Brazilian Journal of Probability and Statistics
2019, Vol. 33, No. 2, 280–300
https://doi.org/10.1214/17-BJPS388

Modified information criterion for testing changes in skew normal model

Khamis K. Said[a], Wei Ning[b] and Yubin Tian[a]

[a]*Beijing Institute of Technology*
[b]*Bowling Green State University*

Abstract. In this paper, we study the change point problem for the skew normal distribution model from the view of model selection problem. The detection procedure based on the modified information criterion (MIC) for change problem is proposed. Such a procedure has advantage in detecting the changes in early and late stage of a data comparing to the one based on the traditional Schwarz information criterion which is well known as Bayesian information criterion (BIC) by considering the complexity of the models. Due to the difficulty in deriving the analytic asymptotic distribution of the test statistic based on the MIC procedure, the bootstrap simulation is provided to obtain the critical values at the different significance levels. Simulations are conducted to illustrate the comparisons of performance between MIC, BIC and likelihood ratio test (LRT). Such an approach is applied on two stock market data sets to indicate the detection procedure.

1 Introduction

In statistics, a change point is defined as place or time point which the observations before and after that point follow different distributions. The study of the change point problem was dated back to Page (1954, 1955), who first proposed a procedure to detect only one change in a parameter. The identification of change points plays an important role in financial time series analysis, economy, quality control, genome research, signal processing, medical research, statistical calibration, etc. For instance, Chernoff and Zacks (1964), Gardner (1969), Hawkins (1992), studied the testing and estimation of a change in the mean of a normal model. Hsu (1977), Inclán (1993), studied change point problem for the variance in a normal model. Readers are referred to Csörgő and Horváth (1997) and Chen and Gupta (2012) for more details of parametric and nonparametric methods on different types of change point problems. Recently, Zou et al. (2007) proposed a procedure based on the empirical likelihood method by Owen (1988) to detect changes in distributions, and established the asymptotic distribution as one of the gumbel distributions.

In general, a change point problem involves with two consecutive steps (1) testing null hypothesis without changes versus the alternative hypothesis with at least

Key words and phrases. Skew normal distribution, change points, model selection, Bayesian information criterion, modified information criterion, likelihood ratio test.

Received May 2017; accepted November 2017.

one change; (2) estimating the change location or locations if we reject the null hypothesis. Therefore, a change point problem can be treated as the problem of selecting a better one from the models under the null and the alternative hypotheses respectively. The choice of the model under the null hypothesis corresponds to the scenario of no change while the choice of the one under the alternative hypothesis corresponds to the scenario of having at least one change. The use of the information criteria for the model selection has been extensively studied since 1970s. See Akaike (1973), Schwarz (1978) and Hannan and Quinn (1979). To adopt these information criteria to change point problems, there have been fruitful research done in this direction, such as Hirotsu, Kuriki and Hayter (1992), Chen and Gupta (1997), Chen, Gupta and Pan (2006), Ngunkeng and Ning (2014), Hasan, Ning and Gupta (2014) and Cai, Said and Ning (2016), to name a few.

The skew normal (SN) distribution refers to a parametric class of probability distributions that extends the normal distribution by adding an additional shape parameter λ that regulates the skewness of the data. Azzalini (1985) proposed the skew normal distribution and defined the probability density function of the standard skew normal distribution as

$$f(z) = 2\phi(z)\Phi(\lambda z), \tag{1.1}$$

and the general skew normal distribution is then defined as

$$f(x; \mu, \sigma, \lambda) = \frac{2}{\sigma}\phi\left(\frac{x-\mu}{\sigma}\right)\Phi\left(\lambda\frac{x-\mu}{\sigma}\right), \tag{1.2}$$

where $\phi(\cdot)$, $\Phi(\cdot)$ are the p.d.f. and c.d.f. of the standard normal distribution $x \in \Re$, μ is the location, σ is the scale and $\lambda \in \Re$ is the shape parameter.

The behavior of skew normal model has been studied by many authors, to name a few, Henze (1986) provided a probabilistic representation of the skew normal distribution family in terms of a normal random variable and a truncated normal random variable. Azzalini and Dalla Valle (1996) extended the univariate case to the multivariate case. Azzalini and Capitanio (1999) studied further probabilistic properties of the multivariate skew normal distribution with applications to some multivariate statistics problems. Arellano-Valle et al. (2008) and Arellano-Valle, Genton and Loschi (2009) considered shape mixtures in the skew-normal class. They addressed the inference problem in skewed regression models and discussed theoretical issues regarding to Bayesian inference in the skew normal family such as conjugacy and robustness. Ning and Gupta (2012) generalized the univariate extended skew normal distribution family to the matrix variate case. Readers are referred to see Azzalini and Capitanio (2014) for more recent results of the skew normal distribution family.

To the best of our knowledge, only a few work has been done on the change point problem of the skew normal distribution. Arellano-Valle, Castro and Loschi (2013) proposed a Bayesian approach for the detection at most one change in the skew normal distribution family. Ngunkeng and Ning (2014) proposed a testing procedure based on Bayesian information criterion (BIC) and applied on several stock market data sets. Most recently, Said, Ning and Tian (2017) proposed a likelihood ratio testing procedure for the change point problem of the skew normal distribution and established asymptotic properties of the associated test statistic. As Chen, Gupta and Pan (2006) pointed out, when using the traditional information criterion such as AIC and BIC in the context of change point problem, the complexity of the model should be reconsidered due to the lack of the consideration of the contributions by change locations.

The rest of paper is organized as follows. In Section 2, we adopt the modified information criterion (MIC) proposed by Chen, Gupta and Pan (2006) to the change point problem of the skew normal distribution. Simulations are conducted in Section 3 to investigate the performance of the proposed method and the comparison with the one based on BIC. Such a method is applied to Chile and Mexico stock returns to illustrate the detecting procedure. Discussion is provided in Section 5.

2 Changes in skew normal parameters

In this section, we apply the modified information approach (MIC) to detect changes in a skew normal model. The MIC was proposed by Chen, Gupta and Pan (2006) which is the modification of BIC approach by refining the model complexity in the context of change point problems in order to involve with the contributions by change locations.

In general, multiple changes in data will be considered. Vostrikova (1981) proposed the binary segmentation method which can detect multiple changes in several consecutive steps with at most one change in each step. She also showed such a procedure is consistent. With the binary segmentation method, the multiple change problem can always be treated as the single change problem. Therefore, through the rest of the paper, we only develop the testing procedure for a single change.

Let $X_1, \dots, X_n$ be a sequence of independent random variables from a skew normal distribution $\mathrm{SN}(\mu, \sigma, \lambda)$. We are interested in testing the changes in the location μ, scale σ^2 and shape λ parameters simultaneously. Thus we are interesting in testing the null hypothesis

$$H_0: \underbrace{\mu_1 = \mu_2 = \cdots = \mu_n}_{\mu}, \qquad \underbrace{\sigma_1 = \sigma_2 = \cdots = \sigma_n}_{\sigma}, \qquad \underbrace{\lambda_1 = \lambda_2 = \cdots = \lambda_n}_{\lambda}$$

versus the alternative hypothesis

$$\begin{aligned} H_1: &\underbrace{\mu_1 = \mu_2 = \cdots = \mu_k}_{\mu_1} \neq \underbrace{\mu_{k+1} = \mu_{k+2} = \cdots = \mu_n}_{\mu_n}, \\ &\underbrace{\sigma_1 = \sigma_2 = \cdots = \sigma_k}_{\sigma_1} \neq \underbrace{\sigma_{k+1} = \sigma_{k+2} = \cdots = \sigma_n}_{\sigma_n}, \\ &\underbrace{\lambda_1 = \lambda_2 = \cdots = \lambda_k}_{\lambda_1} \neq \underbrace{\lambda_{k+1} = \lambda_{k+2} = \cdots = \lambda_n}_{\lambda_n}. \end{aligned}$$

