



---

## **Analysis of Malaysia Stock Return Using Mixture of Normal Distributions**

Zetty Ain Kamaruzzaman<sup>a\*</sup>, Zaidi Isa<sup>b</sup>

<sup>a</sup>*School of Mathematical Sciences, Faculty of Science and Technology, Universiti Kebangsaan Malaysia, 43600  
UKM Bangi, Selangor, Malaysia*

<sup>b</sup>*School of Mathematical Sciences, Faculty of Science and Technology, Universiti Kebangsaan Malaysia, 43600  
UKM Bangi, Selangor, Malaysia*

<sup>a</sup>*Email: zettyain@yahoo.com*

<sup>b</sup>*Email: zaidiisa@ukm.my*

### **Abstract**

In this paper, two component univariate mixtures of Normal distributions is proposed to accommodate the non-normality and asymmetry characteristics of financial time series data as found in the distribution of monthly rates of returns for Bursa Malaysia Index Series namely the FTSE Bursa Malaysia Composite Index (FBMKLCI) from July 1990 until July 2010. Firstly, we give some basic definitions and concepts of mixtures of Normal distributions. Next, we explore some of its distribution properties. In support of determining the number of components, we use the information criterion for model selection. The measures provide supporting evidence in favour of the two-component mixtures of Normal distributions. For parameter estimation, we apply the most commonly used Maximum Likelihood Estimation (MLE) via the EM algorithm to fit the two-component mixtures of Normal distributions using data set on logarithmic stock return of Bursa Malaysia index.

**Keywords:** Bursa Malaysia stock market index; behavior of financial time series; distributional properties of Normal mixtures; EM algorithm; mixture of Normal distributions; model selection criteria; stock market return modeling

---

\* Corresponding author.

E-mail address: zettyain@yahoo.com .

## 1. Introduction

Assumption regarding the returns distribution on financial assets plays a vital role in both financial modelling as well as in its applications. For instance, one may underestimate the occurrence of extreme financial events such as market crashes if ally with a wrong distributional assumption. According to [1], it is difficult to use a single distribution family to estimate and model return distributions with various distributional characteristics as found in numerous financial assets. Therefore, this paper paves the way for an easy applied estimation of returns distribution using mixtures of Normal distributions. Rather than simply conclude that financial markets are normally distributed in the traditional sense, we could utilize this paper to provide a better understanding of mixtures of Normal distributions, its flexibility and its application in finance. Although the premise of this paper (mixtures of Normal distributions) may sound obvious and well known in finance literature [2], the significance of this paper lies in the application of modelling stock market return using mixtures of Normal distributions particularly in Malaysia case.

In this paper, we focus both on the statistical and financial properties of the mixtures of Normal distributions. The paper is structured as follows. We present the stylized facts of stock market return of FTSE Bursa Malaysia Composite Index (FBMKLICI) in Section 2. In Section 3, we use the information criterion for model selection for determining the number of components where the supporting evidence in favour of the two-component mixtures of Normal distributions. In Section 4, we give some basic definitions and concepts of mixtures of Normal distributions and explore some of its distribution properties. For parameter estimation, we use the Maximum Likelihood Estimation (MLE) via the EM algorithm to fit the two-component mixtures of Normal distributions using data set on logarithmic stock return of Bursa Malaysia index in Section 5. Lastly, Section 6 concludes.

## 2. Stylized Facts of Stock Market Return

The data set used in this paper is the monthly closing price covers a twenty years period from July 1990 to July 2010 for Malaysia stock market index namely the FTSE Bursa Malaysia Composite Index (FBMKLICI) as obtained from DataStream. The FBMKLICI is formerly known as the Kuala Lumpur Composite Index (KLICI) and adopts the FTSE global index standard. The enhancements were implemented from 6 July 2009 onwards (<http://www.bursamalaysia.com>). The FBM KLICI comprises the largest 30 companies listed on the Main Board. The series is denominated in Malaysian Ringgit (MYR). The sample consists of 241 observations.

