

Research Article

Using Statistical Model to Study the Daily Closing Price Index in the Kingdom of Saudi Arabia (KSA)

Hassan M. Aljohani and Azhari A. Elhag 

Department of Mathematics and Statistics, College of Science, Taif University, P.O. Box 11099, Taif 21944, Saudi Arabia

Correspondence should be addressed to Azhari A. Elhag; azhri_elhag@hotmail.com

Received 22 January 2021; Revised 10 February 2021; Accepted 26 February 2021; Published 22 March 2021

Academic Editor: Ahmed Mostafa Khalil

Copyright © 2021 Hassan M. Aljohani and Azhari A. Elhag. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

Classification in statistics is usually used to solve the problems of identifying to which set of categories, such as subpopulations, new observation belongs, based on a training set of data containing information (or instances) whose category membership is known. The article aims to use the Gaussian Mixture Model to model the daily closing price index over the period of 1/1/2013 to 16/8/2020 in the Kingdom of Saudi Arabia. The daily closing price index over the period declined, which might be the effect of corona virus, and the mean of the study period is about 7866.965. The closing price is the last regular deal that took place during the continuous trading period. If there are no transactions on the stock during the day, the closing price is the previous day's closing price. The closing auction period comes after the continuous trading period (from 3:00 PM to 3:10 PM), during which investors can enter by buying and selling the stocks at this period. The experimental results show that the best mixture model is E (equal variance) with three components according to the BIC criterion. The expectation-maximization (EM) algorithm converged in 2 repetitions. The data source is from Tadawul KSA.

1. Introduction

The stock market index's direction indicates the movement of the price index or the future trend of fluctuation in the stock market index [1, 2]. Guessing the trend is a practical issue that heavily influences a financial trader buying or selling an instrument [3, 4]. An accurate forecast of the stock index trends can help investors acquire opportunities for gaining profit in the stock exchange [5–7]. Hence, precise forecasting of the stock price index trends can be extremely advantageous for investors [8]. It is essential to study the extent to which the stock price index's movement can be predicted using the data Tadawul from emerging markets such as the Saudi stock market, since its inception on 6 June 2003, corresponding to 2/6/1424 AH. On March 19, 2007, the Council of Ministers approved it [3]. The Saudi Stock Exchange Company “Tadawul” under Article 20 of the Capital Market Authority. The passage of years is involved the incredible expansion of the local economy and companies which need to reach a wide range of investors. It obtained the

total of listings in the market. The leading and parallel 262 companies and securities and debt instruments as many as 190 companies were listed in the leading market and ten firms in growth-parallel market and 62 instrument issues' debt by the end of 2018. In addition to listing the shares, Tadawul is also listed trading in bonds and Sukuk and funds (RITs) [9] and index traded funds (ETF). The leading market index (TASI) also includes all companies listed on the leading “Tadawul” market, one of the leading indicators trusted by investors. It depends on the performance of the companies listed in the stock market in Saudi Arabia. Tadawul also publishes different types of sector indices following the Global Industry Classification Standard (GICS).

2. Gaussian Mixture Model

In this section, we introduced mixture models. Recall that, if our observations X_i come from a mixture model with K mixture components, the marginal probability distribution of X_i is of the form

$$p(X_i = x) = \sum_{k=1}^K \pi_k p(x_i = x | Z_i = k), \quad (1)$$

where $Z_i \in \{1, \dots, K\}$ is the latent variable representing the mixture component for X_i , $P(X_i | Z_i)$ which is the mixture component and π_k is the mixture proportion representing the probability that X_i belongs to the k th mixture component [10].

2.1. Expectation Maximization (EM). It is an algorithm within the Gaussian mixture models. Consider $N(\mu, \sigma^2)$ represents the probability distribution function for a normal random variable. Thus, we get that the conditional distribution $X_i | Z_i = k \sim N(\mu_k, \sigma_k^2)$ so that the marginal distribution of X_i is

$$\begin{aligned} p(X_i = x) &= \sum_{k=1}^K p(Z_i = k) p(X_i = x | Z_i = k) \\ &= \sum_{k=1}^K \pi_k N(x, \mu_k, \sigma_k^2). \end{aligned} \quad (2)$$

Similarly, the joint probability of observations X_1, \dots, X_n is therefore

$$P(X_1 = x_1, \dots, X_n = x_n) = \prod_{i=1}^n \sum_{k=1}^K \pi_k N(x_i, \mu_k, \sigma_k^2). \quad (3)$$

See [11], for more details. This note defines the EM algorithm that aims to determine the maximum likelihood estimates of π_k, μ_k , and σ_k^2 given a dataset of observations $\{x_1, \dots, x_n\}$.