Then the corresponding likelihood function under H_0 is given as

$$L_{H_0} = 2^n \sigma^{-n} \prod_{i=1}^{n} \phi\Big(\frac{x_i - \mu}{\sigma}\Big) \prod_{i=1}^{n} \Phi\Big(\lambda \frac{x_i - \mu}{\sigma}\Big). \tag{2.1}$$

Consequently, the log-likelihood function is given by

$$\log L_{H_0} = n \ln 2 - n \ln \sigma - \frac{1}{2} \sum_{i=1}^{n} \Big(\frac{x_i - \mu}{\sigma}\Big)^2 + \sum_{i=1}^{n} \ln \Phi\Big(\lambda \frac{x_i - \mu}{\sigma}\Big). \tag{2.2}$$

To find maximum likelihood estimators (MLEs) of μ, σ and λ, we need to solve the following nonlinear equations.

$$\frac{\partial}{\partial \mu}(\log L_{H_0}) = \sum_{i=1}^{n} \Bigg(\frac{(x_i - \mu)}{\sigma^2} - \frac{\lambda}{\sigma} \frac{\phi(\lambda \frac{x_i - \mu}{\sigma})}{\Phi(\lambda \frac{x_i - \mu}{\sigma})} \Bigg) = 0, \tag{2.3}$$

$$\frac{\partial}{\partial \sigma}(\log L_{H_0}) = -\frac{n}{\sigma} + \sum_{i=1}^{n} \Bigg(\frac{(x_i - \mu)^2}{\sigma^3} - \frac{\lambda (x_i - \mu)}{\sigma^2} \frac{\phi(\lambda \frac{x_i - \mu}{\sigma})}{\Phi(\lambda \frac{x_i - \mu}{\sigma})} \Bigg) = 0, \tag{2.4}$$

$$\frac{\partial}{\partial \lambda}(\log L_{H_0}) = \sum_{i=1}^{n} \Bigg(\frac{(x_i - \mu)}{\sigma} \frac{\phi(\lambda \frac{x_i - \mu}{\sigma})}{\Phi(\lambda \frac{x_i - \mu}{\sigma})} \Bigg) = 0. \tag{2.5}$$

Similarly, the log-likelihood function under H_1 is given by

$$\begin{aligned} \log L_{H_1} = &\Bigg\{ k \ln 2 - k \ln \sigma_1 - \frac{1}{2} \sum_{i=1}^{k} \Big(\frac{x_i - \mu_1}{\sigma_1}\Big)^2 + \sum_{i=1}^{k} \Phi\Big(\lambda_1 \frac{x_i - \mu_n}{\sigma_n}\Big) \Bigg\} \\ &+ \Bigg\{ (n-k) \ln 2 - (n-k) \ln \sigma_n \\ &- \frac{1}{2} \sum_{i=k+1}^{n} \Big(\frac{x_i - \mu_n}{\sigma_n}\Big)^2 + \sum_{i=k+1}^{n} \Phi\Big(\lambda_n \frac{x_i - \mu_n}{\sigma_n}\Big) \Bigg\}. \end{aligned}$$

Then, the MLEs of $\mu_1, \sigma_1, \lambda_1$ and $\mu_n, \sigma_n, \lambda_n$ are the solutions of the following nonlinear equations.

$$\frac{\partial}{\partial \mu_1}(\log L_{H_1}) = \sum_{i=1}^{k} \Bigg(\frac{(x_i - \mu_1)}{\sigma_1^2} - \frac{\lambda_1}{\sigma_1} \frac{\phi(\lambda_1 \frac{x_i - \mu_1}{\sigma_1})}{\Phi(\lambda_1 \frac{x_i - \mu_1}{\sigma_1})} \Bigg) = 0, \tag{2.6}$$

$$\begin{aligned} \frac{\partial}{\partial \sigma_1}(\log L_{H_1}) &= -\frac{k}{\sigma_1} + \sum_{i=1}^{k} \Bigg(\frac{(x_i - \mu_1)^2}{\sigma_1^3} - \frac{\lambda_1 (x_i - \mu_1)}{\sigma_1^2} \frac{\phi(\lambda_1 \frac{x_i - \mu_1}{\sigma_1})}{\Phi(\lambda_1 \frac{x_i - \mu_1}{\sigma_1})} \Bigg) \\ &= 0, \end{aligned} \tag{2.7}$$

$$\frac{\partial}{\partial \lambda_1}(\log L_{H_1}) = \sum_{i=1}^{k} \Bigg(\frac{(x_i - \mu_1)}{\sigma_1} \frac{\phi(\lambda_1 \frac{x_i - \mu_1}{\sigma_1})}{\Phi(\lambda_1 \frac{x_i - \mu_1}{\sigma_1})} \Bigg) = 0, \tag{2.8}$$

and

$$\frac{\partial}{\partial \mu_n}(\log L_{H_1}) = \sum_{i=k+1}^{n} \left(\frac{(x_i - \mu_n)}{\sigma_n^2} - \frac{\lambda_n}{\sigma_n} \frac{\phi(\lambda_n \frac{x_i-\mu_n}{\sigma_n})}{\Phi(\lambda_n \frac{x_i-\mu_n}{\sigma_n})} \right) = 0, \tag{2.9}$$

$$\frac{\partial}{\partial \sigma_n}(\log L_{H_1}) = -\frac{n-k}{\sigma_n} + \sum_{i=k+1}^{n} \left(\frac{(x_i - \mu_n)^2}{\sigma_n^3} - \frac{\lambda_1 (x_i - \mu_n)}{\sigma_n^2} \frac{\phi(\lambda_n \frac{x_i-\mu_n}{\sigma_n})}{\Phi(\lambda_n \frac{x_i-\mu_n}{\sigma_n})} \right) = 0, \tag{2.10}$$

$$\frac{\partial}{\partial \lambda_n}(\log L_{H_1}) = \sum_{i=k+1}^{n} \left(\frac{(x_i - \mu_n)}{\sigma_n} \frac{\phi(\lambda_n \frac{x_i-\mu_n}{\sigma_n})}{\Phi(\lambda_n \frac{x_i-\mu_n}{\sigma_n})} \right) = 0. \tag{2.11}$$

The existence of nonlinear functions $\phi(\cdot)$ and $\Phi(\cdot)$ prevents us to obtain the explicit forms for the MLEs $\hat{\mu}, \hat{\sigma}, \hat{\lambda}$ under H_0 and MLEs $\hat{\mu}_1, \hat{\sigma}_1, \hat{\lambda}_1, \hat{\mu}_n, \hat{\sigma}_n, \hat{\lambda}_n$ under H_1, so we employ **R** package **sn** (version 1.4.0, Azzalini (2016)) to obtain the numerical solutions of the MLEs. Then we define the modified information criteria MIC(n) under H_0 and MIC(k) under H_1 respectively, according to Chen, Gupta and Pan (2006) as follows

$$\text{MIC}(n) = -2\ln L_{H_0}(\hat{\mu}, \hat{\sigma}, \hat{\lambda}) + 3\ln(n), \tag{2.12}$$

$$\text{MIC}(k) = -2\ln L_{H_1}(\hat{\mu}_1, \hat{\sigma}_1, \hat{\lambda}_1, \hat{\mu}_n, \hat{\sigma}_n, \hat{\lambda}_n) + \left\{ 6 + \left(\frac{2k}{n} - 1 \right)^2 \right\} \ln(n), \tag{2.13}$$

where k is the possible change location in the range of $1 \le k < n$. Then we accept H_0 if

$$\text{MIC}(n) \le \min_{1 \le k < n} \text{MIC}(k),$$

which indicates there is no change point, and we reject H_0 if

$$\text{MIC}(n) > \min_{1 \le k < n} \text{MIC}(k),$$

which indicates that there exists a change point. Consequently, we can estimate the change point location $\hat{k}$ by

