Journal of Statistical Theory and Practice (2020) 14:42
https://doi.org/10.1007/s42519-020-00108-5

ORIGINAL PAPER

Confidence Distributions for Skew Normal Change-Point Model Based on Modified Information Criterion

Suthakaran Ratnasingam[1] · Wei Ning[1,2]

Published online: 12 June 2020

Abstract
In this paper, we consider a skew normal change-point model. Instead of only providing the point estimate of the change location, we propose an estimating procedure based on the confidence distribution combining with the modified information criterion to construct the confidence set for the change location. The simulations indicate the advantages of the proposed method comparing to the existing method in terms of coverage probabilities and average lengths of the confidence sets, especially when the change occurs at the very beginning or in the very end. The proposed method is applied to two stock market data to illustrate the detection and the estimation procedures.

Keywords Confidence distribution · Skew normal distribution · Change point detection · Information criterion · Coverage probability · Average lengths

1 Introduction

The concept of a confidence distribution (CD) has its roots in Fisher's fiducial distribution. A CD is similar to a point estimator or an interval estimator, but it uses a sample-dependent distribution function on the parameter space to estimate the parameter of interest. It also can provide confidence intervals of all nominal levels for a parameter of interest through confidence curves. Xie and Singh [44] gave a detailed review of recent developments in confidence distributions. The first time the terminology "confidence distribution" was used in a formal publication dated back to [15]. In his paper, Cox suggested that a confidence distribution "can either be defined directly or can be introduced in terms of the set of all confidence intervals at different levels of probability." The formal modern definition for CD can be found

✉ Wei Ning
wning@bgsu.edu

1 Department of Mathematics and Statistics, Bowling Green State University, Bowling Green, OH 43403, USA

2 School of Mathematics and Statistics, Beijing Institute of Technology, Beijing 100081, China

in [35, 40, 41]. The concept of CD is broad, and it has wide range of applications including bootstrap distributions, p value functions, normalized likelihood functions and Bayesian posteriors, among others. For example, [35] used the CD to obtain the reduced likelihood function. Singh et al. [40] proposed a method for combining information from independent studies through confidence distributions. Singh et al. [41] provided a formal definition of a CD and an asymptotic confidence distribution (aCD). Singhand Xie [39] proposed a function called a CD posterior which uses the information from observed data with its corresponding prior information. Xie and Singh [44] proposed a CD approach which incorporates the expert opinions while analyzing clinical data with binary outcomes. Shen et al. [38] proposed a predictive distribution function with CDs. Celine et al. [8] proposed a parametric method by using confidence distributions for detecting change points and obtained confidence curves for change locations.

The detection of change points is a process which attempts to identify points in time when the probability distribution of a stochastic process or time series changes. Change-point analysis plays an important role in financial time series analysis, economy, quality control, genome research, signal processing, medical research, statistical calibration and so on. The study of the change-point problem was dated back to [31, 32] who first proposed a procedure to detect only one change in a parameter and has been extensively studied since then. For instance, [13, 17, 20, 21] studied the testing and estimation of a change in the mean of a normal model. Hsu [26] and Inclan [28] studied change-point problem for the variance in a normal model. Readers are referred to [10, 11, 14] for more details of parametric and nonparametric methods on different types of change-point problems. The use of the information criteria in the view of the model selection as an alternative for change-point detection has been extensively studied such as [7, 9, 12, 19, 23, 30, 33], to name a few.

As mentioned earlier, [8] suggested a change-point detection procedure based on a CD incorporating the likelihood function by profiling over the other parameters and obtained the confidence curve for the change location, consequently the confidence sets at any confidence levels. However, as [12] pointed out, the estimation of the change location through the regular likelihood function does not consider the contribution of the change location to the model complexity as a parameter. Clearly, when the change location k is located in the middle of the data, all the parameters are effective. However, the parameter space of the model becomes redundant when the location of change k is near 1 or n. To tackle this issue, [12] proposed the modified information criterion (MIC) by considering the complexity of the penalty term which could relate to the change-point locations. This method assigns larger penalty when the change point is near the beginning or the end of the data. In this paper, we propose a CD-based procedure for a skew normal change-point model incorporating modified information criterion (MIC). Furthermore, the presence of noise in the data can influence the intrinsic nature of data and cause changes. Therefore, we also verify the statistical significance of the detected change point through MIC-based test statistic.

This paper is organized as follows: In Sect. 2, we go over the ideas of MIC and construction of confidence curves through CDs briefly and then introduce the procedure based on CDs associated with MIC for a skew normal change-point model.

In Sect. 3, simulations are conducted to investigate the performance of the proposed method and compare with some other existing method in terms of coverage probabilities and average lengths of confidence sets. The proposed method is applied to two stock market data to illustrate the detection and the estimation procedures in Sect. 4. Some discussion is provided in Sect. 5.

2 Methodology

2.1 Modified Information Criterion (MIC)

Let $x_1, \dots, x_n$ be a random sample drawn from the density function $f(x;\Theta)$. The Schwarz information criterion (SIC) proposed by [34] is given as follows:

$$\mathrm{SIC} = -2l_n(\hat{\Theta}) + \dim(\hat{\Theta})\log(n), \tag{1}$$

where $l_n(\cdot)$ is the log-likelihood function of the random sample, $\hat{\Theta}$ is the maximum likelihood estimate (MLE) of the parameter Θ and $\dim(\hat{\Theta}_k)$ is the dimension of the parameter space. We denote Θ_L, Θ_R to be the pre-change and post-change parameters, respectively, and $\hat{\Theta}_L, \hat{\Theta}_R$ to be the MLEs of the pre-change and post-change parameters. In general, the change-point problem can be treated as the model selection problem by selecting a better model between the null hypothesis of no change and the alternative hypothesis of at least one change existing. Therefore, the SIC in the context of having at least one change can be written as

$$\mathrm{SIC}(k) = -2l_n(\hat{\Theta}_L(k), \hat{\Theta}_R(k), k) + \left[2\dim(\hat{\Theta}_L(k)) + 1\right]\log(n), \tag{2}$$

where $1 \le k < n$ and (1) defines the SIC under the null hypothesis of no change which we denote it as SIC(n). However, as Chen et al. [12] pointed out, (2) does not consider the change location to be a parameter which may cause the redundancy of the parameter space when the change occurs near the beginning or the end of data. Therefore, the modified information criterion (MIC) proposed by Chen et al. [12] is given as follows. Under the null hypothesis of no change, the MIC is defined as:

$$\mathrm{MIC}(n) = -2l_n(\hat{\Theta}) + \dim(\hat{\Theta})\log(n), \tag{3}$$

where $\hat{\Theta}$ maximizes $l_n(\Theta)$. Therefore, under H_0, both SIC(n) and MIC(n) are same. Under the alternative hypothesis, the MIC is defined as:

$$\mathrm{MIC}(k) = -2l_n(\hat{\Theta}_L(k), \hat{\Theta}_R(k), k) + \left[2\dim(\hat{\Theta}_L(k)) + \left(\frac{2k}{n} - 1\right)^2\right]\log(n), \tag{4}$$

where $1 \le k < n$. The difference between (2) and (4) is that (4) considers the contribution of the change location k to the model as a parameter. If $\mathrm{MIC}(n) > \min_{1\le k<n}\mathrm{MIC}(k)$, then we select the model with a change point and the estimate of the change point is given by

$$\mathrm{MIC}(\hat{k}) = \min_{1 \le k < n} \mathrm{MIC}(k). \tag{5}$$

Moreover, for the purpose of verifying the statistical significance of the detected change point, the associated MIC-based test statistic is defined as:

$$S_n = \mathrm{MIC}(n) - \min_{1 \le k < n} \mathrm{MIC}(k) + \dim(\Theta) \log(n), \tag{6}$$

where MIC(n) and MIC(k) are defined in (3) and (5). Chen et al. [12] showed, under Wald conditions and the regularity conditions, as $n \longrightarrow \infty$,

$$S_n \longrightarrow \chi^2_d, \tag{7}$$

in distribution under H_0, where d is the dimension of Θ.