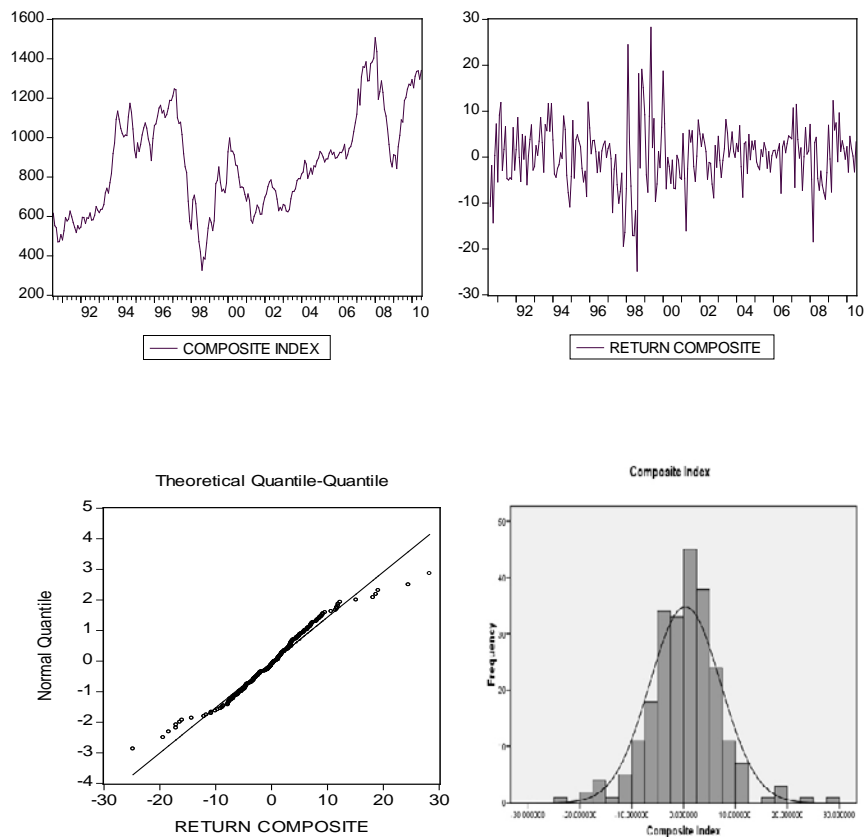
Prior to analysis, the series is analyzed in return, which is the first difference of natural logarithms multiplied by 100 over the whole period. This is done to express things in percentage terms. Let  $P_{it}$  be the observed monthly closing price of market index  $i$  on day  $t$ ,  $i = 1, \dots, n$  and  $t = 1, \dots, T$ . The monthly rates of return is defined as

the percentage rate of return by  $y_{it} = 100 \times \log \left( \frac{P_{it}}{P_{i,t-1}} \right)$ .

Figure 1 depicts the time series plot of monthly stock market index of Bursa Malaysia together with the monthly return of Bursa Malaysia stock market index plot for the time span July 1990 until July 2010, the histogram for

the Malaysia's stock return rate with the corresponding normal curve with same mean and standard deviation as well as the Normal QQ-plot for the Malaysia logarithmic return.

From Figure 1(a) and 1(b), it can be seen that the price rise and fall over time. There are periods of quiet and periods of wild variation in the monthly return. The period analyzed can be characterized as a period of market instability as it reflects the upturn and downturn of Malaysia stock market. It can also be seen that the growth pattern in Malaysia has not been smooth at all times. Malaysia undergoes sequence of upward and downward economical episodes due to global crisis. Just to mention a few; in 1997, Malaysia experienced a reduction in economic growth due to the Asian financial crisis where the rapid growth in Asian economies had come to a halt; The September 11 attacks had a significant economic impact on world market; bird flu epidemic especially in Asia in 2003; subprime mortgage crisis in 2006; the price of petroleum spiked in 2008 as well as the rapid increase in food price on the same year; and H1N1 attacks in 2009. Those can be identified as shocks in the stock market of Malaysia [3], [4]. From Figure 1(c), it can easily be seen that the empirical distribution has higher peak and heavier tails than the Normal distribution. Also note that the return distribution with thicker tail has a thinner and higher peak in the center compared to Normal distribution. The fit of the Normal curve to the histogram is poor. Figure 1(d) depicts the Normal QQ-plot for the corresponding logarithmic return. For the Malaysia stock return distribution, both tails are heavier than the Normal distribution, and the right tail is heavier than the left tail.



**Figure 1:** Clockwise from upper left: (a) time series plot of monthly stock market index, (b) time series plot of index rate for Malaysia stock market, (c) empirical distribution plot of Malaysia index, (d) QQ-plot

Table 1 reports the descriptive statistics and tests for the monthly stock price over the whole sample period. First, the means of the series, in general, not significantly different from zero ( $H_0: \mu = 0$ ). Second, there is evidence of positive skewness ( $\beta_1$ , defined as the 3<sup>rd</sup> standardized moment) in the monthly Composite Index. Third, it has been found that stock return in financial market have excess kurtosis, i.e. kurtosis which is significantly greater than 3 (the value for a Normal distribution). Kurtosis  $\beta_2$ , defined as the ratio of the 4<sup>th</sup> central moment to the square of the variance, increases both with excessive mass in the tails or at the centre of the distribution. The Jarque-Bera test rejects the null hypothesis of normality for the stock market return. We also applied the well-known Kolmogorov-Smirnov (K-S) test. As Table 1 show, the result from the K-S test, we find weak evidence of normality in monthly Composite Index. In order to further investigate deviations from normality, we applied the Anderson-Darling (A-D) test. As Table 1 show, the result from the A-D test, normality is overwhelmingly rejected for the monthly Composite Index.