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

Further, we define the test statistic S_n based on MIC(n) and MIC(k) to test the null hypothesis of no change versus the alternative hypothesis of one change as follows,

$$S_n = \text{MIC}(n) - \min_{1 \le k < n} \text{MIC}(k) + 3\ln(n). \tag{2.14}$$

By substituting the equations (2.12) and (2.13) into the equation (2.14), we obtain

$$\begin{aligned}
S_n &= -2\ln L(\hat{\theta}) + 3\ln(n) \\
&\quad - \min_{1\le k<n}\left[-2\ln L(\hat{\theta}_1,\hat{\theta}_n) + \left\{6 + \left(\frac{2k}{n}-1\right)^2\right\}\ln(n)\right] + 3\ln(n) \\
&= -2\ln L(\hat{\theta}) + 3\ln(n) \\
&\quad - \min_{1\le k<n}\left[-2\ln L(\hat{\theta}_1,\hat{\theta}_n) + 6\ln(n) - \left(\frac{2k}{n}-1\right)^2\ln(n)\right] + 3\ln(n) \\
&= -2\ln L(\hat{\theta}) - \min_{1\le k<n}\left[-2\ln L(\hat{\theta}_1,\hat{\theta}_n) + \left(\frac{2k}{n}-1\right)^2\ln(n)\right],
\end{aligned}$$

where $\hat{\theta} = (\hat{\mu},\hat{\sigma},\hat{\lambda})$, $\hat{\theta}_1 = (\hat{\mu}_1,\hat{\sigma}_1,\hat{\lambda}_1)$ and $\hat{\theta}_n = (\hat{\mu}_n,\hat{\sigma}_n,\hat{\lambda}_n)$. We reject the null hypothesis for a sufficient large value of S_n. The standardization term $3\ln(n)$ removes the constant term in the difference of MIC(n) and MIC(k).

3 Simulation results

In this section, we investigate the critical values and performance of the proposed test in terms of powers through simulations due to the difficulty in deriving the analytic properties of S_n.

3.1 Critical values

In our simulation study, we set up the null distribution to be SN(2, 2, 1) and choose sample sizes $n = 50$, 100, 150, 200 and 300 with significance levels $\alpha = 0.01, 0.05$ and 0.1. We also construct the test statistic T_n based on the classical Bayesian information criterion (BIC) which is defined as below to make a comparison.

$$T_n = \mathrm{BIC}(n) - \min_{1\le k<n}\mathrm{BIC}(k) + 3\log n,$$

where BIC(n) under H_0 and BIC(k) under H_1 are given by

$$\begin{aligned}
\mathrm{BIC}(n) &= -2\ln L_{H_0}(\hat{\mu},\hat{\sigma},\hat{\lambda}) + 3\ln(n), \\
\mathrm{BIC}(k) &= -2\ln L_{H_1}(\hat{\mu}_1,\hat{\sigma}_1,\hat{\lambda}_1,\hat{\mu}_n,\hat{\sigma}_n,\hat{\lambda}_n) + 6\log n.
\end{aligned}$$

The similar idea based BIC has also been considered by Ngunkeng and Ning (2014) to detect multiple changes in a skew normal distribution. We note here that the only difference between S_n and T_n is reflected in the difference between MIC(k) in (2.13) and BIC(k). The penalty term in the former one considers the contribution of the change location k associated with the complexity of the model, while the latter one does not. In order to make a fair power comparison between

S_n and T_n, we simulate the critical values for both test statistics under the same null distributions with the same sample sizes for given significance levels.

We would like to make some notes on the simulations to obtain critical values. To simulate the critical values of test statistics under the null hypothesis, one way is to conduct certain number of simulations for a given null distribution and calculate test statistic values S_n and T_n for each simulation. Then the critical values correspond to percentiles in sorted values calculated from the simulations. Another approach is bootstrap method by resampling certain number of bootstrap samples from a generated null distributional sample with replacement, then the percentiles of sorted test statistics values from the bootstrap samples are critical values for given significance levels.

When using the bootstrap method to obtain simulated critical values of a test statistic, we need to ensure that the bootstrap samples are resampled from a data under the null distribution. In simulations, this is not an issue because the null distribution has been determined before resampling. Therefore, it is known to satisfy H_0 which can be used to generate a sample. Thus, in simulations, both approaches will obtain similar critical values. However, for a real data, it would be an issue for bootstrap method since we do not know whether the data satisfies H_0 or H_1. Therefore, we can not perform resampling directly on the data. The following strategy will be taken. We first assume the data satisfying H_0, which indicates it should be fitted by a single skew normal distribution, say, $\mathrm{SN}_0 = \mathrm{SN}(\hat{\mu}, \hat{\sigma}, \hat{\lambda})$, where $\hat{\mu}, \hat{\sigma}$ and $\hat{\lambda}$ can be obtained by R package **sn** (Azzalini (2016)). Then we generate a random sample based on SN_0 denoted by $x_1, x_2, \dots, x_n$. Then B bootstrap samples are drawn from this generated sample with replacement, denoted by $y_1^{(i)}, y_2^{(i)}, \dots, y_n^{(i)}, i = 1, 2, \dots, B$. For each bootstrap sample, we calculate S_n denoted by $S_n^{(i)}, i = 1, 2, \dots, B$. Thus, the p-value can be approximated as follows

$$p\text{-value} = \frac{1}{B} \sum_{i=1}^{B} I\big(S_n^{(i)} \geq S_n^{(*)}\big),$$

where $I(\cdot)$ is the indicator function and $S_n^{(*)}$ is the value of S_n calculated from the original real data. The following Table 1 and Table 2 list critical values of S_n and T_n obtained from the first approach.

3.2 Power comparison

In this section, we conduction simulations under different scenarios to investigate the performance of test procedures based S_n and T_n in terms of power. Furthermore, we also compare the power of the likelihood ratio test (LRT) proposed by Said, Ning and Tian (2017) according to the suggestion by one referee. We set the distribution following $\mathrm{SN}(\mu_1, \sigma_1, \lambda_1)$ to be $\mathrm{SN}(2, 2, 1)$ before the change and $\mathrm{SN}(\mu_n, \sigma_n, \lambda_n)$ after the change with the parameter $\boldsymbol{\theta_n} = (\mu_n, \sigma_n, \lambda_n)$, where $\boldsymbol{\theta_n} = (3, 3, 0)$, $(2.5, 2.5, 2)$, $(3, 3, 2)$ and $(1.5, 1.5, 1.5)$ with the sample size

Table 1 *Approximate critical values for MIC under different parameters*

		α		
n	SN(·)	0.1	0.05	0.01
50	(2, 2, 1)	13.8599	16.0408	20.9073
	(2, 1, 3)	13.7189	15.9038	20.3248
	(3, 3, 2)	15.0880	17.2305	21.3023
100	(2, 2, 1)	14.2453	16.3676	20.4829
	(2, 1, 3)	14.6955	16.9056	21.8547
	(3, 3, 2)	14.0461	16.3581	19.9698
150	(2, 2, 1)	13.4338	15.7396	20.5235
	(2, 1, 3)	13.3606	15.8911	19.0863
	(3, 3, 2)	13.2416	15.9811	21.1312
200	(2, 2, 1)	12.3516	14.3856	19.0478
	(2, 1, 3)	13.4566	15.3791	22.2769
	(3, 3, 2)	12.7815	15.1346	20.7744
300	(2, 2, 1)	13.4768	15.1166	19.6444
	(2, 1, 3)	12.3190	14.0795	18.2067
	(3, 3, 2)	12.5156	14.1769	17.8763