2.2 Profile log-likelihood and Deviance Function

Suppose observations $x_1, \ldots, x_k$ coming from the population with the density function $f(x, \Theta_L)$ and $x_{k+1}, \ldots, x_n$ coming from the population with the density function $f(x, \Theta_R)$. The log-likelihood function is:

$$l(k, \Theta_L, \Theta_R) = \sum_{i \le k} \log(f(x, \Theta_L)) + \sum_{i \ge k+1} \log(f(x, \Theta_R)). \tag{8}$$

The profile log-likelihood function can be obtained by maximizing the log-likelihood function (8) over Θ_L and Θ_R for a given k. It can be defined as:

$$l_{prof}(k) = \max_{\Theta_L, \Theta_R}(l(k, \Theta_L, \Theta_R)) = l(k, \hat{\Theta}_L, \hat{\Theta}_R), \tag{9}$$

where $\hat{\Theta}_L$ and $\hat{\Theta}_R$ are MLEs of Θ_L and Θ_R for a given k, respectively. Then, the estimated change location $\hat{k}$ is given by $l_{prof}(\hat{k}) = \max_k(l_{prof}(k))$. After $\hat{k}$ is obtained, the deviance function is given by

$$D(k, \boldsymbol{x}) = 2\{l_{prof}(\hat{k}) - l_{prof}(k)\}, \tag{10}$$

where $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)$. To construct a confidence curve for k based on the deviance function, we consider the estimated distribution of $D(k, \boldsymbol{x})$ at position k as follows:

$$\Psi_k(x) = P_{k, \hat{\Theta}_L, \hat{\Theta}_R}\{D(k, \boldsymbol{x}) < x\}, \tag{11}$$

where $x \in \mathbb{R}$. In the case of continuous parameters, Wilks theorem states that $\Psi_k(x)$ is approximately the distribution function of a χ^2_1. However, Wilks theorem does not hold for a discrete parameter k. Therefore, we compute Ψ_k through the simulations. The confidence curve can be constructed as:

$$cc(k, \boldsymbol{x}_{obs}) = \Psi_k(D(k, \boldsymbol{x}_{obs})) = P_{k, \hat{\Theta}_L, \hat{\Theta}_R}\{D(k, \boldsymbol{x}) < D(k, \boldsymbol{x}_{obs})\}. \tag{12}$$

The probability that $cc(k, \boldsymbol{x}_{obs}) < \alpha$, under the true value of k, is often approximated well with α. Then, the confidence sets for k can be visualized using the plot $cc(k, \boldsymbol{x}_{obs})$. The $cc(k, \boldsymbol{x}_{obs})$ is the acceptance probability for k, or one minus the p value for testing that value of k by using the deviance-based test which rejects the null hypothesis for high values of $D(k, \boldsymbol{x})$. We compute Ψ_k and hence $cc(k, \boldsymbol{x}_{obs})$ by simulations as follows:

$$cc(k, \boldsymbol{x}_{obs}) = \frac{1}{B} \sum_{j=1}^{B} I\{D(k, \boldsymbol{x}_j^*) < D(k, \boldsymbol{x}_{obs})\}, \tag{13}$$

for large number of B of simulated copies of dataset $\boldsymbol{x}^*$. For each possible value of k, we simulate data $\boldsymbol{x}_j^*, j = 1, \dots, B$ from $f(x, \Theta_L)$ and $f(x, \Theta_R)$ to the left and right side of k, respectively. See [8] for more details. In our proposed procedure to construct the confidence curves, it is different from the [8] approach here to estimate k. Instead of estimating the change location k by maximizing the profile-likelihood function over all possible values of k, we estimate k using (5) by considering the impact of change locations. The MIC-based statistics S_n in (6) can be used to confirm a significant change statistically to avoid the fluctuations caused by noise.

2.3 Changes in All Three Parameters in a Skew Normal Distribution

The skew normal distribution (*SN*) was introduced by [3] which allows to regulate skewness in the dataset. The probability distribution function of a skew normal random variable X is given by

$$f_X(x) = \frac{2}{\sigma} \phi\left(\frac{x-\mu}{\sigma}\right) \Phi\left(\lambda \frac{x-\mu}{\sigma}\right), \quad x \in \mathbb{R} \tag{14}$$

and the cumulative distribution function (CDF) of the *SN* distribution is:

$$F_X(x) = \Phi\left(\frac{x-\mu}{\sigma}\right) - 2T\left(\frac{x-\mu}{\sigma}, \lambda\right), \quad x \in \mathbb{R} \tag{15}$$

where T is Owen's function, and ϕ and Φ are the probability distribution function and cumulative distribution function of the standard normal distribution. $\mu \in \mathbb{R}$ is the location parameter, $\sigma \in \mathbb{R}^+$ is the scale parameter and $\lambda \in \mathbb{R}$ is the shape parameter. We denote $X \sim SN(\mu, \sigma, \lambda)$. When $\lambda = 0$, the $SN(\mu, \sigma, \lambda)$ reduces to the normal $N(\mu, \sigma)$. Several basic properties of the skew normal distribution were studied by [3]. Readers are referred to [4] for more details of skew normal distribution family and recent developments on this direction. The multivariate case of the skew normal distribution was investigated by [5]. Although many methods have been proposed for making statistical inference for the skew normal distribution family, only a few of the literature is available on the change-point detection. Arellano-Valle et al. [2] proposed a Bayesian approach for detecting changes in parameters in a skew normal model. Ngunkeng and Ning [30] proposed a skew normal change-point model based on the Schwarz information criterion (SIC). Their method was improved by Said et al. [33] where they considered modified information criterion to detect changes in a skew normal model. A change-point problem for a skew normal distribution can be stated as follows:

$$x_i \sim \begin{cases} SN(\mu_L, \sigma_L, \lambda_L) & i = 1, \dots, k, \\ SN(\mu_R, \sigma_R, \lambda_R) & i = (k+1), \dots, n, \end{cases} \tag{16}$$

where $k \in \{1, \dots, (n-1)\}$ is the unknown change-point location and needed to be estimated. We are testing the following hypotheses:

$$\begin{aligned} H_0 \,:\, & \mu_1 = \mu_2 = \cdots = \mu_n = \mu, \\ & \sigma_1 = \sigma_2 = \cdots = \sigma_n = \sigma, \\ & \lambda_1 = \lambda_2 = \cdots = \lambda_n = \lambda, \end{aligned}$$

versus

$$\begin{aligned} H_1 \,:\, & \mu_1 = \cdots = \mu_k \neq \mu_{k+1} = \cdots = \mu_n, \\ & \sigma_1 = \cdots = \sigma_k \neq \sigma_{k+1} = \cdots = \sigma_n, \\ & \lambda_1 = \cdots = \lambda_k \neq \lambda_{k+1} = \cdots = \lambda_n, \end{aligned}$$

where $1 \le k < n$ is the unknown change-point location with $(\mu_L, \sigma_L, \lambda_L)$ and $(\mu_R, \sigma_R, \lambda_R)$ being the parameters of left and right side of the change-point location k. We assume there is at most one change in data since multiple-change detection can always be decomposed to multistages with at most a single change at each stage using the binary segmentation method proposed by [43]. The log-likelihood function under the null hypothesis is defined as follows:

$$l_{H_0}(\mu, \sigma, \lambda) = n \log\left(\frac{2}{\sigma}\right) - \frac{1}{2}\sum_{i=1}^{n} \log\left(\phi\left(\frac{x_i - \mu}{\sigma}\right)^2\right) + \sum_{i=1}^{n} \log\left(\Phi\left(\lambda \frac{x_i - \mu}{\sigma}\right)\right). \tag{17}$$

To obtain the MLEs of μ, σ and λ, we take the first derivative of the log-likelihood function (17) with respect to the parameters μ, σ, λ and set the equations equal to zero. Under the alternative hypothesis,

$$\begin{aligned} l_{H_1}(k) = & \left\{ k \log\left(\frac{2}{\sigma_L}\right) - \frac{1}{2}\sum_{i=1}^{k} \log\left(\phi\left(\frac{x_i - \mu_L}{\sigma_L}\right)^2\right) + \sum_{i=1}^{k} \log\left(\Phi\left(\lambda_L \frac{x_i - \mu_L}{\sigma_L}\right)\right)\right\} \\ & + \left\{ (n-k) \log\left(\frac{2}{\sigma_R}\right) - \frac{1}{2}\sum_{i=k+1}^{n} \log\left(\phi\left(\frac{x_i - \mu_R}{\sigma_R}\right)^2\right) \right. \\ & \left. + \sum_{i=k+1}^{n} \log\left(\Phi\left(\lambda_R \frac{x_i - \mu_R}{\sigma_R}\right)\right)\right\}. \end{aligned} \tag{18}$$

To obtain the MLEs of $\mu_L, \mu_R, \sigma_R, \sigma_L, \lambda_R$ andλ_L, we take the first derivative of the log-likelihood function (18) with respect to the parameters and set the equations equal to zero. Denote $\hat{\mu}_L, \hat{\mu}_R, \hat{\sigma}_R, \hat{\sigma}_L, \hat{\lambda}_R, \hat{\lambda}_L$ to be the MLEs of the $\mu_L, \mu_R, \sigma_R, \sigma_L, \lambda_R$, λ_L, respectively. The profiled log-likelihood function can be derived from (9). The modified information criteria can be obtained from (3). The estimate of the change point $\hat{k}$ is the value which minimizes equation (5). Further, the deviance function and the confidence curve can be obtained from (10) and (11), respectively.