**Table 1:** Statistical properties of Malaysia stock price

Statistics	Composite Index
Mean	0.3237
Standard deviation	6.8727
Skewness	0.0749
Kurtosis	5.2621
Jarque-Bera	51.3942***
Kolmogorov-Smirnov	0.0620**
Anderson-Darling	1.6860***

Significance levels: \* 10 percent, \*\* 5 percent, \*\*\* 1 percent

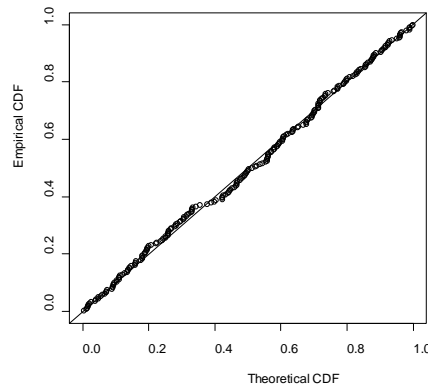
Thus, the return distribution of Bursa Malaysia Composite Index is poorly represents by the Normal distribution. It is a stylized facts that returns distributions of financial assets exhibits non-normality and asymmetry characteristics where it have thick tails, are skewed and leptokurtic relative to the Normal distribution [5], [6], [7]. This happens because it have more values near the mean and in the extreme tails [8] and dramatic falls and spectacular jumps appear with higher frequency than predicted [9]. Stylized facts as defined by [6] are statistical properties of financial time series, common across a wide range of instruments, markets and time periods.

### 3. Model Selection

Determining the number of components,  $k$ , is a major issue in mixture modelling. Two commonly employed techniques in determining the number of components,  $k$ , are the information criterion and parametric bootstrapping of the likelihood ratio test statistic values [10]. Majority of the estimation techniques assume that the number of components,  $k$ , in the mixture is known a priori where it is known before the estimation of parameters is attempted [11].

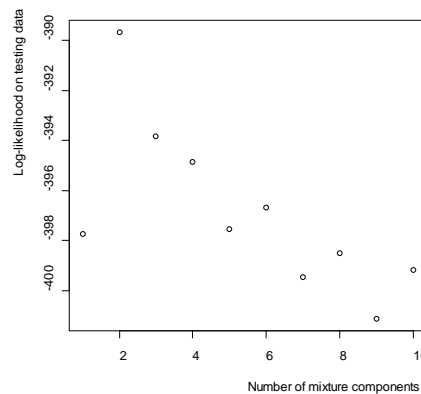
First, we did the calibration checking for the mixtures of Normal distributions. Figure 2 depicts the calibration

plot for the two-component mixtures of Normal distributions. Examining the two-component mixtures of Normal distributions plot, it does look satisfactory.



**Figure 2:** Calibration plot for the two-component mixtures of Normal distributions

Next, we do cross-validation to confirm the selection of number of the components for the mixture model. We do simple data-set splitting, where a randomly-selected half of the data is used to fit the model, and half to test. The basic idea is to split a data set into train and test. We fit the model using the training points, and then calculate the log-likelihood of the test points under the model. We pick the number of component which maximizes the likelihood of the data. Figure 3 and Table 2 depict the log-likelihoods of the 10 fold cross-validation. In Table 2,  $k$  is the number of components, meanwhile,  $\log L$  is the value that maximized log-likelihood. The boldface entry confirms the two components to the mixture model.



**Figure 3:** Log-likelihoods of different sizes of mixture models, fit to a random half of the data for training, and evaluated on the other half of the data for testing

In addition, we apply the two commonly used model selection criteria: the Akaike information criterion, AIC [12] and the Bayesian information criterion, BIC [13] as a tool for model selection. Table 3 depicts the likelihood based goodness-of-fit measures for univariate mixtures of Normal distributions fitted to the Bursa

Malaysia stock market index.  $k$  is the number of components,  $K$  denotes the number of parameters of a model,  $\log L$  is the value that maximized log-likelihood, AIC is the Akaike information criterion,  $AIC = -2 \log L + 2K$ , BIC is the Bayesian information criterion,  $BIC = -2 \log L + K \log N$  where  $N = 241$  is the sample size. Table 3 also reports the AIC and the BIC for model selection criteria. The number of components can be selected based on a minimum value of AIC and BIC. In our case, the best model according to BIC is an unequal variance with two components. The corresponding plots are shown in Figure 4.

**Table 2:** Log-likelihoods of the 10 fold cross-validation

k	log-L
1	-397.7372
<b>2</b>	<b>-389.6632</b>
3	-393.8297
4	-394.8476
5	-397.5470
6	-396.6822
7	-399.4821
8	-398.5211
9	-401.1397
10	-399.1875

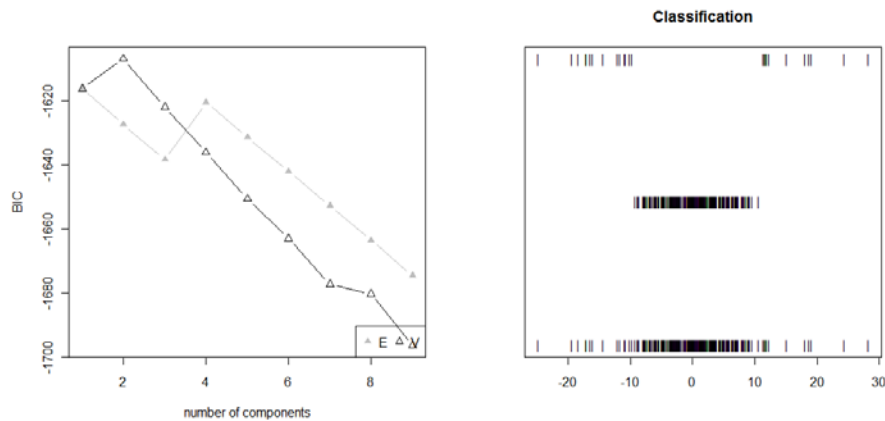
Since a two-component Normal mixture seems good, we should not consider using more components as by going to three, four, etc. components, we improve the in-sample likelihood but we could expose ourselves to the danger of over-fitting. Besides, having so many parameters is not always desirable. It can lead to estimation problems and over-fitting the data can lead to specification problems.

