be calculated as follows:

$$P(\mathbf{\Phi} \mid d, r, q, \boldsymbol{Y}) \propto \prod_{j=1}^{m}\prod_{e=1}^{k_j} \exp\left(-\frac{(\phi_j - \phi_j^*)'(\phi_j - \phi_j^*)}{2\sigma_{\emptyset_j}^2}\right) \quad (13)$$

Therefore, the posterior distribution that ϕ_j follows is:

$$\phi_j \mid d, r, q \sim N(\phi_j^*, \sigma_{\emptyset_j}^2) \quad (14)$$

In which, $\phi_j^* = \sigma_{\emptyset_j}^2\left(\frac{(Y_{j,N_{je}}^*)' Y_{j,N_{je}}}{\sigma_{je}^2} + \frac{u_j}{v_j}\right)$, and $\sigma_{\emptyset_j}^2 = \left(\frac{Y'_{j,N_{je}} Y_{j,N_{je}}}{\sigma_{je}^2} + v_j^{-1}\right)^{-1}$, the posterior mean and variance formulas for $\mathbf{\Phi}$ are consistent with the conclusions derived by Frühwirth [6] for the posterior of a multivariate normal distribution.

3.2: Estimation of $\mathbf{\Sigma}^2$

Assuming that the innovation term of each autoregressive model segment is independent and comes from a discrete distribution G, which is derived from a Dirichlet process, in order to use conjugate priors to calculate the process easily, set G_0 to be an inverse gamma distribution:

$$\begin{aligned} &\varepsilon_i \sim N(0, V_i), i = 1, 2, \cdots, N \\ &V_i \sim G \\ &G \sim DP(\alpha, G_0) \\ &G_0 \sim Inv-gamma(v_1, c_1) \end{aligned} \quad (15)$$

It is worth mentioning that V_i is the variance carried by the i-th data point, while σ_{je}^2 is the variance of the e-th segment within the j-th regime of the SETAR model. G_0 is the base distribution in the Dirichlet process, and α is the precision parameter, which can control the degree of discreteness in the clustering. The posterior estimation of α will be given in the next subsection, and will not be discussed in detail here.

Thus, the innovation term of each segment independently follows an unknown distribution, which can be represented by the following mixture form:

$$\int F_N(0, V)G(dV) \quad (16)$$

Taking the j-th segment as an example, F_N is a normal distribution with mean 0 and variance V, and G is the mixture weight of $V = (V_1, V_2, \cdots, V_{N_j})$. Using the Dirichlet stick-breaking construction method from equation (4), the discrete distribution G is expressed as follows:

International Journal of Scientific Engineering and Science
Volume 8, Issue 12, pp. 68-76, 2024. ISSN (Online): 2456-7361

$$G(dV)=\sum_{i=1}^{\infty}\pi_i\delta_{\sigma_{ji}^2}(dV) \quad (17)$$

Here, δ is the indicator function, which equals 1 only when $\sigma_{ji}^2=V_i$, and 0 otherwise. The length of the i-th stick, π_i, represents the probability that V_i equals some known σ_{ji}^2, and the calculation of each π_i, is given in equation (5).

If the variance carried by each data point is known, then each segment can be classified by having the same variance. However, real data will not conform to any single distribution, so the variance carried by the data cannot be known. Therefore, we introduce the indicator variable $S=(S_1,S_2,\cdots,S_N)$, and assume that the j-th segment has k_j smaller segments. When $V_t=\sigma_{ji}^2, i=1,2,\cdots,k_j+1$, then $S_t=i$. Under the prior of the Dirichlet process, the distribution of S_t is as follows:

$$S_t\sim\sum_{i=1}^{\infty}\pi_i\delta_i \quad (18)$$

Here, the weights π_i, are defined by equation (5).

By introducing S_t into the model, we obtain:

$$\varepsilon_i\sim N(0,\sigma_{S_t}^2)$$
$$S_t\sim\sum_{i=1}^{\infty}\pi_i\delta_i$$
$$\sigma_{S_t}^2\sim G_0 \quad (19)$$

Define $\boldsymbol{\varepsilon}=(\varepsilon_{11},\cdots,\varepsilon_{1N_1},\cdots,\varepsilon_{m1},\cdots,\varepsilon_{mN_m})'$ and $\varepsilon_{ji'}=y_{ji'}-\phi_{j0}-\sum_{i=1}^{q_j}\phi_{ji}y_{ji'-i}, i'=q_j+1,q_j+2,\cdots,N_j$.