Table 2 *Approximate critical values for BIC under different parameters*

		α		
n	SN(·)	0.1	0.05	0.01
50	(2, 2, 1)	15.909	17.896	22.164
	(2, 1, 3)	16.243	18.813	23.722
	(3, 3, 2)	16.309	18.956	23.325
100	(2, 2, 1)	15.202	17.159	21.294
	(2, 1, 3)	15.961	17.792	22.577
	(3, 3, 2)	15.791	17.355	21.207
150	(2, 2, 1)	15.060	17.103	20.326
	(2, 1, 3)	15.504	17.725	21.436
	(3, 3, 2)	15.186	17.268	22.784
200	(2, 2, 1)	14.866	16.928	21.565
	(2, 1, 3)	15.767	17.745	22.867
	(3, 3, 2)	15.437	17.138	22.234
300	(2, 2, 1)	14.957	16.581	21.685
	(2, 1, 3)	15.767	17.502	21.397
	(3, 3, 2)	15.479	17.615	20.991

$n = 50$, 100 and 150. We also set up changes occurring approximately at the beginning ($\frac{1}{4}$th), at the center ($\frac{1}{2}$th) and at the end ($\frac{3}{4}$th) of the sample size n. The LRT statistic is given by Said, Ning and Tian (2017) as follows.

$$Z_n = \max_{1\le k<n}\left\{-2\ln\left(\frac{\sup_{\mu,\sigma,\lambda} L_{H_0}(x_i;\mu,\sigma,\lambda)}{\sup_{\mu_1,\sigma_1,\lambda_1,\mu_n,\sigma_n,\lambda_n} L_{H_1}(x_i;\mu_1,\sigma_1,\lambda_1,\mu_n,\sigma_n,\lambda_n)}\right)\right\}.$$

The results are listed in Tables 3, 4 and 5.

From the comparison, we can observe that the proposed MIC procedure is very competitive comparing to the BIC and the LRT. In general, we observe that the powers of all three tests increase as the increase of the sample size. For example,

Table 3 *Power comparison between MIC, BIC and LRT for* $\alpha = 0.1$

				$(\mu_n, \sigma_n, \lambda_n)$			
n	k	$(\mu_1, \sigma_1, \lambda_1)$	Model	(3, 3, 0)	$(\frac{5}{2}, \frac{5}{2}, 2)$	(3, 3, 2)	$(\frac{3}{2}, \frac{3}{2}, \frac{3}{2})$
50	10	(2, 2, 1)	MIC	0.204	0.184	0.438	0.261
		(2, 2, 1)	BIC	0.173	0.089	0.265	0.289
		(2, 2, 1)	LRT	0.320	0.293	0.575	0.277
	25	(2, 2, 1)	MIC	0.439	0.334	0.711	0.395
		(2, 2, 1)	BIC	0.267	0.288	0.656	0.336
		(2, 2, 1)	LRT	0.555	0.430	0.806	0.370
	40	(2, 2, 1)	MIC	0.319	0.185	0.233	0.093
		(2, 2, 1)	BIC	0.276	0.148	0.435	0.058
		(2, 2, 1)	LRT	0.383	0.267	0.560	0.258
100	25	(2, 2, 1)	MIC	0.731	0.476	0.882	0.407
		(2, 2, 1)	BIC	0.580	0.481	0.863	0.235
		(2, 2, 1)	LRT	0.639	0.528	0.922	0.514
	50	(2, 2, 1)	MIC	0.861	0.530	0.991	0.641
		(2, 2, 1)	BIC	0.704	0.423	0.975	0.535
		(2, 2, 1)	LRT	0.849	0.645	0.976	0.620
	75	(2, 2, 1)	MIC	0.658	0.380	0.924	0.502
		(2, 2, 1)	BIC	0.510	0.368	0.909	0.355
		(2, 2, 1)	LRT	0.735	0.509	0.935	0.474
150	35	(2, 2, 1)	MIC	0.809	0.717	0.992	0.652
		(2, 2, 1)	BIC	0.803	0.610	0.983	0.617
		(2, 2, 1)	LRT	0.836	0.728	0.984	0.678
	75	(2, 2, 1)	MIC	0.983	0.914	1.000	0.881
		(2, 2, 1)	BIC	0.928	0.857	0.999	0.849
		(2, 2, 1)	LRT	0.973	0.874	0.996	0.824
	110	(2, 2, 1)	MIC	0.935	0.724	0.998	0.695
		(2, 2, 1)	BIC	0.926	0.711	0.993	0.651
		(2, 2, 1)	LRT	0.924	0.765	0.997	0.654

Table 4 *Power comparison between MIC, BIC and LRT for* $\alpha = 0.05$

				$(\mu_n, \sigma_n, \lambda_n)$			
n	k	$(\mu_1, \sigma_1, \lambda_1)$	Model	(3, 3, 0)	$(\frac{5}{2}, \frac{5}{2}, 2)$	(3, 3, 2)	$(\frac{3}{2}, \frac{3}{2}, \frac{3}{2})$
50	10	(2, 2, 1)	MIC	0.126	0.089	0.307	0.173
		(2, 2, 1)	BIC	0.081	0.040	0.151	0.160
		(2, 2, 1)	LRT	0.193	0.186	0.430	0.183
	25	(2, 2, 1)	MIC	0.328	0.221	0.603	0.216
		(2, 2, 1)	BIC	0.161	0.158	0.486	0.186
		(2, 2, 1)	LRT	0.403	0.285	0.688	0.234
	40	(2, 2, 1)	MIC	0.187	0.081	0.114	0.034
		(2, 2, 1)	BIC	0.175	0.077	0.312	0.017
		(2, 2, 1)	LRT	0.268	0.148	0.433	0.132
100	25	(2, 2, 1)	MIC	0.543	0.373	0.784	0.278
		(2, 2, 1)	BIC	0.426	0.358	0.820	0.141
		(2, 2, 1)	LRT	0.498	0.414	0.865	0.392
	50	(2, 2, 1)	MIC	0.758	0.373	0.983	0.525
		(2, 2, 1)	BIC	0.598	0.518	0.929	0.404
		(2, 2, 1)	LRT	0.776	0.533	0.958	0.497
	75	(2, 2, 1)	MIC	0.535	0.283	0.851	0.376
		(2, 2, 1)	BIC	0.392	0.252	0.856	0.240
		(2, 2, 1)	LRT	0.653	0.386	0.899	0.343
150	35	(2, 2, 1)	MIC	0.693	0.627	0.979	0.561
		(2, 2, 1)	BIC	0.713	0.445	0.959	0.596
		(2, 2, 1)	LRT	0.740	0.613	0.973	0.591
	75	(2, 2, 1)	MIC	0.949	0.841	1.000	0.771
		(2, 2, 1)	BIC	0.875	0.759	0.996	0.751
		(2, 2, 1)	LRT	0.942	0.818	0.995	0.726
	110	(2, 2, 1)	MIC	0.888	0.590	0.995	0.560
		(2, 2, 1)	BIC	0.860	0.532	0.987	0.542
		(2, 2, 1)	LRT	0.893	0.647	0.993	0.531

when n increases from 50 to 150 and the true change location is at the beginning of the sample with $\alpha = 0.05$, the power of MIC, BIC and LRT with $\theta_n = (3, 3, 2)$ increase from 0.307, 0.151, 0.430 to 0.979, 0.959, 0.973, respectively. Furthermore, in general the MIC procedure outperforms the BIC procedure with various change locations. For example, in Table 4, when $n = 50, 100, 150$ and the true change occurs at the beginning of the sample size, with $\theta_n = (2.5, 2.5, 2)$ the powers of MIC are 0.089, 0.373, 0.627 which are higher than 0.040, 0.358, 0.445 of the BIC, respectively. Similarly for the true change occurs at the end of sample size, that is, the performance of the MIC is better than that of the BIC. It indicates the advantage of the MIC over the BIC by considering the model complexity associated with the change location as described in previous section.