#### 4. Mixtures of Normal Distributions

The Normal distribution is commonly used in the 1700's and successfully applied to astronomical data analysis by Karl Gauss in the 1800 and mostly assumption in previous research on returns distribution of financial assets are approximately normally distributed. However, from the late 1960's, the empirical finance analyses failed to support the normality assumption on estimating the returns distribution of financial assets. [14] and [15] have pointed out that excess kurtosis and heavy tails exist and established the empirical evidence on the non-normality in returns distribution of financial assets. Later, strong evidence by numerous empirical finance studies has indicated that most of the returns distributions of financial assets are non-normal where normality is overwhelmingly rejected in many returns distribution [1], [16].

**Table 3:** Assessing the number of components, k

k	1	2	3
K	2	6	9
logL	-397.7372	<b>-389.6632</b>	-393.8297
AIC	799.4744	<b>791.3264</b>	805.6594
BIC	800.2384	<b>793.6185</b>	811.1334



**Figure 4:** (left) the BIC plot for the dataset, with vertical axes adjusted to display the maximum values (right) the classification plot, all of the data is displayed at the bottom, with the separated classes shown on different levels above

As discussed earlier, we know that the assumption of normally distributed returns is not valid. Two features that account for non-normality in finance and economic series are; one is the presence of big shocks or outlying observations or rare events, another is abrupt regime changes over different sub-periods [17]. The worst part of the assumption of a Normal distribution is particularly with respect to the tail behaviour of the series where the tails of a Normal distribution taper very rapidly, hence the Normal assumption will exclude the possibility of extreme returns where such events are frequently seen in financial markets [18]. Moreover, according to [17], Normal distribution has tails that are too thin to accommodate shocks in financial markets. [19] showed an illustration of a normal distribution model that is quite inappropriate for fitting market data since its density does not take into accounts the fat tails and skewness.

One way to accommodate this stylized fact is to introduce a more flexible distribution model. Mixtures of Normal distributions have gain increasing attention in various disciplines of knowledge. The earliest recorded application of the mixtures of Normal distributions was undertaken by Simon Newcomb in his study in Astronomy in 1886 followed by Karl Pearson in his classic work on Method of Moments in 1894. In empirical finance applications, the use of mixtures of Normal distributions to handle fat tails was first considered by [20]. [21] proves that a mixture of Normal distributions is leptokurtic, when all regimes have the same mean. There exists a long history of modelling asset returns with mixtures of Normal distributions [22], [23], [24], [25]. [26] examined the daily returns from 30 different stocks in the Dow-Jones Industrial Average, estimated mixtures of

Normal distributions with two up to four components which were found to fit appropriately and showed that the mixtures of Normal distributions have more descriptive validity than a Student-t distribution. Others are [8], [27] and [28]. And the lists of literatures continue to emerge until at this moment.

Most financial markets returns are both skewed and leptokurtic. Based on the above analysis, the FBMKLCI is no exception; the monthly log return is far from being normally distributed. Hence, a number of alternatives skewed and leptokurtic distributions have been applied. The mixtures of Normal distributions is by far the most extensively applied and the simplest case is a mixture of two univariate Normal distributions may be considered as the most widely applied. A flexible and tractable alternative of departures from normality is a mixture of two Normal distributions. A mixture of two log Normal distributions fit financial data better than a single Normal distribution. [14] claims that a mixture of several Normal distributions with same mean but different variances are the most popular approach to describe long-tailed distribution of price changes.

One of the most appealing features of the mixtures of Normal distributions for modelling assets returns is that it has the flexibility to approximate various shapes of continuous distributions by adjusting its component weights, means and variances [1]. Other advantages of using mixtures of Normal distributions are they maintain the tractability of Normal, have finite higher order moments, plus can capture excess kurtosis [29]. Besides, mixtures of Normal distributions can capture the structural change both in the mean and variance and it can be asymmetric [30]. Also the mixtures of Normal distributions are easy to interpret if the asset returns are viewed as generated from different information distributions where the mixture proportion can accommodate parameter cyclical shifts or switches among a finite number of regimes [31].

Some attractive property of the mixtures of Normal distributions is that it is flexible to accommodate various shapes of continuous distributions, and able to capture leptokurtic, skewed and multimodal characteristics of financial time series data. Also it is believed that mixtures of Normal distributions are appropriate in order to accommodate certain discontinuities in stock returns such as the 'weekend effect', the 'turn-of-the month effect' and the 'January effect' [32]. [33] discuss how skewness and excess kurtosis in financial time series can be deal using finite mixtures of Normal distributions? and illustrate why mixtures of normal distributions provide a flexible way of dealing with skewness and kurtosis?

A good introduction to the theory and applications of mixtures of Normal distributions can be found in [11], [34], [35], [36], [37]. Meanwhile, various applications of the mixtures of Normal distributions in empirical finance are documented in [2] and [38].