First, based on the values of the data y_t and the threshold values, it can be determined which segment of the threshold model y_t belongs to. Then, combined with equation (6), the conditional posterior distribution of V_t can be obtained:

$$V_t|\{V_{t'},\ t'\neq t\},\varepsilon_t,\alpha\sim$$
$$\frac{\alpha}{n+\alpha-1}g(\varepsilon_t)G(dV|\varepsilon_t)+\frac{1}{n+\alpha-1}\sum_{t'\neq t}f_N(\varepsilon_t|V_{t'})\delta_{V_{t'}}(dV) \quad (20)$$

The first term on the right side represents the scenario where the variance V_t carried by the data y_t comes from a new class, from which a value is drawn and multiplied by the probability of being in this new class. The second term on the right side represents the scenario where the variance V_t carried by the data y_t comes from an existing class, from which a value is drawn and multiplied by the probability of being in this existing class. And $G(dV|\varepsilon_t)\propto f_N(\varepsilon_t|0,V)G_0(dV)$, $g(\varepsilon_t)\equiv\int f_N(\varepsilon_t|0,V)G_0(V)dV$. Combining this with equation (15), we get:

$$G(dV|\varepsilon_t)\propto V^{-\frac{1}{2}}\exp(-\frac{\varepsilon_t^2}{2V})\times V^{-(v_1+1)}\exp(-\frac{c_1}{V})$$
$$\propto V^{-\left(v_1+\frac{1}{2}+1\right)}\exp(-\frac{2c_1+\varepsilon_t^2}{2V}) \quad (21)$$

Based on the form of the inverse gamma distribution's probability density function, it is known that:

$$G(dV|\varepsilon_t)\sim Inv-gamma\left(v_1+\frac{1}{2},\frac{2c_1+\varepsilon_t^2}{2}\right) \quad (22)$$

Marginal likelihood function is:

$$g(\varepsilon_t)=\int f_N(\varepsilon_t|0,V)G_0(V)dV=\frac{1}{\sqrt{2\pi}}\times\frac{(v_1)^{c_1}}{\Gamma(v_1)}\times\frac{\Gamma\left(v_1+\frac{1}{2}\right)}{\left(\frac{2c_1+\varepsilon_t^2}{2}\right)^{v_1+\frac{1}{2}}} \quad (23)$$

It is worth noting that in the above formula, the first and second terms on the right side are the constant parts of $f_N(\varepsilon_t|0,V)$ and $G_0(dV)$, respectively, and the third term is obtained by integrating the distribution of equation (22) with respect to V. Since the true variance carried by the data is unknown, we cannot classify it ideally based on whether the variances are equal. Therefore, we introduce the indicator variable S, assuming that there are initially k_j classes for the j-th segment. The indicator S_t for the t-th data point will be sampled probabilistically from $1,2,\cdots,k_j,k_j+1$. Drawing a number from $1,2,\cdots,k_j$ indicates that it belongs to one of the existing classes, while drawing the number k_j+1 indicates that it belongs to a new category. After normalization, we get:

$$S_t|\sigma_{je}^2,\varepsilon_t,\alpha,e=1,\cdots,k_j\sim$$
$$\begin{cases}\dfrac{\alpha g(\varepsilon_t)\delta_{k_j+1}(dS_t)}{\sum_{e=1}^{k_j}f_N(\varepsilon_t|\sigma_{je}^2)\delta_e(dS_t)+\alpha g(\varepsilon_t)\delta_{k_j+1}(dS_t)}\\[2ex]\dfrac{\sum_{e=1}^{k_j}f_N(\varepsilon_t|\sigma_{je}^2)\delta_e(dS_t)}{\sum_{e=1}^{k_j}f_N(\varepsilon_t|\sigma_{je}^2)\delta_e(dS_t)+\alpha g(\varepsilon_t)\delta_{k_j+1}(dS_t)}\end{cases} \quad (24)$$

Therefore, the category label for the t-th data S_t will be drawn from $1,2,\cdots,k_j,k_j+1$ with probability of (24). If S_t draws the number k_j+1, then let $k_j=k_j+1$. After knowing the distribution of S_t for the j-th segment, we begin to update the variances within each class:

International Journal of Scientific Engineering and Science
Volume 8, Issue 12, pp. 68-76, 2024. ISSN (Online): 2456-7361

$$\pi(\sigma_{je}^2|\boldsymbol{S},\boldsymbol{\varepsilon},k_j)=\prod_{t:S_t=e}^{N_j} f_N(\varepsilon_t|0,\sigma_{jS_t}^2)G_0(d\sigma^2)$$
$$\propto\prod_{t:S_t=e}^{N_j}(\sigma_{jS_t}^2)^{-\frac{1}{2}}\exp(-\frac{\varepsilon_t^2}{2\sigma_{jS_t}^2})\times(\sigma_{jS_t}^2)^{-(v_1+1)}\exp(-\frac{c_1}{\sigma_{jS_t}^2})$$
$$\propto(\sigma_{jS_t}^2)^{-\left[\frac{N_{je}+2v_1}{2}+1\right]}\exp\left(-\frac{N_{je}\varepsilon_t^2+2c_1}{2\sigma_{jS_t}^2}\right)$$
$$\sim Inv-gamma\left(\frac{N_{je}+2v_1}{2},\frac{N_{je}\varepsilon_t^2+2c_1}{2}\right) \quad (25)$$

Where $j=1,2,\cdots,m$, $e=1,2,\cdots,k_j$, $\boldsymbol{S}=\{S_1,S_2,\cdots,S_{N_j}\}$, $\boldsymbol{\varepsilon}=\{\varepsilon_1,\varepsilon_2,\cdots,\varepsilon_{N_j}\}$. Therefore, using equation (25), we can estimate the variance values of the normal distributions that each subclass in every segment follows.