Table 5 *Power comparison between MIC, BIC and LRT for $\alpha = 0.01$*

n	k	$(\mu_1, \sigma_1, \lambda_1)$	Model	$(\mu_n, \sigma_n, \lambda_n)$			
				(3, 3, 0)	$(\frac{5}{2}, \frac{5}{2}, 2)$	(3, 3, 2)	$(\frac{3}{2}, \frac{3}{2}, \frac{3}{2})$
50	10	(2, 2, 1)	MIC	0.030	0.035	0.152	0.060
		(2, 2, 1)	BIC	0.019	0.009	0.035	0.038
		(2, 2, 1)	LRT	0.056	0.057	0.190	0.065
	25	(2, 2, 1)	MIC	0.135	0.087	0.393	0.088
		(2, 2, 1)	BIC	0.070	0.054	0.263	0.048
		(2, 2, 1)	LRT	0.182	0.104	0.430	0.086
	40	(2, 2, 1)	MIC	0.050	0.026	0.047	0.007
		(2, 2, 1)	BIC	0.030	0.042	0.118	0.002
		(2, 2, 1)	LRT	0.118	0.052	0.221	0.038
100	25	(2, 2, 1)	MIC	0.231	0.130	0.545	0.080
		(2, 2, 1)	BIC	0.217	0.181	0.625	0.054
		(2, 2, 1)	LRT	0.264	0.220	0.720	0.208
	50	(2, 2, 1)	MIC	0.512	0.203	0.918	0.321
		(2, 2, 1)	BIC	0.269	0.302	0.768	0.156
		(2, 2, 1)	LRT	0.558	0.308	0.897	0.278
	75	(2, 2, 1)	MIC	0.219	0.106	0.685	0.154
		(2, 2, 1)	BIC	0.262	0.063	0.733	0.066
		(2, 2, 1)	LRT	0.469	0.204	0.778	0.138
150	35	(2, 2, 1)	MIC	0.445	0.474	0.883	0.373
		(2, 2, 1)	BIC	0.342	0.168	0.899	0.385
		(2, 2, 1)	LRT	0.462	0.386	0.922	0.395
	75	(2, 2, 1)	MIC	0.827	0.647	0.987	0.492
		(2, 2, 1)	BIC	0.753	0.490	0.986	0.407
		(2, 2, 1)	LRT	0.852	0.603	0.991	0.500
	110	(2, 2, 1)	MIC	0.725	0.344	0.970	0.262
		(2, 2, 1)	BIC	0.680	0.408	0.947	0.296
		(2, 2, 1)	LRT	0.779	0.414	0.968	0.285

3.3 Consistency of the estimator $\hat{k}$

We also investigate the consistency of the estimator $\hat{k}$ of the true change location k through a numerical study. The results are listed in Table 6. 1000 simulations have been conducted under different sample sizes $n = 50, 100, 150, 200$ and $n = 300$ with true change location k at $n/2$ and $n/4$. The parameter is set to be $\boldsymbol{\theta}_1 = (2, 2, 1)$ before the change and to be $\boldsymbol{\theta}_n = (-1, 1.5, 1.5)$ after the change. Furthermore, we compare the bias and mean square error (MSE) of the $\hat{k}$ for the MIC and the BIC. The results are listed in Table 6. Generally, we have been observed that, the MSE of MIC are smaller than the MSE of BIC, and also we observed that in most cases the bias of MIC is smaller than the bias of BIC.

Table 6 *The consistency of change location estimator $\hat{k}$*

δ	n	k	$P(\lvert\hat{k}-k\rvert \le \delta)$		Bias($\hat{k}$)		MSE($\hat{k}$)	
			MIC	BIC	MIC	BIC	MIC	BIC
1	50	12	0.851	0.854	0.205	0.182	0.205	0.182
		25	0.870	0.860	0.187	0.194	0.187	0.194
	100	25	0.883	0.879	0.196	0.199	0.196	0.199
		50	0.887	0.886	0.206	0.205	0.206	0.205
	150	37	0.914	0.887	0.201	0.185	0.201	0.185
		75	0.910	0.909	0.173	0.173	0.173	0.173
	200	50	0.895	0.910	0.191	0.199	0.191	0.199
		100	0.913	0.913	0.210	0.210	0.210	0.210
	300	75	0.892	0.904	0.169	0.196	0.169	0.196
		150	0.910	0.910	0.223	0.223	0.223	0.223
2	50	12	0.922	0.923	0.347	0.320	0.489	0.458
		25	0.940	0.930	0.327	0.334	0.467	0.474
	100	25	0.956	0.952	0.342	0.345	0.488	0.491
		50	0.957	0.955	0.346	0.343	0.486	0.481
	150	37	0.972	0.956	0.317	0.323	0.433	0.461
		75	0.969	0.969	0.291	0.293	0.409	0.413
	200	50	0.966	0.962	0.333	0.303	0.475	0.407
		100	0.961	0.961	0.306	0.306	0.402	0.402
	300	75	0.964	0.966	0.303	0.320	0.457	0.444
		150	0.975	0.975	0.353	0.353	0.483	0.483
3	50	12	0.962	0.965	0.467	0.446	0.849	0.836
		25	0.966	0.961	0.405	0.427	0.701	0.753
	100	25	0.973	0.976	0.393	0.417	0.641	0.707
		50	0.979	0.979	0.412	0.415	0.684	0.697
	150	37	0.985	0.981	0.356	0.398	0.550	0.686
		75	0.988	0.988	0.348	0.350	0.580	0.584
	200	50	0.988	0.989	0.399	0.384	0.673	0.650
		100	0.987	0.987	0.384	0.384	0.636	0.636
	300	75	0.985	0.981	0.376	0.365	0.646	0.579
		150	0.989	0.989	0.395	0.395	0.609	0.609

It is noticed that in Table 5, MSE($\hat{k}$) and Bias($\hat{k}$) are exactly same for $\delta = 1$. The reason is that when the estimated change location $\hat{k}$ satisfied the condition $|\hat{k} - k| \le 1$, the value of $\hat{k} - k$ is either $0, 1$ or -1. Therefore, $|\hat{k} - k|$ is equal to $(\hat{k} - k)^2$. Consequently, MSE and Bias are same for $\delta = 1$.

4 Applications

We apply our method to Chile and Mexico stock market datasets to detect possible changes. The stock returns for both countries were recorded weekly from October 31, 1995, to October 31, 2000. Both data have previously been analyzed by Arellano-Valle, Castro and Loschi (2013) and Ngunkeng and Ning (2014) by applying a Bayesian method and a method based on Bayesian information criterion (BIC) approach respectively, for skew normal models.