3 Simulation Study

In this section, simulations will be conducted to investigate the performance of the proposed change-point detection method based on a CD and compare with the one proposed by [8]. To make fair comparisons, we perform two methods under the same settings and study their coverage probabilities and the average sizes of the confidence sets of the change points. A confidence set of a change point is defined by $\{k : cc(k, x) \leq \alpha\}$. Correspondingly, the size of a confidence set is defined by the number of k belonging to the confidence set for a given nominal level α.

Through all the simulations, we only consider a single change scenario since the multiple-change case can always be dealt with the binary segmentation method. The data before the change point k are always generated from $SN(0, 1, 1)$, and the data after the change point are generated from $SN(\mu, \sigma, \lambda)$ where $\mu = \{1, 1.5, 2\}$, $\sigma = \{2, 2.5, 3\}$ and $\lambda = \{3, 3.5, 4\}$. We consider two sample sizes $n = 50$ and $n = 100$. For the first sample size, we set up the changes occurring at $k = \{10, 20, 25\}$. For the second sample size $n = 100$, we set up the changes occurring at $k = \{10, 20, 40, 50\}$. The choices of k are approximately corresponding to the scenarios that a chance occurs at the very beginning, in the middle and in the very end of data. We do not consider the changes occurring in the second half of the data, for example $k = 30, 40$ and $k = 60, 80, 90$, since the performances will be similar due to the symmetric property. In our simulations, we consider three different approaches to obtain the estimated change location $\hat{k}$ to calculate the deviance function $D(k, x)$ given in (10). The first approach is simply based on the modified information criterion (MIC) given in (5). The second approach is based on the test statistic S_n which verifies statistically significance of the estimated change location $\hat{k}$ to avoid the impact of noise in the data. The third approach log-like method is by [8] which obtained $\hat{k}$ by maximizing the profile likelihood function $l_{prof}(k)$ given in (9) for all possible values of k without considering the impact of the location of the change. One thousand simulations are conducted for each scenario.

Table 1 lists simulation results of coverage probabilities for $n = 50$ and various confidence levels 0.50, 0.90, 0.95 and 0.99. From the results, we observe that the MIC and S_n provide comparable coverage probabilities to the log-like method by [8], and in general, the MIC performs slightly better than the log-like method. As the increases in the differences among parameters, three methods perform similarly and the coverage probabilities get closer to the confidence levels we set up. Table 2 provides the average of sizes of confidence sets. MIC and S_n methods have similar average sizes. However, both of them provide smaller average sizes of confidence sets than the ones by log-like method by [8], especially when the change occurs at the beginning (equivalently, in the end) of the data. For example, for $SN(1, 2, 3)$ case with $k = 10$ and $\alpha = 0.99$, the average size of confidence sets by MIC method is 24.86 and is 24.68 by S_n method, comparing to 30.06 by

Table 1 Comparisons of coverage probabilities, $n = 50$

	α	$k = 10$			$k = 20$			$k = 25$		
		MIC	loglik	S_n	MIC	loglik	S_n	MIC	loglik	S_n
SN(1, 2, 3)	0.50	0.40	0.37	0.35	0.40	0.37	0.32	0.41	0.39	0.35
	0.90	0.73	0.72	0.70	0.68	0.65	0.63	0.69	0.66	0.62
	0.95	0.79	0.78	0.77	0.76	0.71	0.69	0.74	0.73	0.70
	0.99	0.91	0.89	0.89	0.89	0.86	0.83	0.86	0.86	0.82
SN(1.5, 2.5, 3.5)	0.50	0.43	0.41	0.40	0.44	0.43	0.40	0.46	0.44	0.38
	0.90	0.76	0.74	0.73	0.78	0.75	0.71	0.80	0.80	0.74
	0.95	0.83	0.81	0.79	0.87	0.85	0.81	0.88	0.87	0.83
	0.99	0.94	0.93	0.91	0.96	0.94	0.92	0.96	0.95	0.93
SN(2, 3, 4)	0.50	0.49	0.47	0.46	0.49	0.45	0.49	0.50	0.49	0.45
	0.90	0.82	0.81	0.78	0.90	0.80	0.82	0.90	0.89	0.81
	0.95	0.89	0.88	0.85	0.93	0.85	0.88	0.95	0.94	0.89
	0.99	0.96	0.96	0.95	0.99	0.93	0.96	0.99	0.99	0.96

Table 2 The comparisons of average sizes of confidence sets, $n = 50$

	α	$k = 10$			$k = 20$			$k = 25$		
		MIC	loglik	S_n	MIC	loglik	S_n	MIC	loglik	S_n
SN(1, 2, 3)	0.50	7.15	7.81	7.04	5.75	6.62	5.72	5.84	6.68	5.86
	0.90	13.78	17.04	13.48	9.71	11.82	9.65	9.32	11.54	9.34
	0.95	17.10	21.23	16.81	11.56	14.28	11.51	10.93	13.84	10.97
	0.99	24.86	30.06	24.68	16.33	20.38	16.25	15.38	19.81	15.38
SN(1.5, 2.5, 3.5)	0.50	4.91	5.61	4.87	4.79	5.13	4.77	5.02	5.17	5.04
	0.90	8.64	10.10	8.56	6.98	7.79	6.95	7.02	7.69	7.04
	0.95	10.59	12.53	10.47	7.99	9.16	7.96	7.98	8.89	7.99
	0.99	15.79	18.93	15.60	10.80	12.84	10.77	10.46	12.04	10.48
SN(2, 3, 4)	0.50	4.29	5.88	4.29	4.58	5.12	4.56	5.03	5.19	5.01
	0.90	6.97	8.58	6.96	6.22	7.21	6.21	6.61	7.17	6.61
	0.95	8.28	10.46	8.26	6.97	8.16	6.95	7.30	8.06	7.30
	0.99	12.14	15.56	12.12	8.93	10.79	8.92	9.11	10.44	9.11

log-like method. For $k = 25$, the average sizes of confidences are 15.38 for both MIC and S_n but 19.81 for log-like method. We also observe that, as the increases in differences among parameters, all three methods obtain similar average sizes of confidence sets. Figures 1 and 2 show the graphs of coverage probability and average size of confidence sets comparisons between S_n and the log-like, the