A mixture of two Normal densities is defined by

$$g(x) = \pi\varphi_1(x) + (1-\pi)\varphi_2(x), \quad 0 < \pi < 1, \quad (1)$$

where  $\varphi_1$  and  $\varphi_2$  are two Normal densities with different expectations and variances. In general, the cumulative distribution function (cdf) of a mixture of  $k$  normal random variable  $X$  can be represented by



$$F(x) = \sum_{i=1}^K \pi_i \Phi\left(\frac{x - \mu_i}{\sigma_i}\right) \tag{2}$$

where  $\Phi$  is the cdf of  $N(0,1)$ . Therefore its probability density function (pdf) is

$$f(x) = \sum_{i=1}^K \pi_i \phi(x; \mu_i, \sigma_i) \tag{3}$$

where, for  $i = 1, \dots, K$

$$\phi(x; \mu_i, \sigma_i) = \frac{1}{\sqrt{2\pi}\sigma_i} e^{-\frac{(x-\mu_i)^2}{2\sigma_i^2}}$$

$$\sum_{i=1}^K \pi_i = 1 \quad \text{and} \quad 0 \leq \pi_i \leq 1$$

If  $X$  is a mixture of  $K$  Normals with pdf in (3), then its mean, variance, skewness and kurtosis are

$$\begin{aligned} \mu &= \sum_{i=1}^K \pi_i \mu_i \\ \sigma^2 &= \sum_{i=1}^K \pi_i (\sigma_i^2 + \mu_i^2) - \mu^2 \\ \tau &= \frac{1}{\sigma^3} \sum_{i=1}^K \pi_i (\mu_i - \mu) [3\sigma_i^3 + (\mu_i - \mu)^2] \\ \kappa &= \frac{1}{\sigma^4} \sum_{i=1}^K \pi_i [3\sigma_i^4 + 6(\mu_i - \mu)^2 \sigma_i^2 + (\mu_i - \mu)^4] \end{aligned} \tag{4}$$

[2] defines finite mixture distribution as a probability-weighted sum of other distribution functions where the density function of a mixture distribution is the same probability-weighted sum of the component density function. In mixtures of Normal distributions, the return distribution is approximated by a mixture of Normal each has unique mean  $\mu_i$ , standard deviation  $\sigma_i$  and weight (sometimes also called as probability or mixing parameter)  $\pi_i$  [39]. The mixture setting, according to [2] is design to capture different market regimes. The typical interpretation of a mixture of two Normal distributions is that there are two regimes for returns. One where the return has mean  $\mu_1$  and variance  $\sigma_1^2$  another where the return has mean  $\mu_2$  and variance  $\sigma_2^2$ . The weight is the probability  $\pi$  which the first regime occurs while the second regime occurs with probability  $1 - \pi$ . For example, in a two-component mixture ( $k = 2$ ), the first component, with a relatively high mean and small variance, may be interpreted as the bull market regime, occurring with probability  $\pi$  whereas the second regime, with a lower expected return and a greater variance, represent the bear market.

Thus, the five parameters  $(\pi, \mu_1, \mu_2, \sigma_1^2, \sigma_2^2)$  of mixtures of Normal distributions allow a very flexible definition of departures from symmetry and normality. Therefore, mixtures of Normal distributions have the ability to deal with skewness and kurtosis in analyzing financial time series. By using mixtures of Normal distributions, we can obtain densities with higher peaks and heavier tails than Normal distribution.

## 5. Parameter Estimation

Fitting mixture distributions can be handled by a wide variety of techniques, such as graphical methods, method of moments, minimum-distance methods, maximum likelihood and Bayesian approaches (see [34] for an exhaustive review of these methods). Considerable advances have been made in the fitting of mixture models especially via the maximum likelihood method. The maximum likelihood method has focused many attentions and by far has been the most commonly used approach to fit the mixture distributions mainly due to the existence of an associated statistical theory and since the advent of the EM algorithm. The key property of the EM algorithm has been established by [40]. The EM algorithm is a popular tool for simplifying maximum likelihood problems in the context of a mixture model. The EM algorithm has become the method of choice for estimating the parameters of a mixture model, since its formulation leads to straightforward estimators [41].

Fitting the parameters of the mixtures of Normal distributions is one of the oldest estimation problems in the statistical literature. A variety of approaches have been used to estimate the mixtures of Normal distributions as discussed earlier. The maximum likelihood method (MLE) is the most commonly preferred method for the estimation problem of mixtures of Normal distributions. Unfortunately, the MLEs have no closed forms; hence they have to be computed iteratively. However, the computation becomes straightforward using the expectation-maximization (EM) algorithm.

The EM algorithm is widely used as it is an easy and implementable method as well as a popular tool for simplifying difficult maximum likelihood problems plus has shown great performance in practice where it has the ability to deal with missing data, unobserved variables and mixture density problems. The EM algorithm will find the expected value as well as the current parameter estimates at the E step and maximizes the expectation at the M step. By repeating the E and M step, the algorithm will converge to a local maximum of the likelihood function. Various EM-type algorithms can be found in the literature (for example, see [37] and [42] for references).