2.2. MLE of Normal Distribution. Suppose we have mn observations X_1, \dots, X_n from a Gaussian distribution with an unidentified mean μ and a recognized variance σ^2 . To define the maximum likelihood estimate for μ , we get the log-likelihood $\ell(\mu)$ to take the derivative concerning μ , set it equal zero, and resolve for μ :

$$\begin{aligned} L(\mu) &= \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma^2}} \exp - \frac{(x_i - \mu)^2}{2\sigma^2} \\ \implies \ell(\mu) &= \sum_{i=1}^n \left[\log \left(\frac{1}{\sqrt{2\pi\sigma^2}} \right) - \frac{(x_i - \mu)^2}{2\sigma^2} \right] \\ \implies \frac{d}{d\mu} \ell(\mu) &= \sum \frac{x_i - \mu}{\sigma^2}. \end{aligned} \quad (4)$$

Defining the result equal to zero and resolving for μ , we have that $\mu_{MLE} = (1/n) \sum_{i=1}^n x_i$. Furthermore, applying the log function to the likelihood helped decompose the product and eliminated the exponential function. Thus, the MLE could be resolved easily.

2.3. MLE of Gaussian Mixture Model. Now, we attempt the same strategy for deriving the MLE of the Gaussian mixture model. Our unknown parameters are

$$\theta = \{\mu_1, \dots, \mu_k, \sigma_1, \dots, \sigma_k, \pi_1, \dots, \pi_k, \quad (5)$$

based on the first section of the note, and our likelihood is

$$L(\theta | X_1, \dots, X_n) = \prod_{i=1}^n \sum_{k=1}^K \pi_k N(\pi_k; \mu_k, \sigma_k^2). \quad (6)$$

So, our log-likelihood is

$$\ell(\theta) = \sum_{i=1}^n \log \sum_{k=1}^K \pi_k N(\pi_k; \mu_k, \sigma_k^2). \quad (7)$$

Considering the expression above, we already see a difference between this scenario and the simple setup in the preceding section. The summation over the k constituents "blocks" our log function from application to the ordinary densities. When following the same earlier steps, differentiating concerning μ_k and setting the expression equal to zero, the result would be

$$\sum_{i=1}^n \frac{1}{\sum_{k=1}^K \pi_k N(\pi_k; \mu_k, \sigma_k^2)} \pi_k N(x_k, \mu_k, \sigma_k) \frac{(x_k - \mu_k)}{\sigma_k^2} = 0. \quad (8)$$

We are currently stuck due to the inability to resolve analytically for μ_k , but a significant observation is made when we defined the latent variables Z_i . After that, we could collect all samples X_i such that $Z_i = k$ and utilize the estimate from the preceding section to estimate μ_k .

3. Numerical Results

The available historical data consisted of the daily closing price index over the period of 1/1/2013 to 16/8/2020 in the Kingdom of Saudi Arabia. The data source is from Tadawul KSA [12].

Figure 1 displays the data that is taken daily over 1/1/2013 to 16/8/2020. It can be seen that the data contains numerous of information. However, the proposed methodology can answer questions that we need it. On the contrary, Table 1 shows the summary statistics. This information can be obtained quickly from any software. From Table 1, the mean is 7866.965 and the standard deviation is 1126.780, where the first impression can be given from this primary information.

The comparison among these models can be found using BIC in Table 2, where this method is used in the maximum likelihood and then compared between them. It can be seen that the model number 3 has the smallest figure, although the model number 2 gives negative result.

Since mixed model is used, it is beneficial to study the proportion between two normal distributions, where the model numbers 1 and 2 have the same number of the balance and larger than model 3, see Table 2.

Table 3 shows the means after the proposed methods are implemented. It is easy to see that the model number 3 gives a

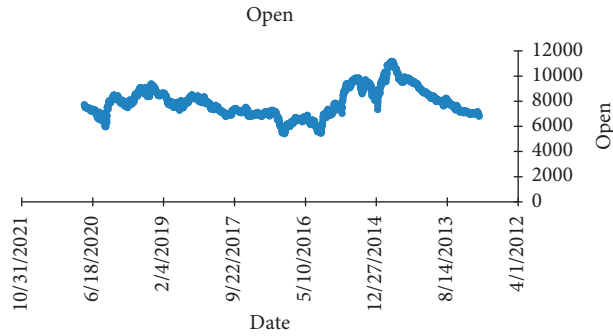


FIGURE 1: The series for the daily closing price index over the period of 1/1/2013 to 16/8/2020.