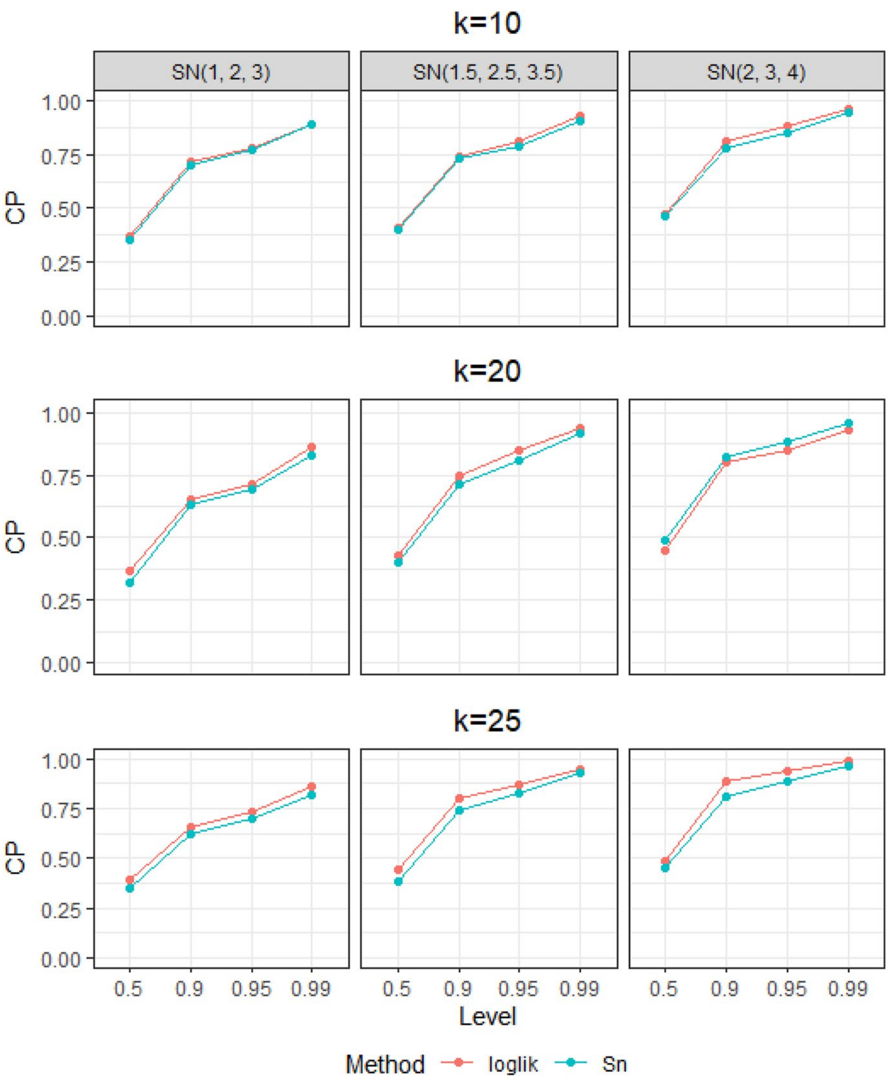

Fig. 1 The coverage probability comparison when all parameters change at various change-point positions, $k = 10, 20, 25$ and the sample size $n = 50$

likelihood method used in [8]. Since MIC and S_n perform very similarly, therefore, we only graph the curves for S_n and log-like methods.

Tables 3 and 4 list the simulation results for the coverage probabilities and average sizes of confidence sets for $n = 100$ by three approaches with various confidence levels and change points. Same pattern is observed as the one observed from Tables 1 and 2. Figures 3 and 4 show the graphs for coverage probabilities and average sizes, respectively.

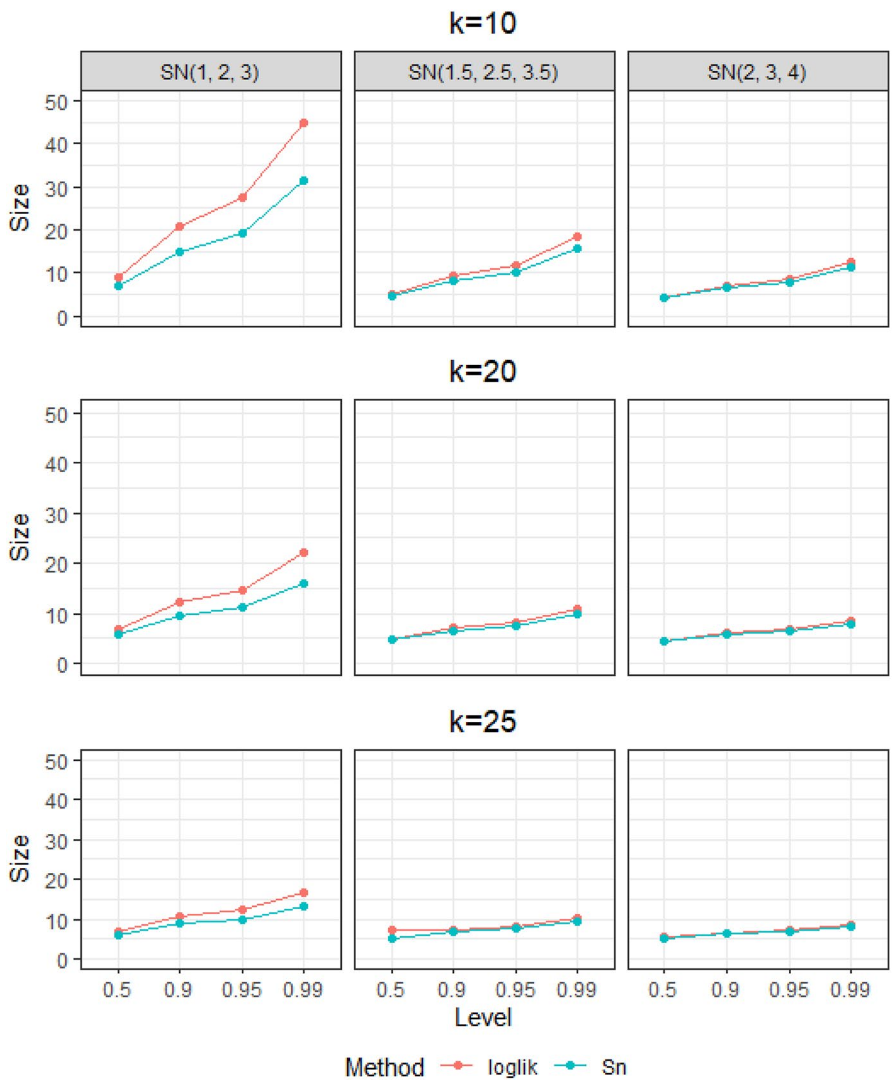

Fig. 2 The confidence set comparison when all three parameters change at various change-point positions, $k = 10, 20, 25$ and the sample size $n = 50$

4 Application

In this section, we consider the stock returns for Brazilian and Chilean markets to apply the proposed method. The stock returns for both countries were recorded weekly from October 31, 1995, to October 31, 2000. These datasets were used in [2, 30]. Instead of using the data directly, we use the stock return ratio as recommended in [30]. In both analyses, we also assume changes occurring in all three parameters simultaneously. The binary segmentation method proposed by [43]

Table 3 The comparisons of coverage probability, $n = 100$

	α	$k = 10$			$k = 20$			$k = 40$			$k = 50$		
		MIC	loglik	S_n	MIC	loglik	S_n	MIC	loglik	S_n	MIC	loglik	S_n
SN(1, 2, 3)	0.50	0.42	0.39	0.43	0.43	0.40	0.42	0.45	0.42	0.44	0.45	0.43	0.44
	0.90	0.68	0.66	0.68	0.69	0.68	0.69	0.70	0.67	0.69	0.70	0.68	0.70
	0.95	0.73	0.72	0.73	0.75	0.74	0.75	0.78	0.75	0.77	0.79	0.76	0.78
	0.99	0.83	0.82	0.85	0.86	0.84	0.86	0.87	0.86	0.86	0.89	0.85	0.88
SN(1.5, 2.5, 3.5)	0.50	0.45	0.44	0.45	0.46	0.44	0.46	0.48	0.47	0.48	0.49	0.47	0.48
	0.90	0.72	0.70	0.72	0.73	0.71	0.72	0.82	0.78	0.82	0.88	0.85	0.88
	0.95	0.77	0.76	0.75	0.77	0.76	0.77	0.88	0.85	0.88	0.91	0.88	0.91
	0.99	0.86	0.85	0.85	0.89	0.82	0.89	0.95	0.94	0.95	0.97	0.96	0.97
SN(2, 3, 4)	0.50	0.49	0.47	0.48	0.49	0.47	0.48	0.49	0.46	0.48	0.52	0.50	0.50
	0.90	0.82	0.80	0.82	0.83	0.85	0.83	0.85	0.83	0.85	0.88	0.86	0.87
	0.95	0.88	0.86	0.88	0.88	0.86	0.86	0.93	0.85	0.81	0.93	0.92	0.92
	0.99	0.96	0.95	0.96	0.96	0.96	0.96	0.97	0.94	0.96	0.98	0.96	0.97

Table 4 The comparisons of average sizes of confidence sets, $n = 100$

	α	$k = 10$			$k = 20$			$k = 40$			$k = 50$		
		MIC	loglik	S_n	MIC	loglik	S_n	MIC	loglik	S_n	MIC	loglik	S_n
SN(1, 2, 3)	0.50	7.02	8.86	6.94	5.95	6.91	5.90	5.61	6.55	5.58	5.90	6.89	5.90
	0.90	14.82	20.88	14.71	9.69	12.22	9.64	8.48	10.56	8.45	8.74	10.74	8.74
	0.95	19.31	27.63	19.16	11.43	14.71	11.37	9.70	12.36	9.68	9.96	12.43	9.96
	0.99	31.71	45.11	31.34	15.98	21.92	15.88	12.71	16.75	12.64	12.99	16.78	12.99
SN(1.5, 2.5, 3.5)	0.50	4.85	5.17	4.75	4.70	4.87	4.69	4.77	4.92	4.75	5.38	7.29	5.20
	0.90	8.35	9.28	8.21	6.61	7.02	6.60	6.43	6.75	6.40	6.82	7.24	6.82
	0.95	10.40	11.60	10.22	7.50	8.08	7.48	7.13	7.64	7.10	7.50	8.02	7.50
	0.99	15.98	18.48	15.73	9.94	10.84	9.92	8.89	9.58	8.87	9.24	10.01	9.24
SN(2, 3, 4)	0.50	4.14	4.38	4.11	4.31	4.41	4.30	4.72	4.67	4.71	5.21	5.46	5.16
	0.90	6.57	7.12	6.53	5.71	5.97	5.70	5.88	5.95	5.88	6.38	6.45	6.38
	0.95	7.88	8.62	7.85	9.07	6.70	6.37	6.39	6.54	6.38	6.90	7.03	6.90
	0.99	11.47	12.68	11.42	8.02	8.42	8.00	7.63	7.93	7.62	8.20	8.47	8.20