3.2: Estimation of d,r,q

Esmail Amiri [7] provided the joint posterior density of d,r and q with data as conditions:

$$f(d,r,q|\boldsymbol{Y})\propto\prod_{j=1}^{m}2^{\left(-\frac{v_j}{2}+1\right)}\pi^{\frac{v_j}{2}}\Gamma(\frac{v_j}{2})(\frac{v_js_j^2}{2})^{-\frac{v_j}{2}}|Y_i'Y_i|^{-\frac{1}{2}} \quad (26)$$

In the above formula, $q=\{q_1,\cdots,q_m\}$, $r=\{r_1,\cdots,r_{m-1}\}$, $d=0,1,2,\cdots,I$, where I is an integer, $v_j=\{v_{je},e=1,2,\cdots,k_j\}$ and $s_j^2=\{s_{je}^2\}$ are given by: $s_{je}^2=(Y_{j,N_{je}}^*-Y_{j,N_{je}}\phi_j^*)'(Y_{j,N_{je}}^*-Y_{j,N_{je}}\phi_j^*)/v_{je}$, $v_{je}=N_{je}-q_j-1$. Using the MCMC method to estimate the values of d,r and q, the closed-form of each hierarchical posterior probability function or the delay given other parameters is available. The conditional posterior of r is as follows:

$$f(r|\boldsymbol{Y},d,q)\propto\prod_{j=1}^{m}2^{\left(-\frac{v_j}{2}+1\right)}\pi^{\frac{v_j}{2}}\Gamma(\frac{v_j}{2})(\frac{v_js_j^2}{2})^{-\frac{v_j}{2}}|Y_j'Y_j|^{-\frac{1}{2}} \quad (27)$$

The conditional posterior of d is given by:

$$f(d|\boldsymbol{Y},r,q)\propto\frac{\prod_{j=1}^{m}\Gamma(\frac{v_j}{2})(\frac{v_js_j^2}{2})^{-\frac{v_j}{2}}|Y_i'Y_i|^{-\frac{1}{2}}}{\sum_{d=0}^{I}\prod_{j=1}^{m}\Gamma(\frac{v_j}{2})(\frac{v_js_j^2}{2})^{-\frac{v_j}{2}}|Y_i'Y_i|^{-\frac{1}{2}}} \quad (28)$$

The parameter I represents the maximum lag parameter, which is a positive integer. The conditional posterior of q is:

$$f(q|\boldsymbol{Y},r,d)\propto\frac{\prod_{j=1}^{m}\Gamma(\frac{v_j}{2})(\frac{v_js_j^2}{2})^{-\frac{v_j}{2}}|Y_i'Y_i|^{-\frac{1}{2}}}{\sum_{q_1=0}^{n_1}\cdots\sum_{q_m=0}^{n_m}\prod_{j=1}^{m}\Gamma(\frac{v_j}{2})(\frac{v_js_j^2}{2})^{-\frac{v_j}{2}}|Y_i'Y_i|^{-\frac{1}{2}}} \quad (29)$$

In the formula, $n_1,\cdots,n_m$ represent the maximum order of autoregression for each segment, which are less than the actual number of each time series segment

3.4: The posterior estimation of α

Firstly, it is assumed that the prior distribution of α follows a gamma distribution, $\alpha\sim G(c,d)$, Given the prior density function $P(\alpha)=\frac{d^c}{\Gamma(c)}\alpha^{c-1}e^{-d\alpha}$, using the conclusion from Antoniak [8]:

$$P(k|\alpha,n)=C_n(k)n!\alpha^k\frac{\Gamma(\alpha)}{\Gamma(\alpha+n)} \quad (30)$$

In the given context, $C_n(k)=P(k|\alpha=1,n)$, where n is the number of data points and k is the number of clusters. If π represents the weight of each data point and D_n represents the collection of n data points, then when π and k are known, the data D_n is conditionally independent of α, i.e., $P(\alpha|k,\pi,D_n)=P(\alpha|k,\pi)$. When k is known, π is Conditionally independent of α, i.e., $P(\alpha|k,\pi)=P(\alpha|k)$, Therefore, we have:

$$P(\alpha|k,\pi,D_n)=P(\alpha|k)\propto P(\alpha)P(k|\alpha)=P(\alpha)\alpha^k\frac{\Gamma(\alpha)}{\Gamma(\alpha+n)} \quad (31)$$