TABLE 1: The descriptive statistics for the daily closing price index over the period of 1/1/2013 to 16/8/2020.

Summary statistics							
Std. Deviation	Mean	Maximum	Minimum	Obs. without missing data	Obs. with missing data	Observations	Variable
1126.780	7866.965	11149.360	5416.470	1899	0	1899	The daily closing price index

TABLE 2: Selection criterion is Bayesian information criterion.

Evolution of the BIC for each model				
5	4	3	2	The daily closing price index over <i>E</i>
-32040.623	-32025.525	-32010.427	-32107.323	

TABLE 3: The mean by the three components.

Means by class			
3	2	1	Class
9232.995	7385.067	7385.066	Mean (open)

TABLE 4: Selection criterion for selection model. The NEC criterion is more than one which indicates that there is no clustering structure in the data.

Selection criterion for the selected model						
DF	Entropy	NEC	Log-likelihood	ICL	AIC	BIC
6.000	1424.349	25.436	-15982.566	-34859.124	-31977.132	-32010.427

high figure of mean, where model numbers 1 and 2 are closer to 7385, though the variances are similar in the three models.

In contrast, the selection criterion can be used to find the class in the data. It shows that there is more than one clustering in this data, where the method gives 25.436. This means that two or three groups can be given (see Table 4).

Figure 2 shows that there are two clusters: the red line shows first groups and the green shows the other. This means that two groups of companies can be gathered together. More precisely, from the data there are one group going down and the other going up. It is not easy to find this information from the data directly.

The plots of three model can be found in Figure 3, where the mixed model gives the larger maximum likelihood than the other models “1, 2.”

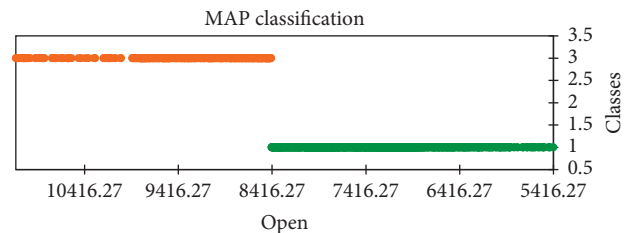


FIGURE 2: The MAP classification shows that there is no assignment to classes 2.

It can be seen that the proposed method gives curve close to the empirical components. This means that the proposed method gives excellent results, even compared to various models, see Figures 4 and 5.

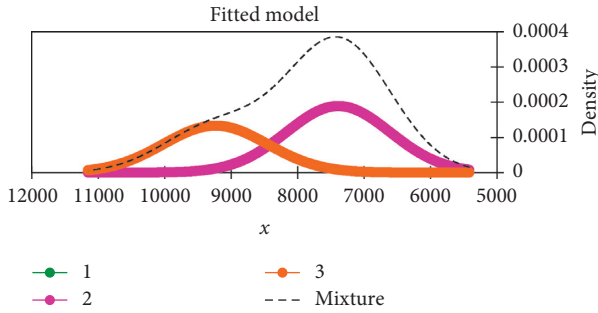


FIGURE 3: The mixture model.

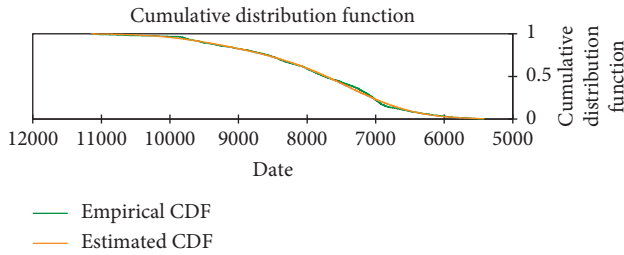


FIGURE 4: The mixture model. It is clear that the estimated CDF is very close to empirical CDF, which confirms the accuracy of the estimation.