Denote  $\theta$  the parameters of the function to be optimized. The algorithm consists of iterating between two steps, the E-step and the M-step. In the Expectation (E) step, the current estimates of the parameters are used to assign responsibilities according to the relative density of the training points under each model. Next, in the Maximization (M) step, these responsibilities are used in weighted maximum-likelihood fits to update the estimates of the parameters. The E-step is repeated, updated with a new value as the current value of  $\theta$  and then the M-step again provides a further updated value for  $\theta$ . Thus, the algorithm proceeds, iterating between the E-step and the M-step until convergence is achieved. Reference [43] introduces a simple procedure of the EM algorithm for the special case of mixtures of Normal distributions.

- Expectation (E) Step

$$\gamma_i = \frac{\pi\phi(x_i; \mu_2, \sigma_2^2)}{\pi\phi_1(x_i; \mu_1, \sigma_1^2) + (1-\pi)\phi_2(x_i; \mu_2, \sigma_2^2)} \quad (5)$$

for  $t = 1, \dots, T$ .

- Maximization (M) Step

$$\begin{aligned} \mu_1 &= \frac{\sum_{t=1}^T (1-\gamma_t)x_t}{\sum_{t=1}^T (1-\gamma_t)}, & \sigma_1^2 &= \frac{\sum_{t=1}^T (1-\gamma_t)(x_t - \mu_1)^2}{\sum_{t=1}^T (1-\gamma_t)}, \\ \mu_2 &= \frac{\sum_{t=1}^T \gamma_t x_t}{\sum_{t=1}^T \gamma_t}, & \sigma_2^2 &= \frac{\sum_{t=1}^T \gamma_t (x_t - \mu_2)^2}{\sum_{t=1}^T \gamma_t}, \\ \pi &= \sum_{t=1}^T \frac{\gamma_t}{T} \end{aligned} \tag{6}$$

As an illustration, we apply the maximum likelihood method via EM algorithm to fit the mixtures of Normal distributions with two components to the Bursa Malaysia monthly return. Table 4 depicts the summary of two components mixtures of Normal distributions using the EM algorithm. There are two components, with weight ( $\pi$ ), two means ( $\mu$ ), two standard deviations ( $\sigma$ ), implied skewness ( $\tilde{\beta}_1$ ), implied kurtosis ( $\tilde{\beta}_2$ ) and the overall log-likelihood (logL). Also stated are the standard errors, ratios and confidence intervals for the five parameters.

Several important observations may be drawn from Table 4. First, in general the low-variance component has a higher probability. The second component has a lower variance in the monthly Composite Index. The high-variance component has the smaller probability for this data. The Composite Index indicates that the first Normal is a low mean high variance regime and the second Normal is a high mean low variance regime. Meanwhile, the weights indicate that the second regime is the more prevalent regime for the stock market index of Bursa Malaysia.

The estimation results for the two-component mixtures of Normal distributions are

$$f(x_t) = 0.2222N(-0.1996, 132.2486) + 0.7778N(0.4732, 22.5891)$$

In order to judge whether the estimated models are compatible with the stylized facts of the data, we compute the implied skewness  $\tilde{\beta}_1$  and the implied kurtosis  $\tilde{\beta}_2$  of the models from

$$\begin{aligned} \tilde{\beta}_1 &= \frac{\sum_{i=1}^I p_i (3\sigma_i^2 \delta_i + \delta_i^3)}{\left[ \sum_{i=1}^I p_i (\sigma_i^2 + \delta_i^2) \right]^{3/2}} \\ \tilde{\beta}_2 &= \frac{\sum_{i=1}^I p_i (3\sigma_i^4 + 6\sigma_i^2 \delta_i^2 + \delta_i^4)}{\left[ \sum_{i=1}^I p_i (\sigma_i^2 + \delta_i^2) \right]^2} \end{aligned} \tag{7}$$

with  $\delta_i = \mu_i - \mu$ , where  $\mu$  is the overall mean.

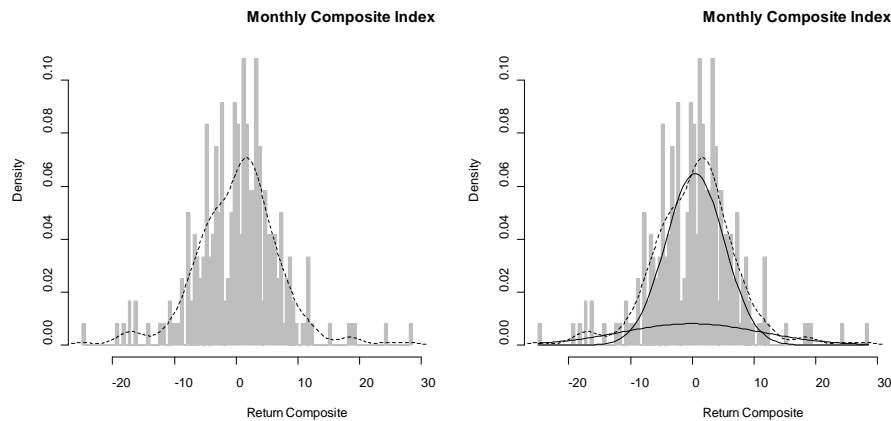
The results, reported in Table 4, show a rather close agreement between the pattern of skewness and kurtosis in the data with the implied skewness and kurtosis. There is a quite close agreement between implied leptokurtosis and actual leptokurtosis for the monthly data.