Using the properties of the beta function and the gamma function, the ratio of the two gamma functions in the above expression can be written as:

$$\frac{\Gamma(\alpha)}{\Gamma(\alpha+n)}=\frac{(\alpha+n)\beta(\alpha+1,n)}{\alpha\Gamma(n)} \quad (32)$$

The notation $\beta(\cdot)$ refers to the beta function. Using the definition of the beta function, the transformation of the above expression is:

$$P(\alpha|k)\propto P(\alpha)\alpha^{k-1}(\alpha+n)\int_0^1 x^\alpha(1-x)^{n-1}dx \quad (33)$$

Without loss of generality,

$$(\alpha,\eta|k)=P(\alpha)\alpha^{k-1}(\alpha+n)\eta^\alpha(1-\eta)^{n-1} \quad (34)$$

Where $\eta\in(0,1)$, so $P(\alpha|k)$ is the marginal distribution of $P(\alpha,\eta|k)$. Therefore, based on equation (34), we can derive two conditional posterior distributions, $P(\alpha|k,\eta)$ and $P(\eta|k,\alpha)$, as follows:

$$P(\alpha|k,\eta)\propto P(\alpha)\alpha^{k-1}(\alpha+n)\eta^\alpha\propto\alpha^{c-1}e^{-d\alpha}\alpha^{k-1}(\alpha+n)e^{\alpha ln\eta}$$
$$=\alpha^{k+c-1}e^{-\alpha(d-ln\eta)}+n\alpha^{k+c-2}e^{-\alpha(d-ln\eta)} \quad (35)$$

Thus, the posterior estimation of α is a mixture of two gamma distributions:

$$\alpha|k,\ \eta\sim\pi_\eta G(c+k,d-ln\eta)+(1-\pi_\eta)G(c+k-1,d-ln\eta) \quad (36)$$

In equation (36), η and π_η are still unknown, so π_η is defined by the following formula: $\frac{\pi_\eta}{1-\pi_\eta}=\frac{c+k-1}{n(d-ln\eta)}$, This implies that:

International Journal of Scientific Engineering and Science
Volume 8, Issue 12, pp. 68-76, 2024. ISSN (Online): 2456-7361

$$\pi_\eta = \frac{c+k-1}{n(d-ln\eta)+c+k-1} \qquad (37)$$

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

$$\eta|k,\alpha \sim Beta(\alpha+1,n) \qquad (38)$$

By estimating π_η and η using equations (37) and (38), and combining them with equation (36), we can calculate the posterior estimate of α.

IV. Data Simulation

In this section, we illustrate the simulation study of the aforementioned model and compare this method with the OLS methods to demonstrate the superiority of our model. We consider the following TAR model, which has a single threshold value, with each segment being a first-order autoregression, and the variance of the innovation term in each segment is derived from a mixture of two variances:

$$Y_1^* = \begin{cases} 7.8+0.2Y_1+\varepsilon_1, & y_{t-1} \le 17 \\ 14+0.2Y_2+\varepsilon_2, & 17 < y_{t-1} \end{cases}$$

Where $\varepsilon_1 \sim \pi_{11}N(0,10)+\pi_{12}N(0,50)$, $\varepsilon_2 \sim \pi_{21}N(0,10)+\pi_{21}N(0,100)$.

The reason for setting $\phi_{10}=7.8$ and $\phi_{20}=14$, as well as the threshold value $r=17$, is that we are considering a situation with larger variances, in order to distinctly separate the data of the first and second segments. We set the innovations of the autoregressive models in each segment to follow a mixture of two normal distributions.

The process of using the Gibbs sampling algorithm is as follows:

Algorithm: Sample each parameter

```
Input: Y, r, d, q, Σ², Θ, α
for i=1 to maxIters
  Y₁ = {y: y ≤ r}; Y₂: = {y, y > r}
  Rearrange r, d and q, update Y₁ and Y₂
  draw Θ and α
  for jj=1 to N1, where N1 denotes the number of Y₁
    S1 = {1,2,⋯,k₁}
    draw Σ₁² = {σ²₁ₑ} for e=1 to k₁
  end for
  for ww=1 to N1, where N1 denotes the number of Y₂
    S2 = {1,2,⋯,k₂}
    draw Σ₂² = {σ²₂ₑ} for e=1 to k₂
  end for
end for
Print r, d, q, Σ², Θ, α
```

Since the initial values of the model parameters in this paper are set arbitrarily, this may result in a large discrepancy between the initial estimates and the true values. If retained, this could introduce some bias into the final estimation results. Therefore, this paper sets the number of iterations to 1000, with the first 500 iterations considered as burn-in values and discarded. The average of the parameters from the last 500 iterations is taken as the estimated result. This paper considers that the variance of the innovation term in each segment is independent and comes from a mixed distribution, and compares the calculation results with the situation where the innovation term in each segment is independent but only follows a single distribution. The parameter α does not have a true value, its magnitude merely restricts the number of variance clusters. The smaller the value of α, the fewer the number of clusters. This paper calculates the clustering parameter for the first segment, $\alpha_1=1.075$, and for the second segment, $\alpha_2=0.592$. The estimated values of the threshold r and the parameter $\mathbf{\Phi}$ are shown in the following table:

TABLE 4-1. Parameter estimation by DPMM

parameter	r	ϕ_{10}	ϕ_{11}	ϕ_{20}	ϕ_{21}
true value	17.000	7.800	0.200	14.000	0.200
DPMM	17.563	7.845	0.235	14.268	0.220

TABLE 4-2. Parameter estimation by OLS

parameter	r	ϕ_{10}	ϕ_{11}	ϕ_{20}	ϕ_{21}
true value	17.000	7.800	0.200	14.000	0.200
OLS	16.690	7.130	0.273	12.860	0.271

In tables 4-1 and 4-2, the second row represents the true values of the parameters, and the third row represents the estimated average values. Here, r is the threshold value, ϕ_{10} and ϕ_{11} represent the constant term and the coefficient of the first-order autoregressive term for the first segment, respectively, and ϕ_{20}, ϕ_{21} represent the constant term and the coefficient of the first-order autoregressive term for the second segment, respectively. From the table, it can be seen that our results for the parameter estimates are closer to the true values, indicating that our method is more effective.

We have plotted the estimates from the last 500 iterations of the threshold value, constant term, and coefficients of the autoregressive components in the heteroscedastic threshold model that we developed as follows:

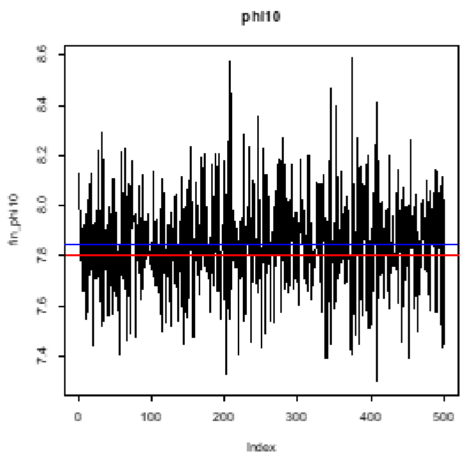

Fig. 4-1. Estimation of the constant term for the first autoregressive model segment

International Journal of Scientific Engineering and Science
Volume 8, Issue 12, pp. 68-76, 2024. ISSN (Online): 2456-7361

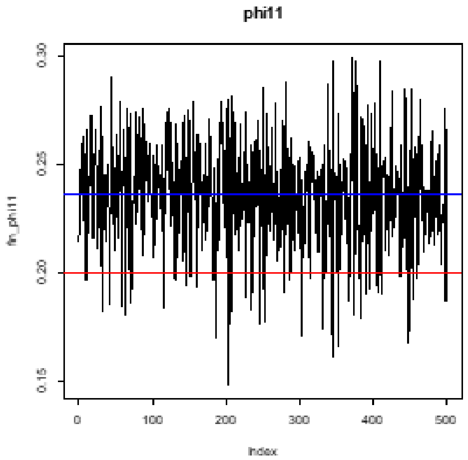

Fig. 4-2. Estimation of the coefficients for the first autoregressive model segment

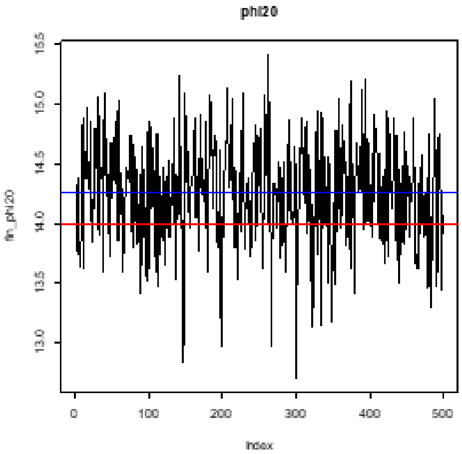

Fig. 4-3. Estimation of the constant term for the second autoregressive model segment

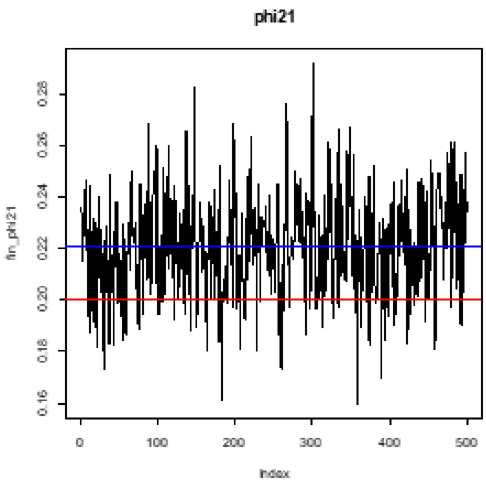

Fig. 4-4. Estimation of the coefficients for the second autoregressive model segment

Figures 4-1, 4-2, 4-3, and 4-4 respectively display the estimates of the constant terms and coefficients for the first and second autoregressive model segments, with the red line representing the true values and the blue line representing the estimated means. Figure 4-1 shows that the true value of the constant term for the first autoregressive model is 7.8, and our estimated mean is 7.845. Figure 4-2 shows that the true value of the constant term for the first autoregressive model is 0.20, and our estimated mean is 0.235. Figure 4-3 shows that the true value of the constant term for the second autoregressive model is 14, and our estimated mean is 14.268. Figure 4-4 shows that the true value of the constant term for the first autoregressive model is 0.20, and our estimated mean is 0.220.