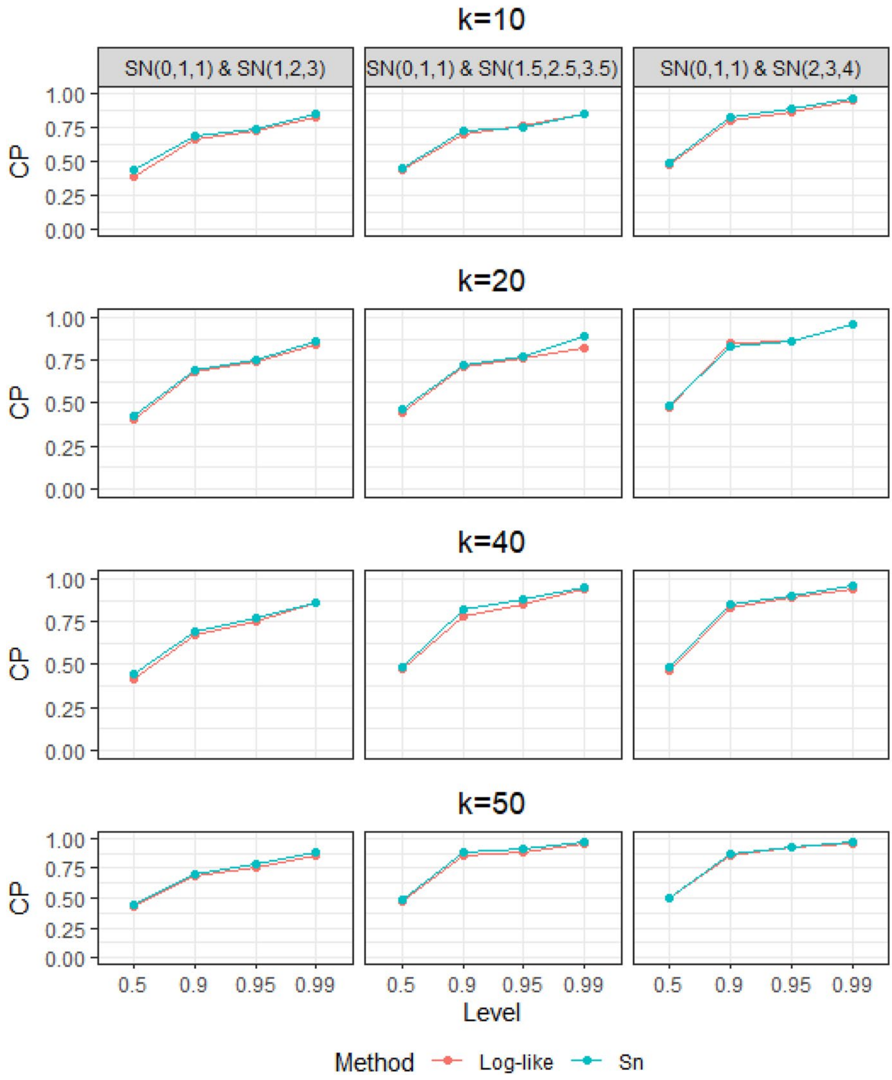

Fig. 3 The coverage probability comparison when all parameters change at various change-point positions, $k = 10, 20, 40, 50$ and the sample size $n = 100$

is used to detect possible multiple changes in the data. The stock return ratio is obtained by the following transformation:

$$R_t = \frac{P_{t+1} - P_t}{P_t},\ t = 1, 2, \cdots, n - 1.$$

The independence of the transformed data can be verified by the Portmanteau test. See [27, 30] for details.

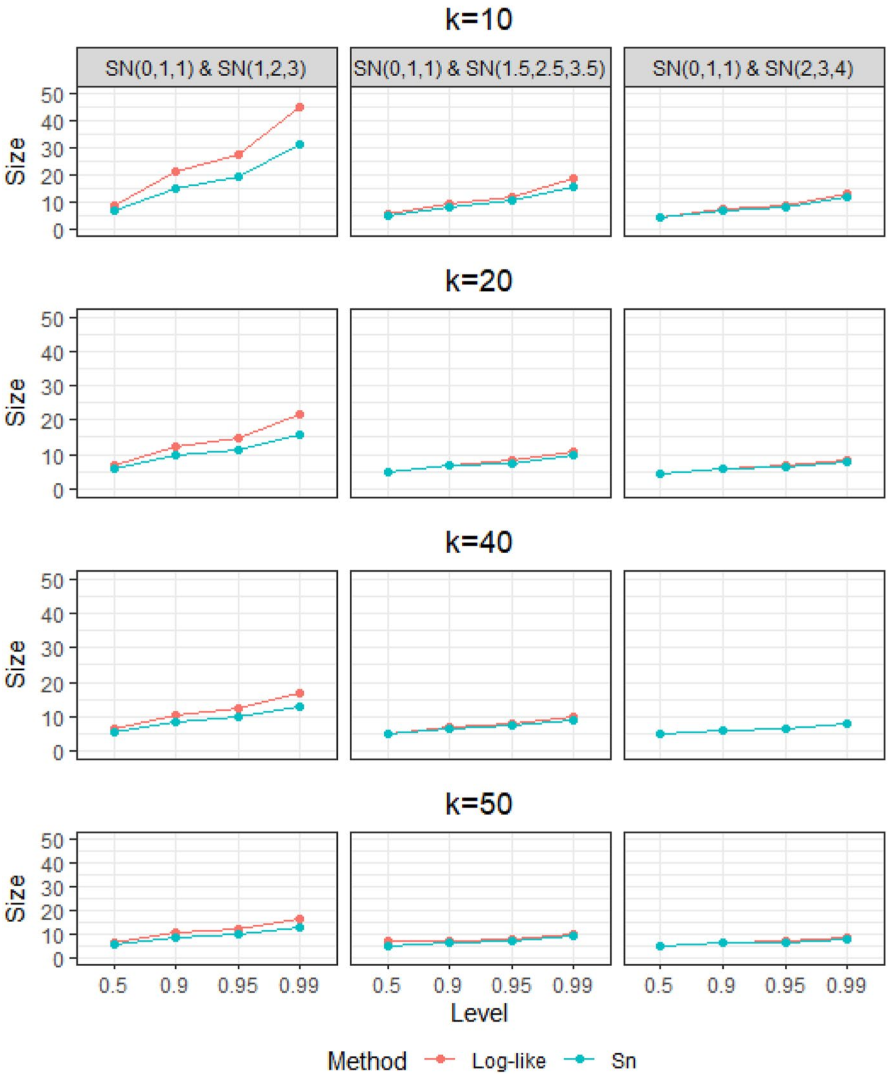

Fig. 4 The confidence sets comparison when all three parameters change at various change-point locations, $k = 10, 20, 40, 50$ and the sample size $n = 100$

4.1 Brazilian Market Return Ratio Data

The Brazilian market return and return ratio data are graphed in Fig. 5.