**Table 4:** Summary of two components mixtures of Normal distributions using EM algorithm

Parameter	Statistics	Component 1	Component 2
Weight $\pi$	Estimate	0.2222	0.7778
	Standard error	19.1709	-
	Ratio	0.0116	-
	95% CI	(0.0407,0.4038)	-
Mean $\mu$	Estimate	-0.1996	0.4732
	Standard error	0.5764	2.4166
	Ratio	-0.3463	0.1958
	95% CI	(-3.7684,3.3692)	(-0.3799,1.3263)
Variance $\sigma^2$	Estimate	132.2486	22.5891
	Standard error	0.0355	0.3267
	Ratio	3725.3127	69.1433
	95% CI	(47.1284,217.3687)	(14.7898,30.3883)
Implied skewness $\tilde{\beta}_1$			-0.1020
Implied kurtosis $\tilde{\beta}_2$			5.7435
Estimate log-likelihood (log-L)			-789.6755

Mixtures of Normal distributions have an intuitive interpretation when markets display regime-specific behaviour [2]. Markets are stable when the expected return is relatively small and positive and the volatility is relatively low, but market crash as the expected return is relatively large and negative and the returns volatility is larger than when markets are stable. Therefore, from Table 4, during a stable market regime, which occurs 78% of the time, have expectation 0.47 and standard deviation 4.75. However, 22% of the time the market is in turmoil with an expected return of -0.20 and a standard deviation of 11.50.

Next, we plot the histogram of the data and the non-parametric density estimate (Figure 5a). In Figure 5b, we add the density of a given component to the current plot, but scaled by the share it has in the mixture, so that it is visually comparable to the overall density.



**Figure 5:** Figure 5a (left) Histogram (grey) for monthly stock price of Bursa Malaysia. The dashed line is a kernel density estimate, which is not completely satisfactory. Figure 5b (right) As in the previous figure, plus the components of a mixture of two Normal, fitted to the data by the EM algorithm. These are scaled by the mixing weights of the components.

## 6. Conclusion

In this paper, we present the empirical evidence on the stock market return based on its time series patterns. We propose the mixtures of Normal distributions, which is a flexible family of distribution to accommodate the non-normality and asymmetric characteristics of financial time series data as found in the distribution of monthly rates of returns for the FTSE Bursa Malaysia Composite Index (FBMKLCI) from July 1990 until July 2010. We define the mixtures of Normal distributions and explore some of its distribution properties. We fit the two component mixtures of Normal distributions to data sets using the maximum likelihood estimation via EM algorithm. We may conclude from the above analysis that using the mixtures of two Normal distributions to fit the market data of stock market return captures well the stylized facts of non-normality and leptokurtosis.

## Acknowledgements

The authors want to thank anonymous referees whose comments, suggestions, and corrections led to a considerably improvement of this paper.

## References

- [1] K. Tan and M. Chu. "Estimation of portfolio return and value at risk using a class of gaussian mixture distributions." *The International Journal of Business and Finance Research*, vol. 6, pp. 97-107, 2012.
- [2] C. Alexander. *Market Risk Analysis Volume I: Quantitative Methods In Finance*. West Sussex: John Wiley & Sons Ltd, 2008, pp. 94-97.
- [3] M-Y. Cheng. "Economic fluctuations and growth: an empirical study of malaysian economy." *The Journal of Business in Developing Nations*, vol. 7, pp. 51-74, 2003.