Let F_t be the stock return index values at week t. In general, we study the stock return rates instead of stock returns directly which is defined as

$$R_t = \frac{F_{t+1} - F_t}{F_t}, \qquad t = 1, 2, \dots, n. \tag{4.1}$$

To test the independence of the transformed data, Hsu (1979) proposed several methods to check independence for such a transformed data set. Here we use the Portmanteau test given by

$$Q_m = n \sum_{i=1}^{m} r_i^2, \tag{4.2}$$

where r_i is the autocorrelation coefficient (ACF) at lag i and m is the lag up to which the auto-correlation coefficient function is considered.

4.1 Chilean stock market

For the Chile stock return rates, we obtain,

$$Q_{24} = 261 \times \sum_{i=1}^{24} r_i^2 = 261 \times 0.1210905 = 31.725 < \chi^2_{0.95}(24) = 36.415.$$

Hence, we fail to reject the null hypothesis of independence. Therefore, the R_t series for Chilean stock return rates are independent. Left graph in Figure 1 shows the ACF of the R_t series data and the right graph in Figure 1 shows the normal Q–Q plot of R_t series which indicates that the normality assumption fails. Applying normality tests such as Shapiro–Wilk test also concludes the validity of normality is violated.

We apply the proposed MIC procedure associated with the test statistic S_n to this data to test the hypotheses in Section 2. Figure 2 below shows the stock return index and stock return rate for the Chilean stock market data with possible change locations identified by the proposed procedure and the corresponding MIC values is shown in Figure 3.

The binary segmentation method is implemented in the detecting procedure to detect all possible changes in the data.

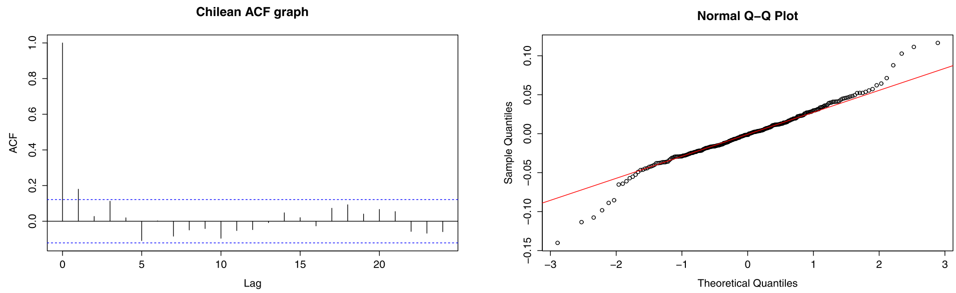

Figure 1 *Left*: *The graph of ACF values*. *Right*: *The normal Q–Q plot of* R_t.

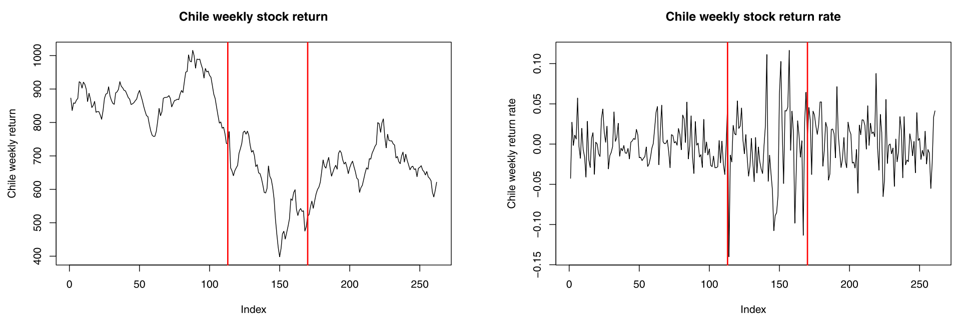

Figure 2 *Left*: *The graph for stock return index*. *Right*: *The graph for stock return rate with change locations for the Chilean data set*.

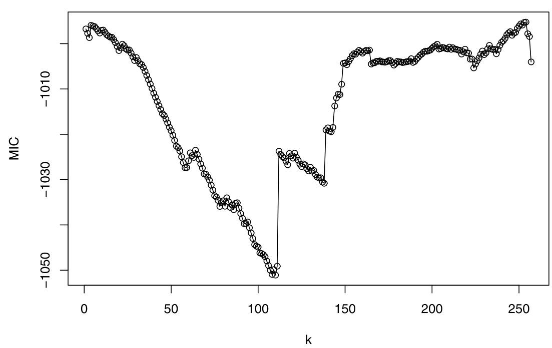

Figure 3 *The graph of MIC values for Chilean stock return rate*.

- Consider the sequence R_t from 1 to 261. The test statistic $\text{MIC}(261) = -1016.9941$ and $\min_{1 \le k < 261} \text{MIC}(k) = -1051.0883$ and hence the statistic $S_n = \text{MIC}(261) - \min_{1 \le k < 261} \text{MIC}(k) + 3\log(261) = 50.7877$. With the discussion in Section 3.1 on the computation of critical values, the bootstrap method proposed in Section 3.1 is used here. We obtain the approximated p-value $<$

0.001 with $B = 2000$. Therefore, we conclude that there is a change in the data and the estimated change location $\hat{k} = \arg\min_{1\le k<261} \text{MIC}(k) = 112$. It corresponds to the change location 113th position in original $\{F_t\}$ series with the associated value 736.133 of the stock return index.

- Then we consider the subsequence R_t from 1 to 112. Similar as the previous step, the test statistic S_n is calculated as 7.824 with the approximated p-value to be 0.5105 under $B = 2000$ bootstrap samples. Thus we accept null hypothesis and conclude there is no change in this subsequence.
- Meanwhile, we consider another subsequence R_t from 113 to 261. We calculate the value of test statistic $S_n = 21.956$ with the approximated p-value $= 0.003$ which leads us to reject the null hypothesis and find another change in R_t series at 169th position. That is, the change occurs at 170th position in the original F_t series corresponding 519.747 of the stock return index.
- With the binary segmentation method, we repeat the above detecting process for 170 to 261 and for 113 to 170 and found out that no more changes exist in both subsequences.

In summary, in the Chilean stock return data set, we are able to locate two change locations which are 113th and 170th positions corresponding to 26th December, 1997 and 29th January, 1999, respectively. These changes may be caused by the result of the 1997 Asian financial crises which reached its climax by mini crash in the October 27, 1997 and may also be caused partly by 1998 Russian financial crises, which led to devaluation of the ruble and its government suspension on foreign creditor payments.

Arellano-Valle, Castro and Loschi (2013) detected a single change in the same data by a Bayesian approach and concluded that the change approximately happened during the first week of February in 1998 but their method can not provide the estimates of change point locations. Ngunkeng and Ning (2014) proposed a detection procedure based on BIC which detected two change locations which are January 29, 1999 and December 26, 1997.

We also apply the likelihood ratio test (LRT) procedure proposed by Said, Ning and Tian (2017) to detect possible multiple changes in the data set. We reject the null hypothesis of test statistic for large value of Z_n, where

$$Z_n = \max_{1\le k<n}\left\{-2\ln\left(\frac{\sup_{\mu,\sigma,\lambda} L_{H_0}(x_i;\mu,\sigma,\lambda)}{\sup_{\mu_1,\sigma_1,\lambda_1,\mu_n,\sigma_n,\lambda_n} L_{H_1}(x_i;\mu_1,\sigma_1,\lambda_1,\mu_n,\sigma_n,\lambda_n)}\right)\right\}.$$

The calculated test statistic $Z_n = 50.8996 > C_{0.05,261} = 16.842$ leads to reject the null hypothesis associated with the estimated change location at 113th position. With the binary segmentation method, we repeat the same procedure for subsequence from 114 to 261 and we obtain $Z_n = 22.4407 > C_{0.05,150} = 16.668$ which leads us to reject null hypothesis. Consequently, the estimated change occurs at 170th position. We repeat the same procedure for remaining subsequences and we fail to locate more changes.