We apply the proposed approach along with the binary segmentation procedure to detect multiple change points and construct corresponding confidence sets. $\mathrm{MIC}(n) = \mathrm{MIC}(262) = -781.8335 > \min_{1 \le k < n} \mathrm{MIC}(k) = \mathrm{MIC}(87) = -824.4513$ provides the estimation of the change location to be $\hat{k} = 87$. To confirm that it is a statistical significant change instead of being caused by noises, we

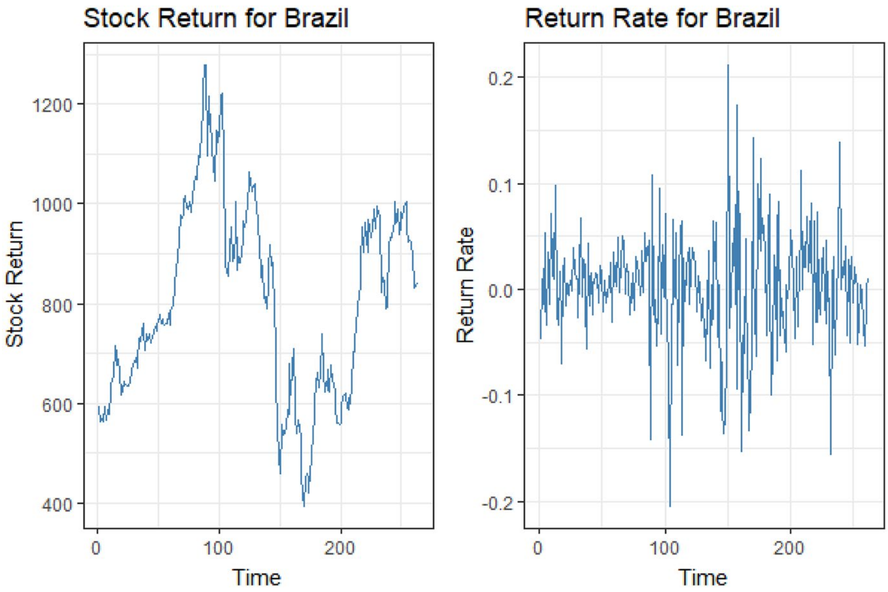

Fig. 5 The weekly stock return and return ratio data for Brazil

calculate the test statistic $S_n = 59.3228$ associated with the critical value being $\chi^2_{0.95,3} = 7.815$ and the p value being 8.202×10^{-13}. Therefore, we confirm that the change occurring at 87^{th} location which corresponds to the change 88^{th} location in the raw data is statistically significant. The maximum likelihood estimates (MLEs) of the parameters before the change point are $(\hat{\mu}_L, \hat{\sigma}_L, \hat{\lambda}_L) = (-0.0079, 0.0328, 0.8732)$, and the MLEs of the parameters are $(\hat{\mu}_R, \hat{\sigma}_R, \hat{\lambda}_R) = (-0.0006, 0.0611, 0.0022)$ after the change. Furthermore, the 95% confidence set for the change point is $\{71, 78, \dots, 90\}$.

We then divide the datasets into two subsequences that are below k (≤ 87) and above k (> 87) and repeat the above detection process to detect changes in these two subsequences. Such an iterative process stops till no further change points being found. We detect all possible change points being $\{19, 88, 144, 170, 240\}$. The confidence curves for all change-point estimates are shown in Figs. 6, 7 and 8. 95% confidence sets are marked by red dashed lines. Comparing to the ones obtained by [30], we detect an additional change at 170th. These change points are shown in Fig. 9. Moreover, as suggested by one of the reviewers, we apply the normal model by Celine et al. [8] to the dataset to detect changes and obtain the change-point set {19, 88, 144, 192, 240}. It has the change point 192 different from our method and the one by Ngunkeng and Ning [30]. However, the violation of normality of the data was verified by Ngunkeng and Ning [30]. Therefore, the model based on the skew normal distribution is more appropriate than the one based on the normal distribution by Celine et al. [8]. Consequently, the results obtained based on the skew normal model are more convincible.

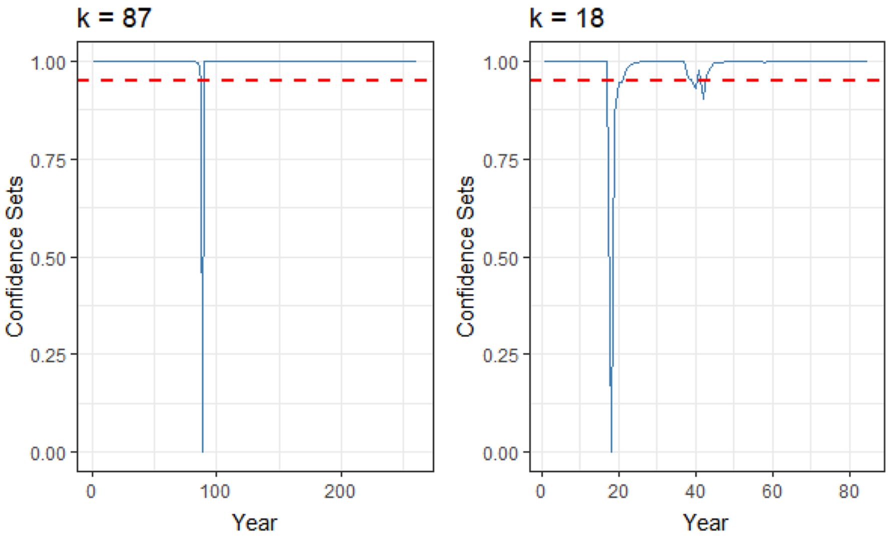

Fig. 6 Left: confidence curve for change point at $\hat{k} = 87$ and right: confidence curve for the first subset below ($k \leq 87$), the $\hat{k} = 18$

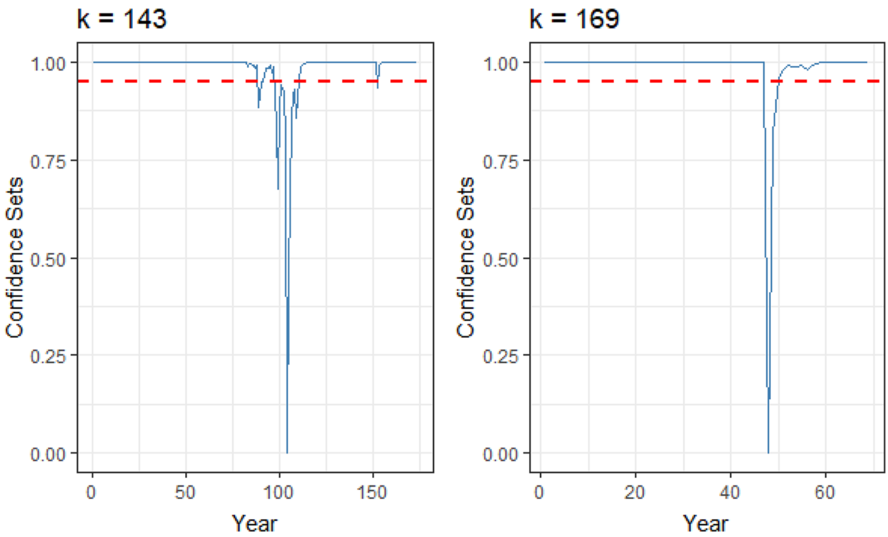

Fig. 7 Left: confidence curve for the second subset ($k \geq 88$) after change point at $\hat{k} = 143$ and right: confidence curve for the subset ($k \geq 144$), the $\hat{k} = 169$