TABLE 4-3. DPMM estimation of mixed variances

parameter	σ_{11}^2	σ_{12}^2	σ_{21}^2	σ_{22}^2
true value	10	50	10	100
DPMM	11.322	40.394	10.254	14.377

TABLE 4-4. OLS estimation of variance

parameter	σ_{11}^2	σ_{12}^2	σ_{21}^2	σ_{22}^2
true value	10	50	10	100
OLS	35.861	35.861	35.607	35.607

In tables 4-3 and 4-4, the second row represents the true values of the variances, and the third row represents the estimated mean values of the variances. The OLS method assumes that the innovation terms in each segment only follow the same distribution; therefore, it provides the same estimates for σ_{11}^2 and σ_{12}^2, as well as for σ_{21}^2 and σ_{22}^2. It is worth mentioning that when we use the Dirichlet process to cluster the innovation terms, the number of clusters is often more than two. We choose to take the two clusters with the most members as the clustering results and discard the others with very few members. The results show that the DPMM method, when estimating the number of mixed variances, yields the same number as the initially designed number of mixed distributions. When setting up the data, the first segment's innovation term follows a mixed distribution given by:

$\varepsilon_1 \sim 0.370N(0,10)+0.630N(0,100)$. Our estimated mixed distribution for the first segment is:

$\varepsilon_1 \sim 0.372N(0,11.322)+0.628N(0,40.394)$.

The second segment's innovation term follows a mixed distribution given by:

$\varepsilon_2 \sim 0.667N(0,10)+0.333N(0,100)$. Our estimated mixed distribution for the second segment is:

$\varepsilon_2 \sim 0.738N(0,10.254)+0.262N(0,14.377)$.

V. REAL EXAMPLE

This subsection applies the proposed method to a real dataset to better demonstrate the practicality of our method with actual data. The analysis is conducted using the closing prices of Nanjing Photoelectric, with the raw data consisting of closing prices every 5 minutes from January 14, 2021, to November 10, 2023, totaling 2,704 valid data points. The data source is Eastmoney Choice: https://choice.eastmoney.com. The original data of Nanjing Photoelectric is shown in Figure 5-1.

We set the number of iterations to 1000, with the first 500 being discarded as burn-in values. We then use the results from the last 500 iterations for our study, and plot the estimated values of the constant terms and coefficients for each segment as shown in the following figure.

International Journal of Scientific Engineering and Science
Volume 8, Issue 12, pp. 68-76, 2024. ISSN (Online): 2456-7361

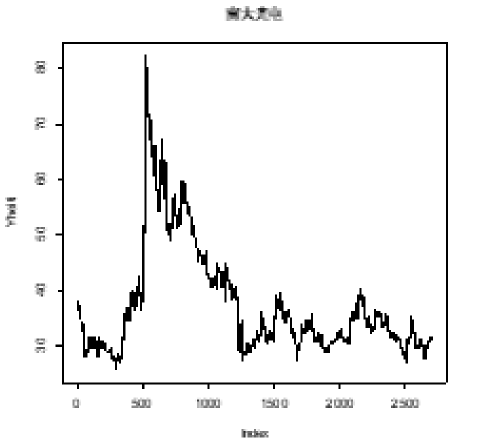

Fig. 5-1. The plot of all data for Nanjing Photoelectric.