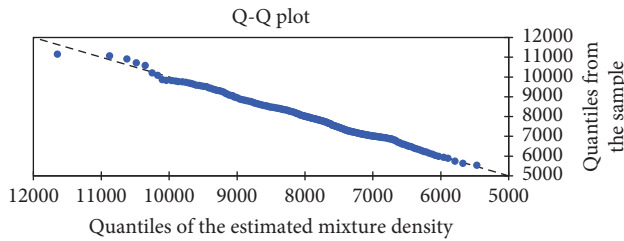


FIGURE 5: The quintiles of the estimated mixture density.

TABLE 5: The proportions of three components.

Proportions			
3	2	1	Class
0.261	0.370	0.370	Proportions

TABLE 6: The variance by the three components.

Variance by class			
3	2	1	Class
610676.444	610676.444	610676.444	Variance

4. Conclusion

This paper used the Gaussian mixtures model to classify the daily closing price index over 1/1/2013 to 16/8/2020 in KSA and describe the problem of predicting the daily closing price index in KSA which represents a huge problem. The

mean of the daily closing price index over the study period is 7866.965. The decline of the daily closing price index in KSA, which occurred last year, might have been due to COVID-19 pandemic. The EM algorithm converged in 124 iterations; according to Bayesian information criterion, the best mixture model is the equal variance with three components (see Table 2). The proportions of the three ingredients are varied between 0.261 and 0.370 (see Table 5). The mean of the three components is run between 9232.995 and 7385.066 (see Table 3). The variance of the four ingredients is 610676.444 (see Table 6). Finally, we must point out that implementing such a mechanism to predict the daily closing price index in the KSA is beneficial.

Data Availability

No data were used to support the findings of this study.

Conflicts of Interest

The authors declare that there are no conflicts of interest related to this article.

Acknowledgments

This work was supported by Taif University Researchers Supporting Project (no. TURSP2020/279), Taif University, Taif, Saudi Arabia.

References

- [1] G. Hu and J. Geng, "Heterogeneity learning for SIRS model: an application to the COVID-19," 2020, <http://arxiv.org/abs/2007.08047v1>.
- [2] M. Qiu and Y. Song, "Predicting the direction of stock market index movement using an optimized artificial neural network model," *PLoS ONE*, vol. 11, 2016.
- [3] D. Gholamiangonabadi, S. D. Mohseni Taheri, A. Mohammadi, and M. B. Menhaj, Eds., in *Proceedings of the 2014 5th Conference on Thermal Power Plants*, pp. 75–82pp. 75–, Tehran, Iran, June 2014.
- [4] M. A. I. Khan, "Financial volatility forecasting by nonlinear support vector machine heterogeneous autoregressive model: evidence from Nikkei 225 stock index," *International Journal of Finance & Economics*, vol. 3, no. 4, pp. 138–150, 2014.
- [5] Z. Wang, D.-W. Zhao, L. Wang, G.-Q. Sun, and Z. Jin, "Immunity of multiplex networks via acquaintance vaccination," *EPL (Europhysics Letters)*, vol. 112, no. 4, pp. 48002–48007, 2015.
- [6] Z. Wang, M. A. Andrews, Z.-X. Wu, L. Wang, and C. T. Bauch, "Coupled disease-behavior dynamics on complex networks: a review," *Physics of Life Reviews*, vol. 15, pp. 1–29, Article ID 2621171, 2015.
- [7] M. Qiu, Y. Song, and F. Akagi, "Application of artificial neural network for the prediction of stock market returns: the case of the Japanese stock market," *Chaos, Solitons & Fractals*, vol. 85, pp. 1–7, 2016.
- [8] T. Fischer and C. Krauss, "Deep learning with long short-term memory networks for financial market predictions," *European*

Journal of Operational Research, vol. 270, no. 2, pp. 654–669, 2018.

- [9] [http://file:///C:/Users/DELL/Downloads/Llisting-Guide-2019%20\(1\).pdf](http://file:///C:/Users/DELL/Downloads/Llisting-Guide-2019%20(1).pdf).
- [10] https://stephens999.github.io/fiveMinuteStats/intro_to_em.html#mle_of_Gaussian_mixture_model.
- [11] Z. Yao, B. Zhao, L. Qin, Y. Jiang, B. Ran, and Bo Peng, “An efficient heterogeneous platoon dispersion model for real-time traffic signal control,” *Physica A: Statistical Mechanics and Its Applications*, vol. 539, 2020.
- [12] <https://www.tadawul.com.sa/wps/portal/tadawul/home/>.