Compared to the proposed test procedure to their methods, the proposed MIC procedure detected two different changes in data set in 26th December, 1997 and 29th January, 1999 respectively which concur with all of the three approaches, especially in Ngunkeng and Ning (2014) and Said, Ning and Tian (2017). There are two major differences between the proposed MIC procedure and the one based on BIC proposed by Ngunkeng and Ning (2014). First, the MIC procedure incorporates the effect of change locations by considering the complexity of the model, which leads to more powerful than the one based on traditional BIC especially for the change locations toward to the beginning or the end of a data. Simulations results in Section 3 also indicate this. Another difference is that the proposed MIC procedure is associated with a test statistic which can reveal the significance of changes statistically in terms of the critical values and p-values, instead of simply comparing BIC values under the null and alternative models as the one in Ngunkeng and Ning (2014).

4.2 Mexican stock market

To test for the independence in Mexican stock market data, we have the following results,

$$Q_{24} = 261 \times \sum_{i=1}^{24} r_i^2 = 261 \times 0.0910859 = 23.77342 < \chi^2_{0.95}(24) = 36.415.$$

Hence, we fail to reject the null hypothesis which indicates the R_t series for Mexican stock market data are independent. The left graph in Figure 4 shows the ACF values of the R_t series data and the right graph in Figure 4 shows the normal Q–Q plot of R_t series which indicates that the normality assumption fails. Applying normality tests such as Shapiro–Wilk test also concludes the validity of normality is violated.

Now, we apply the proposed MIC method with the binary segmentation procedure to R_t time series data.

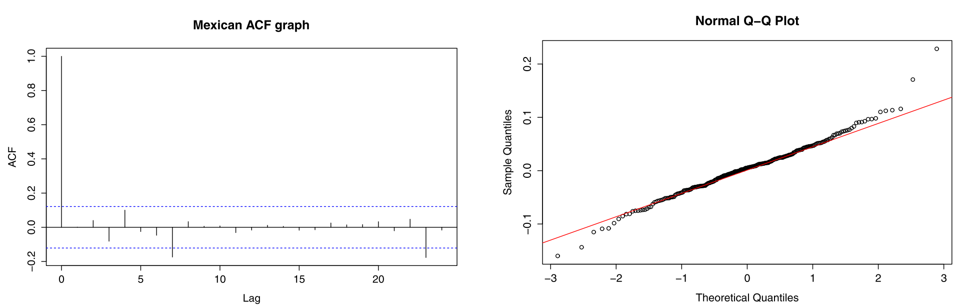

Figure 4 *Left*: *The graph of ACF values*. *Right*: *The Q–Q plot test for the transformed data*.

- We first screen the whole F_t series from 1 to 261 for potential changes. We calculate $S_n = \text{MIC}(n) - \min_{1\leq k<261} \text{MIC}(k) + 3\log(n) = 38.1982$. With the bootstrap method given in Section 3.1, the approximated p-value < 0.001 with $B = 2000$. Therefore, we reject the null hypothesis and conclude there is a change. Simultaneously, we obtain the estimated change location $\hat{k} = \arg\min_{1\leq k<261} \text{MIC}(k) = 94$, which is corresponding to 95th position in F_t series with the value 1284.851. This change occurred in August 22, 1997.
- Similarly, we check all possible subsequences by the binary segmentation method. We find one more change occurring at 142th position in F_t series which corresponding to the value 1055.913 in stock return. Such a change occurred in July 10, 1998.

In summary, we have found the changes at 95th and 142th positions which correspond to August 22, 1997 and July 10, 1998, respectively. These changes may be caused by the results of the 1995 Mexico's crises which involved in emerging markets occurring in January, 1996, Asian financial crises which had global effects and reached its climax in October 1997; and 1998 Russian financial crises. Figure 5 shows the Mexico monthly stock return rate and the monthly sock return index with identified change locations and Figure 6 shows the MIC values of Mexico stock return rate.

Arellano-Valle, Castro and Loschi (2013) analyzed the same data using the Bayesian approach. Their method detected a single change in the data and concluded that the change approximately happened during the first week of September, 1997 but did not give the estimated change point. They discussed that the data presents positive asymmetry before and after the change point. Furthermore, they pointed out that the mean return of the data is positive before and after the change point. Comparing to their method, the proposed MIC procedure can detect multiple changes in data associated with reasonable interpretations. Furthermore, we can obtain the point estimated values of change locations instead of approximate time windows of change locations only as Arellano-Valle, Castro and Loschi (2013) did. Therefore, based on the estimated change points, we can fit the data by

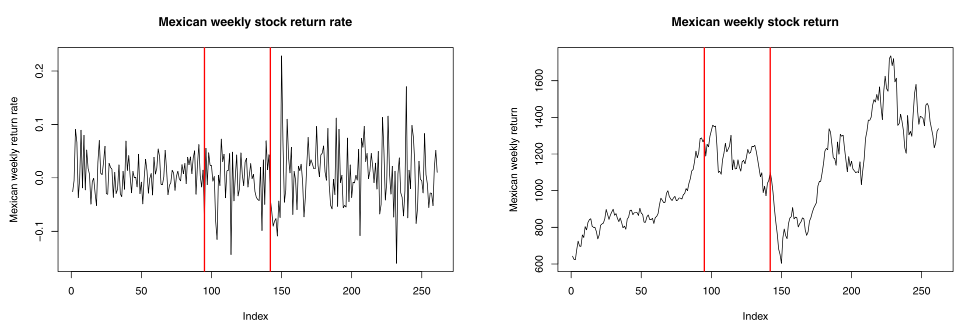

Figure 5 *Left*: *Mexico weekly stock return rate. Right*: *Stock return index with change locations.*

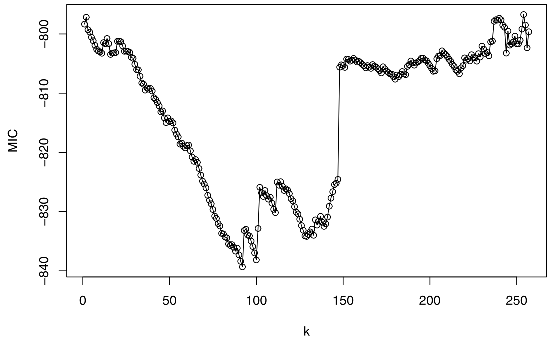

Figure 6 *The graph of MIC values of Mexican stock return rate.*

skew normal distributions with different parameters. Consequently, the estimated values of shape parameters can be obtained through the data fitting which can explicitly indicate the change of asymmetry and the exact magnitude of change in skewness.

Said, Ning and Tian (2017) proposed the detecting procedure for multiple change using likelihood ratio test (LRT) procedure. We analyze the Mexican stock market data using their procedure and we obtain that

$$Z_n = \max_{1\le k<n}\left\{-2\ln\left(\frac{\sup_{\mu,\sigma,\lambda} L_{H_0}(x_i;\mu,\sigma,\lambda)}{\sup_{\mu_1,\sigma_1,\lambda_1,\mu_n,\sigma_n,\lambda_n} L_{H_1}(x_i;\mu_1,\sigma_1,\lambda_1,\mu_n,\sigma_n,\lambda_n)}\right)\right\}$$
$$= 38.6335 > C_{0.05,261} = 16.842,$$

which lead us to reject the null hypothesis and conclude that there is a change occurring at 95th position corresponding with August 22, 1997. We repeat the same procedure with the help of binary segmentation method and no more change is found. Compared to the LRT procedure, the proposed MIC approach locate two different changes on the same data set, which are 95th and 131st positions respectively while LRT procedure only detect single change.