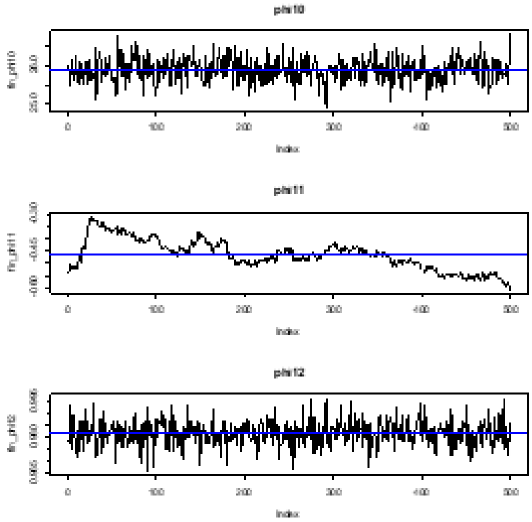

Fig. 5-2. Estimation of the first autoregressive model segment

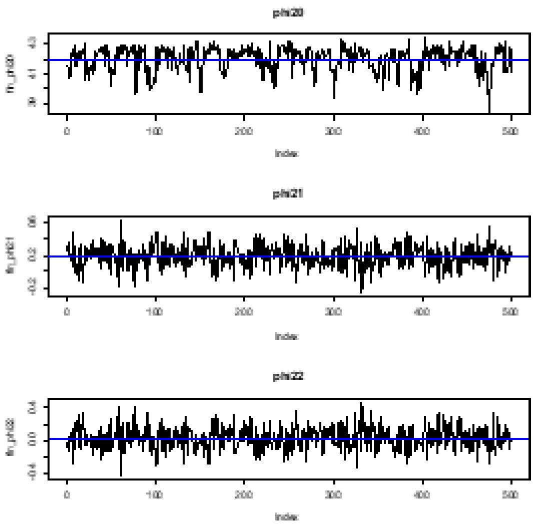

Fig. 5-3. Estimation of the second autoregressive model segment

Figure 5-2, from top to bottom, represents the constant term, the coefficient of the first-order autoregressive term, and the coefficient of the second-order autoregressive term. Figure 5-3, from top to bottom, shows the constant term, the coefficient of the first-order autoregressive term, and the coefficient of the second-order autoregressive term.

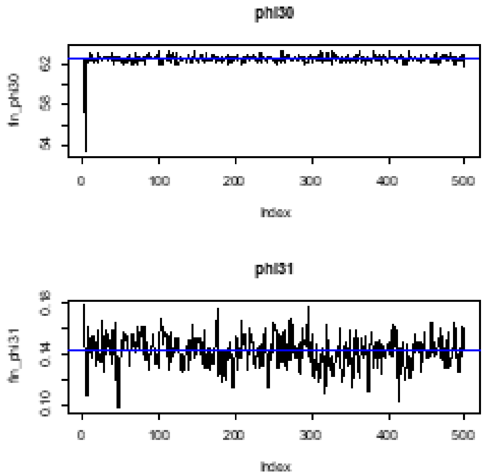

Fig. 5-4. Estimation of the third autoregressive model segment

Figure 5-4, from top to bottom, shows the constant term and the coefficient of the first-order autoregressive term.

In the figure above, the blue line represents the mean of the 500 estimated values, which is also our final estimate for the parameter. It can be observed from the graph that the estimated values fluctuate around the blue line, indicating that the estimation results are stable. This suggests that the number of iterations we selected is feasible.

We modeled the data from Nanjing Photoelectric and obtained a three-segment threshold model with two threshold values. Figures 5-2, 5-3, and 5-4 represent the constant terms and the coefficients of the lagged variables for the first, second, and third segments, respectively. The autoregressive order of the first segment is second-order, the autoregressive order of the second segment is second-order, and the autoregressive order of the third segment is first-order. The clustering parameters for the segments are $\alpha_1 = 1.660$, $\alpha_2 = 0.240$ and $\alpha_3 = 0.017$. The resulting threshold model is as follows:

$$Y_t = \begin{cases} 25.889 - 0.465Y_{t-1} + 0.982Y_{t-2} + \varepsilon_1, Y_{t-1} \le 42.713 \\ 41.980 + 0.188Y_{t-1} + 0.030Y_{t-2} + \varepsilon_2, \\ \qquad 42.713 < Y_{t-1} \le 68.700 \\ 62.577 + 0.143Y_{t-1} + \varepsilon_3, 68.700 < Y_{t-1} \end{cases}$$

The aforementioned model reveals that the constant term of the first autoregressive segment is $\phi_{10} = 25.889$, the coefficient of the first-order autoregressive term is $\phi_{11} = -0.456$, and the coefficient of the second-order autoregressive term is $\phi_{12} = 0.982$. Our model can be applied to various fields, such as the stock market, banking credit risk, healthcare, and insurance claims.

TABLE 5-1. Estimation values of the DPMM

	Σ_1^2	Σ_1^2	Σ_1^2	Σ_2^2	Σ_2^2	Σ_3^2
parameter	σ_{11}^2	σ_{12}^2	σ_{13}^2	σ_{21}^2	σ_{22}^2	σ_{31}^2
Estimation	0.320	0.379	0.480	27.774	0.290	11.225

Table 5-1 displays the estimated variances of the innovation terms for each segment. We have also calculated the mixture forms of the normal distributions that the innovation terms of each segment follow. The innovation term of the first segment is a mixture of three normal distributions, all with a mean of 0:

$$\varepsilon_1 \sim 0.670N(0,\sigma_{11}^2)+0.193N(0,\sigma_{12}^2)+0.137N(0,\sigma_{13}^2).$$

The innovation term of the second segment is a mixture of two normal distributions:

$$\varepsilon_2 \sim 0.604N(0,\sigma_{21}^2)+0.396N(0,\sigma_{22}^2).$$

The innovation term of the third segment comes from a single normal distribution: $\varepsilon_3 \sim N(0,\sigma_{31}^2)$.