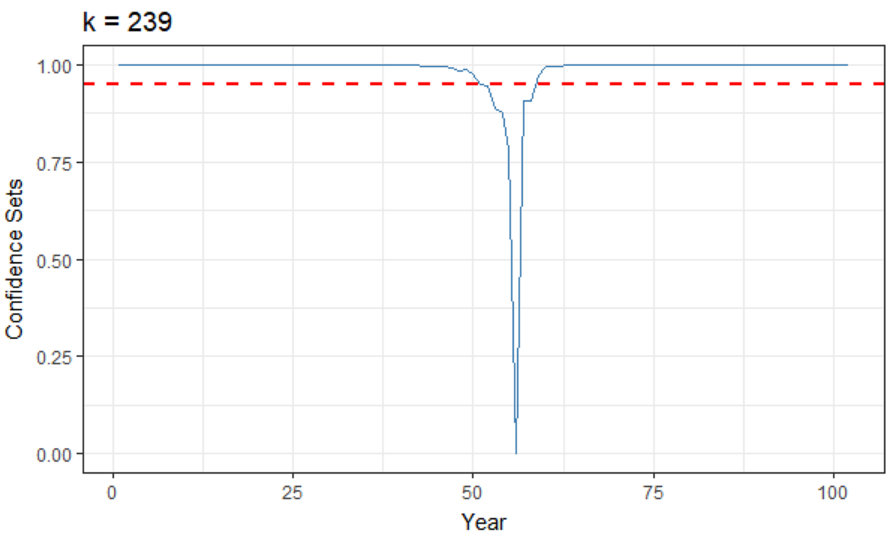

Fig. 8 Confidence curve for the subset ($k \geq 170$), the $\hat{k} = 239$

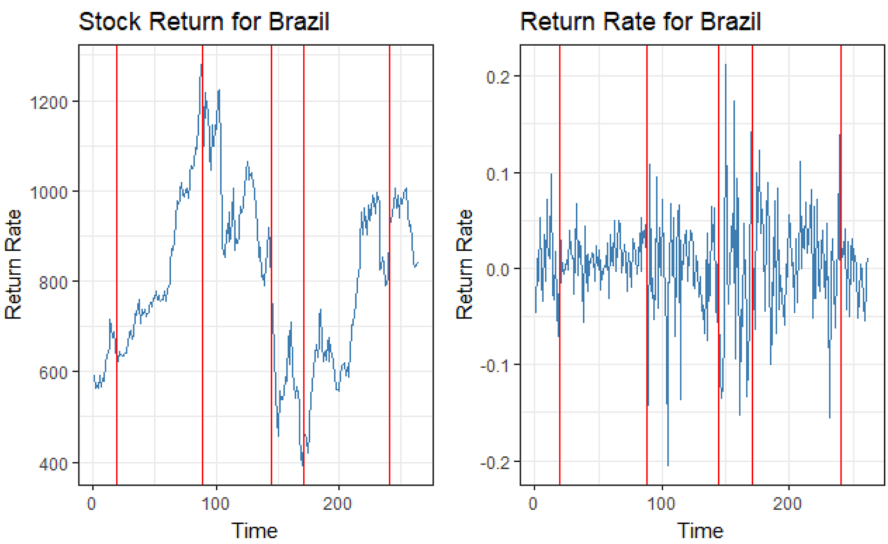

Fig. 9 The weekly stock return data for Brazil with change-point estimates

4.2 Chilean Market Return Ratio Data

The Chilean market return data and return ratio data are graphed in Fig. 10. The $\mathrm{MIC}(n) = \mathrm{MIC}(262) = -1015.536 > \min_{1 \leq k < n} \mathrm{MIC}(k) = \mathrm{MIC}(87) = -1051.643$. The estimated change location is $\hat{k} = 112$. The test statistic $S_n = 52.8011$ with

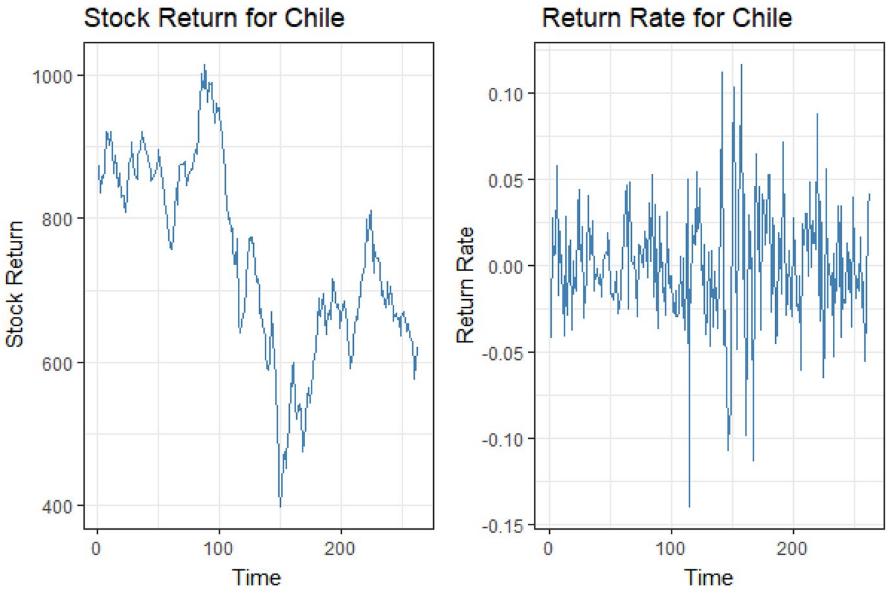

Fig. 10 The weekly stock return and return ratio data for Chile

the critical value $\chi^2_{0.95,3} = 7.815$ and the p value 2.0214×10^{-11} confirms that it is a statistically significant change. The corresponding change point in the data is 113. The MLEs of the parameters are $(\hat{\mu}_L, \hat{\sigma}_L, \hat{\lambda}_L) = (-0.0253, 0.0323, 2.5638)$ and

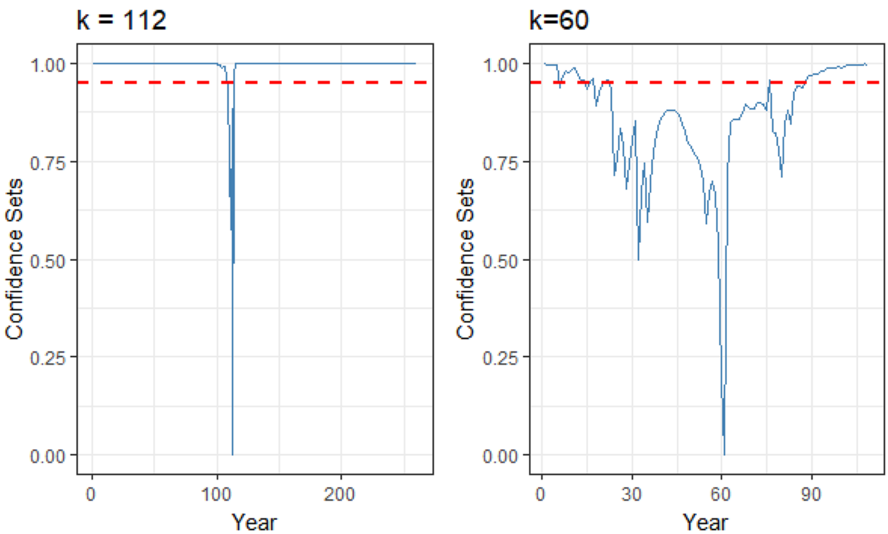

Fig. 11 Left: confidence curve for change point at $\hat{k} = 112$ and right: confidence curve for the first subset below ($k \leq 112$), the change-point estimate $\hat{k} = 60$

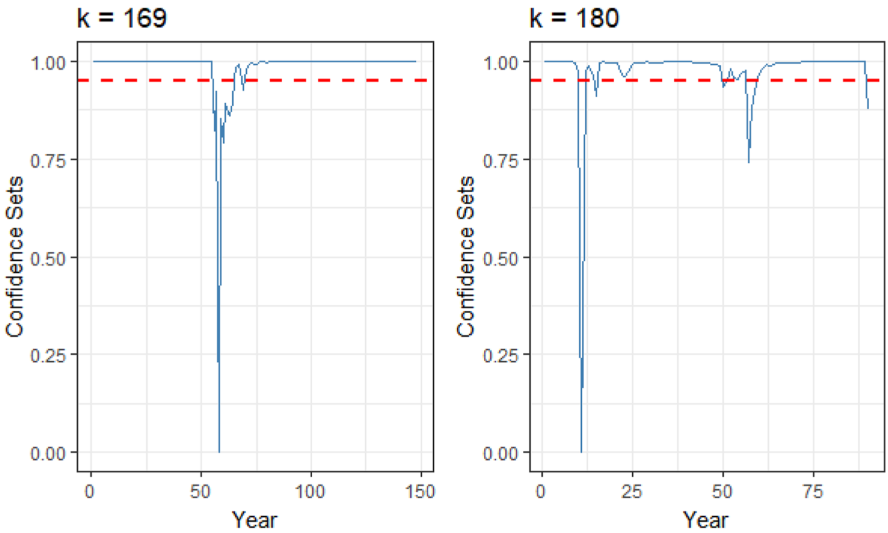

Fig. 12 Left: confidence curve for second subset ($k \geq 113$), the change-point estimate $\hat{k} = 169$ and right: confidence curve for the subsequence after the change point ($k \geq 170$), the change-point estimate $\hat{k} = 180$

$(\hat{\mu}_R, \hat{\sigma}_R, \hat{\lambda}_R) = (-0.0004, 0.0402, 0.0018)$ before and after the change, respectively. With the binary segmentation method, we found all four change points in the Chilean stock data. They are $\{61, 113, 170, 181\}$. Figures 11 and 12 show the confidence curves for all change-point estimates and 95% confidence sets. These change points are graphed in Chilean market data in Fig. 13. The findings from our proposed