5 Discussion

We propose a skew normal change point model based on the modified information criterion (MIC). The procedure for detecting simultaneous changes in all three parameters in a skew normal distribution is established. Simulations are conducted to illustrate the performance of the proposed test procedure under different scenarios. Comparisons with the procedures based on the traditional Schwarz information criterion (BIC) and the likelihood ratio test (LRT) are investigated. The advantages of the proposed MIC procedure are illustrated through simulations. Since the analytic null distribution of the associated test statistic is not available, critical values are simulated at different significance levels. The convergence of the estimated change location is verified numerically. Finally, such a proposed procedure with the binary segmentation method is applied to two stock market data and several changes are identified successfully with interpretations.

In our current work, the associated test statistic S_n is studied only numerically and analytical results are not obtained. Chen, Gupta and Pan (2006) derived the asymptotic distributions of S_n as the standard chi-square distribution with the degree of freedom d where d is the number of parameters in S_n, as long as the conditions $W1$–$W7$ are satisfied by the distribution. However, for the skew normal distribution, it does not satisfy the condition $W5$, therefore, the asymptotic distribution of S_n does not follow the chi-square distribution with the degrees of freedom 3 in our case. We have verified this through the simulations which indicate the failure of the goodness-of-fit test for this distribution. Furthermore, we have conducted extensive simulations to explore the possible asymptotic distribution of S_n with various values of parameters and sample sizes. Interestingly, we find out that, for the large sample sizes (in our simulations, $n \geq 200$), as the increase of the sample size, the behavior of the test statistic gets closer to that of the chi-square distribution with the degrees of freedom 7. It also passes the goodness-of-fit test by Kolmogorov–Smirnov test with such a chi-square distribution. In our ongoing work, we plan to take a deeper look at this relationship and derive its analytic distribution and properties.

Acknowledgments

The authors would like to thank the anonymous referee for his/her constructive comments and suggestions which helped to improve this manuscript significantly.

References

Akaike, H. (1973). Information theory and an extension of the maximum likelihood principle. In *2nd International Symposium of Information Theory* (B. N. Petrov and E. Csaki, eds.), 267–281. Budapest: Akademiai Kiado. MR0483125

Arellano-Valle, R. B., Castro, L. M., Genton, M. G. and Gómez, H. W. (2008). Bayesian inference for shape mixtures of skewed distributions, with application to regression analysis. *Bayesian Analysis* **3**, 513–540. MR2434401

Arellano-Valle, R. B., Genton, M. G. and Loschi, R. H. (2009). Shape mixtures of multivariate skew-normal distributions. *Journal of Multivariate Analysis* **100**, 91–101.

Arellano-Valle, R. B., Castro, L. M. and Loschi, R. H. (2013). Change point detection in the skew-normal model parameters. *Communications in Statistics—Theory and Methods* **42**, 603–618. MR3211938

Azzalini, A. (1985). A class of distributions which includes the normal ones. *Scandinavian Journal of Statistics* **12**, 171–178.

Azzalini, A. (2016). The skew-normal and related distributions such as the skew-t. R package version 1.4.0. MR3468021

Azzalini, A. and Capitanio, A. (1999). Statistical applications of the multivariate skew-normal distribution. *Journal of the Royal Statistical Society, Series B* **61**, 579–602. MR1707862

Azzalini, A. and Capitanio, A. (2014). *The Skew-Normal and Related Families*. New York: Cambridge University Press.

Azzalini, A. and Dalla Valle, A. (1996). The multivariate skew-normal distribution. *Biometrika* **83**, 715–726. MR1440039

Cai, X., Said, K. K. and Ning, W. (2016). Change-point analysis with bathtub shape for the exponential distribution. *Journal of Applied Statistics.* **43**, 2740–2750. MR3546112

Chen, J. and Gupta, A. K. (1997). Testing and locating variance change points with application to stock prices. *Journal of the American Statistical Association* **92**, 739–747.

Chen, J. and Gupta, A. K. (2012). *Parametric Statistical Change Point Analysis with Applications to Genetics, Medicine, and Finance*, 2nd ed. Boston: Birkhäuser. MR3025631

Chen, J., Gupta, A. K. and Pan, J. (2006). Information criterion and change point problems for regular models. *Sankhya* **68**, 252–282.

Chernoff, H. and Zacks, S. (1964). Estimating the current mean of a normal distribution which is subject to changes in time. *The Annals of Mathematical Statistics* **35**, 999–1018.

Csörgő, M. and Horváth, L. (1997). *Limit Theorems in Change-Point Analysis*. New York: Wiley.

Gardner, L. A. (1969). On detecting change in the mean of normal variates. *The Annals of Mathematical Statistics* **40**, 116–126.

Hannan, E. J. and Quinn, B. G. (1979). The determination of the order of an autoregression. *Journal of the Royal Statistical Society, Series B* **41**, 190–195.

Hasan, A., Ning, W. and Gupta, A. K. (2014). An information-based approach to the change-point problem of the noncentral skew t distribution with applications to stock market data. *Sequential Analysis: Design Methods and Applications* **33**, 458–474.

Hawkins, D. M. (1992). Detecting shifts in functions of multivariate location and covariance parameters. *Journal of Statistical Planning and Inference* **33**, 233–244.

Henze, N. (1986). A probabilistic representation of the skew normal distribution. *Scandinavian Journal of Statistics* **13**, 271–275.

Hirotsu, C., Kuriki, S. and Hayter, A. J. (1992). Multiple comparison procedure based on the maximal component of the cumulative chi-squared statistic. *Biometrika* **79**, 381–392.

Hsu, D. A. (1979). Detecting shifts of parameter in gamma sequences with applications to stock price and air traffic flow analysis. *Journal of the American Statistical Association* **74**, 31–40.

Hsu, D. A. (1977). Tests for variance shifts at an unknown time point. *Applied Statistics* **26**, 179–184.

Inclán, C. (1993). Detection of multiple changes of variance using posterior odds. *Journal of Business & Economic Statistics* **11**, 189–300.

Ning, W. and Gupta, A. K. (2012). Matrix variate extended skew normal distributions. *Random Operators and Stochastic Equations* **20**, 299–310.

Ngunkeng, G. and Ning, W. (2014). Information approach for the change-point detection in the skew normal distribution and its applications. *Sequential Analysis: Design Methods and Applications* **33**, 475–490.

Owen, A. B. (1988). Empirical likelihood ratio confidence intervals for a single functional. *Biometrika* **75**, 237–249. MR0946049

Page, E. S. (1954). Continue inspection schemes. *Biometrika* **41**, 100–235.

Page, E. S. (1955). A test for a chance in a parameter occurring at an unknown point. *Biometrika* **42**, 523–527.

Said, K. K., Ning, W. and Tian, Y. B. (2017). Likelihood procedure for testing changes in skew normal model with applications to stock returns. *Communications in Statistics—Simulation and Computation* **46(9)**, 6790–6802.

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics* **6**, 461–464.

Vostrikova, L. J. (1981). Detecting 'disorder' in multidimensional random processes. *Soviet Mathematics Doklady* **24**, 55–59.

Zou, C., Liu, Y., Qin, P. and Wang, Z. (2007). Empirical likelihood ratio test for the change-point problem. *Statistics and Probability Letters* **77**, 374–382. MR2339041

Khamis K. Said
Yubin Tian
School of Mathematics and Statistics
Beijing Institute of Technology
Beijing
China

Wei Ning
Department of Mathematics and Statistics
Bowling Green State University
Bowling Green, Ohio 43403
USA
E-mail: wning@bgsu.edu