VI. SUMMARY

This paper, based on the Dirichlet process, estimates the parameters of a heteroscedastic threshold model and derives the posterior distributions of each parameter under reasonable prior settings. Our method is more flexible because we do not restrict the distribution that the autoregressive innovation terms follow; we can use a mixed normal distribution to fit it. The number of components in the mixture distribution does not need to be preset and can be estimated through the Dirichlet process. Simulation experiments have demonstrated that when the variance is large, our method can estimate the number of components and their weights in the mixed distribution, and the estimates of the parameters are closer to the true values.

REFERENCES

[1] Tong H. Ona Threshold Model[J]. 1978.
[2] Wang H, Li G, Jiang G. Robust regression shrinkage and consistent variable selection through the LAD-Lasso[J]. Journal of Business & Economic Statistics, 2007, 25(3): 347-355.
[3] Tong, H. Threshold Models in Non-linear Time Series Analysis, Lecture Notes in Statistics. 1983.
[4] Wu Xizhi. Applied Time Series Analysis: Accompanied by R Software [M]. China Machine Press, 2018.1: 212-215.
[5] Tsay R S. Testing and modeling threshold autoregressive processes[J]. Journal of the American statistical association, 1989, 84(405): 231-240.
[6] Frühwirth-Schnatter S. Finite mixture and Markov switching models[M]. Springer, 2006.
[7] 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.
[8] Antoniak C E. Mixtures of Dirichlet processes with applications to Bayesian nonparametric problems[J]. The annals of statistics, 1974: 1152-1174.
[9] Tong H. Non-linear time series: a dynamical system approach[M]. Oxford university press, 1990.
[10] Chen C W S, Lee J C. Bayesian inference of threshold autoregressive models[J]. Journal of time series analysis, 1995, 16(5): 483-492.
[11] Chen C W S. A Bayesian analysis of generalized threshold autoregressive models[J]. Statistics & probability letters, 1998, 40(1): 15-22.
[12] Tsay R S. Testing and modeling threshold autoregressive processes[J]. Journal of the American statistical association, 1989, 84(405): 231-240.
[13] Pan J, Xia Q, Liu J. Bayesian analysis of multiple thresholds autoregressive model[J]. Computational Statistics, 2017, 32: 219-237.
[14] Zhang R, Liu Q, Shi J. Estimation of generalized threshold autoregressive models[J]. Communications in Statistics-Theory and Methods, 2023, 52(18): 6456-6474.
[15] Ferguson T S. A Bayesian analysis of some nonparametric problems[J]. The annals of statistics, 1973: 209-230.
[16] Escobar M D, West M. Bayesian density estimation and inference using mixtures[J]. Journal of the american statistical association, 1995, 90(430): 577-588.
[17] Hong L, Martin R. A flexible Bayesian nonparametric model for predicting future insurance claims[J]. North American Actuarial Journal, 2017, 21(2): 228-241.
[18] Adesina O S, Agunbiade D A, Oguntunde P E. Flexible Bayesian Dirichlet mixtures of generalized linear mixed models for count data[J]. Scientific African, 2021, 13: e00963.
Example for papers presented at conferences (unpublished):
[19] Zhang Yongxia, Meng Shengwang, Tian Maozai. Semiparametric Bayesian Hierarchical Quantile Regression Model and Its Application in Insurance Company Cost Analysis [J]. Journal of Applied Statistics and Management, 2021, 40(03): 381-394.
[20] Kalli M, Griffin J E. Bayesian nonparametric vector autoregressive models[J]. Journal of econometrics, 2018, 203(2): 267-282.
[21] Liu Wenying. Estimating a Semiparametric Double-Autoregressive Model with a Dirichlet Process Mixture [D]. Xiamen University, 2015.
[22] Lau J W, So M K P. A Monte Carlo Markov chain algorithm for a class of mixture time series models[J]. Statistics and computing, 2011, 21(1): 69-81.
[23] Nakatsuma T. A Markov-chain sampling algorithm for GARCH models[J]. Studies in Nonlinear Dynamics & Econometrics, 1998, 3(2).
[24] Peng Mingyi. Research on Comment Clustering Based on Dirichlet Process and Multinomial Distribution Mixture Model [D]. Shenyang Ligong University, 2021.
[25] Teh Y W , Jordan M I , Beal M J ,et al.Hierarchical Dirichlet Processes[J].Journal of the American Statistical Association, 2006, 101(December):1566-1581.
[26] qunyong wang. Fixed-effect panel threshold model using Stata[J], Stata Journal, 2015, 15(1): 121-134.
[27] D BLACKWELL, and JB MACQUEEN. "Ferguson Distributions Via Polya Urn Schemes", Annals of Statistics 1.2 (1973): 353-355.