- [4] P-C. Athukorala. "Malaysian economy in three crises." *Working Paper in Trade and Development*, Working Paper No. 2010/12, Oct. 2010.
- [5] H. Eijgenhuijsen and A. Buckley. "An overview of returns in Europe." *European Journal of Finance*, vol. 5, pp. 276–297, 1999.
- [6] R. Cont. "Empirical properties of asset returns: stylized facts and statistical issues." *Quantitative Finance*, vol. 1, pp. 223–236, 2001.
- [7] A. Behr and U. Pötter. "Alternatives to the normal model of stock returns: gaussian mixture, generalized logF, and generalized hyperbolic models." *Annals of Finance*, vol. 5, pp. 49-68, 2009.
- [8] J.A. Hall, B.W. Brorsen and S.H. Irwin. "The distribution of future prices: A test of the stable Paretian and the mixture of normals hypotheses." *Journal of Financial Quantitative Analysis*, vol. 24, pp. 105-116, 1989.
- [9] P.H. Frances and D. van Dijk. *Non Linear Time Series Models In Empirical Finance*. Cambridge: Cambridge University Press, 2000.
- [10] G.J McLachlan. "On bootstrapping the likelihood ratio test statistic for the number of components in a normal mixture." *Applied Statistics*, vol. 36, pp. 318-324, 1987.
- [11] B.S Everitt and D.J. Hand. *Finite Mixture Distributions*. London: Chapman and Hall, 1981, pp. 107-127.
- [12] H. Akaike. "Information theory and an extension of the maximum likelihood principle," in *Second international symposium on information theory*, B.N. Petrov and F. Csaki, Ed. Budapest: Akademiai Kiado, 1973, pp. 267-281.
- [13] G. Schwarz. "Estimating the dimension of a model." *The Annals of Statistics*, vol. 6, pp 461-464, 1978.
- [14] E.F Fama. "The behavior of stock-market prices." *Journal of Business*, vol. 38, pp. 34-105, 1965.
- [15] B.B Mandelbrot. "The variation of certain speculative prices." *Journal of Business*, vol. 26, pp. 394-419, 1963.
- [16] D.N. Esch. "Non-normality facts and fallacies." *Journal of Investment Management*, vol. 8, pp. 49-61, 2010.
- [17] P.V. Bidarkota. "Modeling economic time series with stable shocks." PhD Dissertation, The Ohio State University, 1996.
- [18] R.H. Rimmer. "A mathematical description of markets." Internet: <http://www.mathestate.com/tools/Financial/wp1/MarketTheory.html>, Sept. 20, 2008.
- [19] J. Wang. "Generating daily changes in market variables using a multivariate mixture of normal distributions," in *Proceedings of the 2001 Winter Simulation Conference*, 2001, pp. 283-289.
- [20] S. Newcomb. "A generalized theory of the combination of observations so as to obtain the best result," in *American contributions to mathematical statistics in the nineteenth century*, vol. 2, M. Stigler, Ed. New York: Arno Press, 1963, pp. 343–366.
- [21] N.T. Gridgeman. "A comparison of two methods of analysis of mixtures of normal distributions." *Technometrics* , vol. 12, pp. 823–833, 1970.
- [22] S.J. Press. "A compound events model for security prices." *Journal of Business*, vol. 40, pp. 317-335, 1967.
- [23] P.D. Praetz. "The distribution of share price changes." *Journal of Business*, vol. 45, pp. 49-55, 1972.
- [24] P.K. Clark. "A subordinated stochastic process model with finite variance for speculative prices." *Econometrica*, vol. 41, pp. 135–155, 1973.

- [25] R.C. Blattberg and N.J. Gonedes. "A comparison of the stable and student distributions as statistical models for stock prices." *Journal of Business*, vol. 47, pp. 244-280, 1974.
- [26] S.J. Kon. "Models of stock returns – a comparison." *Journal of Finance*, vol. 39, pp. 147-165, 1984.
- [27] A.L. Tucker. "A reexamination of finite- and infinite-variance distributions as models of daily stock returns." *Journal of Business & Economic Statistics*, vol. 10, pp. 73-81, 1992.
- [28] K.C. Tran. "Estimating mixtures of normal distributions via empirical characteristic function." *Econometric Reviews*, vol. 17, pp. 167-183, 1998.
- [29] R.S. Tsay. *Analysis of Financial Time Series*. New Jersey: Wiley Series in Probability and Statistics, 2005, pp. 15-16.
- [30] J. Knight and S. Satchell. *Return Distributions in Finance*. Oxford: Butterworth-Heinemann: Quantitative Finance Series, 2001.
- [31] D. Xu and T. Wirjanto. "An empirical characteristic function approach to VaR under a mixture-of-normal distribution with time-varying volatility." *Journal of Derivatives*, vol. 18, pp. 39-58, 2010.
- [32] B. Klar and S.G. Meintanis. "Tests for normal mixtures based on the empirical characteristic function." *Computational Statistics and Data Analysis*, vol. 49, pp. 227-242, 2005.
- [33] J.G. Dias, J.K. Vermunt and S. Ramos. "Analysis of heterogeneous financial time series using a mixture gaussian hidden markov model." *Elsevier Science*, 2007.
- [34] D.M. Titterton, A.F.M. Smith and U.E. Makov. *Statistical Analysis of Finite Mixture Distributions*. New York: John Wiley & Sons, 1985.
- [35] G.J. McLachlan and K.E. Basford. "Mixture Models: Inference and Applications To Clustering." New York: Marcel Dekker, 1988, pp. 4-5.
- [36] B.G. Lindsay. *Mixture Models: Theory, Geometry and Applications*. NSF-CBMS regional conference series in probability and statistics, Hayward, CA: Institute of Mathematical Statistics, 1995.
- [37] G. McLachlan and D. Peel. *Finite Mixture Models*. New York: John Wiley & Sons, 2000, pp. 35-36.
- [38] S. Frühwirth-Schnatter. *Finite Mixture and Markov Switching Models*. New York: Springer, 2006, pp. 169-177.
- [39] S. Subramanian and U.S Rao. "Sensex and stylized facts an empirical investigation." *Social Science Research Network*, id. 962828, 2007.
- [40] A.P. Dempster, N.M. Laird and D.B. Rubin. "Maximum likelihood from incomplete data via the EM algorithm." *Journal of Royal Statistical Society B*, vol. 39, pp. 1-38, 1977.
- [41] F. Picard. "An introduction to mixture models." *Statistics for Systems Biology Group*, Research Report No 7, 2007.
- [42] G.J. McLachlan and T. Krishnan. *The EM Algorithm and Extensions*. New York: Wiley, 1997.
- [43] T. Hastie, R. Tibshirani and J. Friedman. *The Elements of Statistical learning: Data Mining, Inference and Prediction*. New York: Springer Verlag, 2001, pp. 272-275.