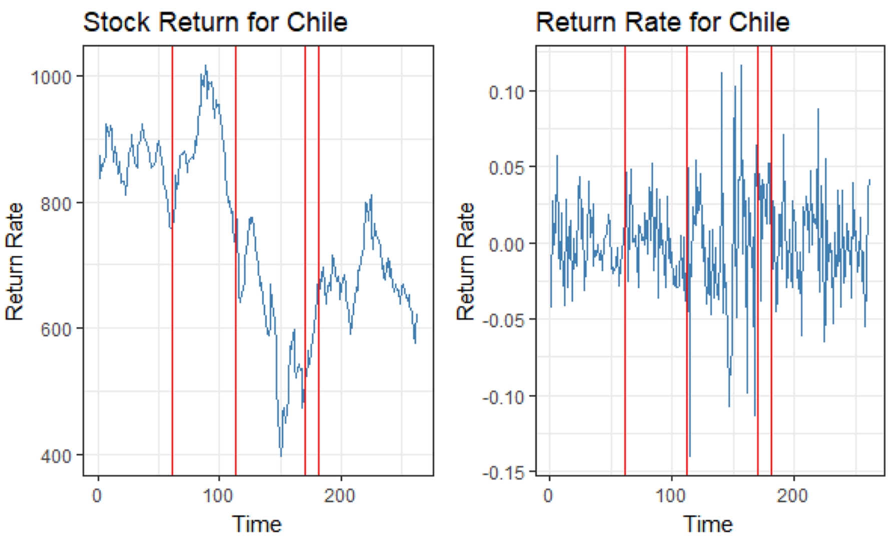

Fig. 13 The weekly stock return data for Chile with change-point estimates

approach are matched with the ones obtained by [30]. Same as the Brazilian data, we also apply the normal model by Celine et al. [8] to detect changes. We obtain the change-point set {7, 113, 170, 259} which has only the change 170 matching our result and the one by Ngunkeng and Ning [30]. The difference is also due to the violation of normality of the data.

5 Conclusion

In this paper, we propose a CD-based procedure incorporating the modified information criterion for a skew normal change-point model. Different from other existing methods, the proposed method can provide confidence sets for the change point for a given nominal level instead of giving the point estimate only. Moreover, the proposed method considers the impact of the location of the change in terms of the model complexity. Consequently, it provides better coverage probability and comparatively smaller average sizes of confidence sets. Simulations are conducted under different scenarios which indicate the advantages of the proposed method. Two stock market data are given to illustrate the detecting procedure by the proposed method.

Acknowledgements The authors would like to thank two anonymous referees for their constructive comments and suggestions which helped to improve this manuscript significantly.

References

1. Abraham B, Wei W (1984) Inferences about the parameters of a time series model with changing variance. Metrika 31:183–194
2. Arellano-Valle RB, Castro LM, Loschi RH (2013) Change point detection in the skew-normal model parameters. Commun Stat Theory Methods 42:603–618
3. Azzalini A (1985) A class of distribution which includes the normal ones. Scand J Stat 12:171–178
4. Azzalini A, Capitanio A (2014) The skew-normal and related families. Cambridge University Press, New York
5. Azzalini A, Dalla Valle A (1996) The multivariate skew-normal distribution. Biometrika 83:715–726
6. Bhattacharya G, Johnson R (1968) Non-parametric tests for shift at an unknown time point. Ann Math Stat 39:1731–1743
7. Cai X, Said K, Ning W (2016) Change-point analysis with bathtub shape for the exponential distribution. J Appl Stat 43:2740–2750
8. Celine C, Gudmund H, Hjort N (2018) Confidence distributions for change-points and regime shifts. J Stat Plan Inference 195:14–34
9. Chen J, Gupta A (1997) Testing and locating variance change points with application to stock prices. J Am Stat Assoc 92:739–747
10. Chen J, Gupta AK (1999) Change point analysis of a Gaussian model. Stat Pap 40:323–333
11. Chen J, Gupta AK (2012) Parametric statistical change point analysis with applications to genetics, medicine, and finance, 2nd edn. Birkhäuser, Boston
12. Chen J, Gupta AK, Pan J (2006) Information criterion and change point problem for regular models. Indian J Stat 68(2):252–282
13. Chernoff H, Zacks S (1964) Estimating the current mean of a normal distribution which is subject to changes in time. Ann Math Stat 35:999–1018
14. Cosörgö M, Horváth L (1997) Limit theorems in change-point analysis. Wiley, New York
15. Cox D (1958) Some problems with statistical inference. Ann Math Stat 29:357–372
16. Davis W (1979) Robust methods for detection of shifts of the innovation variance of a time series. Technometrics 21:313–320
17. Gardner L (1969) On detecting changes in the mean of normal variables. Ann Math Stat 40:116–126

18. Gupta AK, Chen J (1996) Detecting changes of mean in multidimensional normal sequences with application to literature and geology. Comput Stat 11:211–221
19. Hasan A, Ning W, Gupta AK (2014) An information-based approach to the change-point problem of the noncentral skew t distribution with applications to stock market data. Sequent Anal Design Methods Appl 33:458–474
20. Hawkins DM (1977) Testing a sequence of observations for a shift in location. J Am Stat Assoc 72(357):180–186
21. Hawkins DM (1992) Detecting shifts in functions of multivariate location and covariance parameters. J Stat Plan Inference 33:233–244
22. Hinkley DV (1970) Inference about the change-point in a sequence of random variables. Biometrika 57:1–17
23. Hirotsu C, Kuriki S, Hayter AJ (1992) Multiple comparison procedure based on the maximal component of the cumulative chi-squared statistic. Biometrika 79:381–392
24. Horváth L (1993a) The maximum likelihood method of testing changes in the parameters of normal observations. Ann Stat 21(2):671–680
25. Horváth L (1993b) The maximum likelihood method of testing changes in the parameters of normal observations. Ann Stat 21:671–680
26. Hsu D (1977) Tests for variance shifts at an unknown time point. Appl Stat 26(3):279–284
27. Hsu D (1979) Detecting shifts of parameter in gamma sequences with applications to stock price and air traffic flow analysis. J Am Stat Assoc 74:31–40
28. Inclán C (1993) Detection of multiple changes of variance using posterior odds. J Bus Econ Stat 11:189–300
29. Kim H, Siegnumd D (1989) The likelihood ratio test for a change-point in simple linear regression. Biometrika 76(3):409–423
30. Ngunkeng G, Ning W (2014) Information approach for the change-point detection in the skew normal distribution and its applications. Sequent Anal Des Methods Appl 33(4):475–490
31. Page ES (1954) Continue inspection schemes. Biometrika 41:100–135
32. Page ES (1955) A test for a chance in a parameter occurring at an unknown point. Biometrika 42:523–527
33. Said K, Ning W, Tian Y (2019) Modified information criterion for testing changes in skew normal model. Braz J Prob Stat 33:280–300
34. Schwarz G (1978) Estimating the dimension of a model. Ann Stat 6:461–464
35. Schweder T, Hjort N (2002) Confidence and likelihood. Scand J Stat 29(2):309–332
36. Sen A, Srivastava M (1975a) On tests for detecting change in mean. Ann Stat 3:98–108
37. Sen A, Srivastava M (1975b) Some one-sided tests on change in level. Technometrics 17:61–64
38. Shen J, Liu R, Xie M (2018) Prediction with confidence-a general framework for predictive inference. J Stat Plan Inference 195:126–140
39. Singh K, Xie M (2012) CD posterior-combining prior and data through confidence distributions. Contemporary Developments in Bayesian Analysis and Statistical Decision Theory: A Festchrift in Honor of Williams E. Strawderman (D. Fourdrinier et al. eds.). IMS Collection, 8:200–214. IMS, Beachwood, OH
40. Singh K, Xie M, Strawderman WE (2005) Combining information from independent sources through confidence distributions. Ann Stat 33(1):159–183
41. Singh K, Xie M, Strawderman WE (2007) Confidence distribution (cd): distribution estimator of a parameter. In: Lecture Notes-Monograph Series Vol. 54, Complex datasets and inverse problems: tomography, networks and beyond, pp. 132–150
42. Srivastava MS, Worsley KJ (1986) Likelihood ratio tests for a change in the multivariate mean. J Am Stat Assoc 81:199–204
43. Vostrikova LJ (1981) Detecting "disorder" in multidimensional random processes. Soviet Math Doklady 24:55–59
44. Xie M, Singh K (2013) Confidence distribution, the frequentist distribution estimator of a parameter: a review. Int Stat Rev 81(1):2–39