Time Series Behaviour of the Number of Air Asia Passengers: A Distributional Approach

Conference Paper · February 2013
DOI: 10.1063/1.4823977

2 authors:

Norhaidah Mohd Asrah
Universiti Tun Hussein Onn Malaysia
5 PUBLICATIONS  0 CITATIONS

SEE PROFILE

Maman Djauhari
Universiti Putra Malaysia
77 PUBLICATIONS  160 CITATIONS

SEE PROFILE

All in-text references underlined in blue are linked to publications on ResearchGate, letting you access and read them immediately.

Available from: Norhaidah Mohd Asrah
Retrieved on: 31 July 2016
Time Series Behaviour of the Number of Air Asia Passengers: A Distributional Approach

Norhaidah Mohd Asrah¹ and Maman Abdurachman Djauhari²

¹Department of Science and Mathematics, Faculty of Science, Technology and Human Development, Universiti Tun Hussein Onn Malaysia
²Department of Mathematical Science, Faculty of Science, Universiti Teknologi Malaysia

Abstract. The common practice to time series analysis is by fitting a model and then further analysis is conducted on the residuals. However, if we know the distributional behavior of time series, the analyses in model identification, parameter estimation, and model checking are more straightforward. In this paper, we show that the number of Air Asia passengers can be represented as a geometric Brownian motion process. Therefore, instead of using the standard approach in model fitting, we use an appropriate transformation to come up with a stationary, normally distributed and even independent time series. An example in forecasting the number of Air Asia passengers will be given to illustrate the advantages of the method.

Keywords: Geometric Brownian Motion, time series
PACS: Statistics, 02.50.

INTRODUCTION

The time series is a sequence of real numbers, collected at a regular intervals over a period of time. Time series data are widely used in the fields of financial market, medicine, biology, chemistry, computer, physics and many more. It is also ubiquitous and pervasive in science, engineering, business and other areas.

Generally, the time series analysis will begin by fitting a model, then the analysis will be conducted on the residuals. The distributional behavior of time series is very important. If we know the distribution of the time series, the analyses in model identification, parameter estimation, and model checking are much easier.

The geometric Brownian motion (GBM) is one of the method in forecasting share prices in a mathematical model [1]. The criteria in GBM like stationary, normally distributed and independent can be used as standard approach in model fitting.

The future prices for the next two weeks of investment in Bursa Malaysia can be predicted by using GBM. The results shows that this model is highly accurate and can be proven by the mean absolute percentage error (MAPE) and GBM is very suitable for prediction in the short-term investment [1].

The future closing price in Bursa Malaysia can also be forecast by using GBM. In paper by [2], the volatility was calculated by using four different equations. This is to find the better model of GBM for forecasting the stock price. The equations are simple volatility, log volatility, highs and lows volatility and highs-low-close volatility. The results shows that the highs-low-close volatility has the highest accuracy model.

Paper by [3] identified better GBM models in forecasting the indices of FBMKLCI and FBMHS in Bursa Malaysia. The volatility for both indices were calculated using two different equations, simple volatility and log volatility. The results shows that log volatility gives better results compare to simple volatility. The forecast value was highly accurate with the 4 week daily data.

GBM method was also used to predict the coal price [4]. It produced a widest range of predicted prices compare to mean reverting and econometric modeling. GBM process may be appropriate to varied data types, based on the criteria of normality and independence [5]. The examples of data that appropriate to GBM process are electric utility data and the airline passenger data. These data follows GBM assumptions which are stationary, independence and normal.

In this study, we used the data about number of Air Asia passengers to represent it as a GBM process. The checking assumptions of validity for GBM and the forecasting for the number of passengers for the next month period will be discussed.
RESEARCH METHODOLOGY

A. Data Preparation

In this study, the data have been obtained from Malaysia Airport Holdings Berhad (MAHB) Sepang, Selangor. The data obtained were about the number of passengers departure from Kuala Lumpur International Airport (KLIA) to the international destinations daily. There are two set of time series daily data used in this study. Firstly, data for the number of Air Asia passengers from January 2009 until March 2009. Secondly, data for the number of Air Asia passengers from January 2012 until May 2012.

B. Data Analysis

1. Test of GBM Assumptions

According to [5], there are two assumptions must be satisfied for GBM process. There are normality of the log ratios and the independence from the previous data. Both sets of data were tested for GBM process fit.

The two important approach for checking the distribution assumptions are empirical procedures and statistical procedures [6]. The empirical procedures are based on graphical properties of the distribution. This procedures can be used to check and validate the distribution assumptions. While the statistical procedures are the Goodness of Fit tests. The results for Goodness of Fit tests are quantifiable and more reliable from the empirical procedures.

In this study, the QQ plot is used to check for normality. The data is from normal distribution if the points of the plot lie close to a straight line. Then statistical test of normality was conducted by using the Anderson-Darling (AD) test, one of the Goodness of Fit tests. AD is the best distance test for small samples and also can be used for large samples. The results from AD test in the Table 1 below and QQ plot in Figure 1 shows that the data for both sets are normally distributed at 5% significance level.

| Table 1. Anderson-Darling Test of Normality |
|-----------------|----------|----------|
| Year            | AD value | p-value  |
| Air Asia 2009   | 0.407    | 0.342    |
| Air Asia 2012   | 0.330    | 0.513    |

The scatter plot can be used to visualized the relationship between data. If the data scattered and do not show any pattern, it can be concluded that probably no or weak relationship between data. To test the independency of the previous data, the statistics $r$ or called the correlation coefficient of the random sample can be used. The results from correlation test in the Table 2 below and scatter plot in Figure 2 shows that the data for both sets are independent at 5% significance level.

| Table 2. Correlation Coefficient Test of Independency |
|-----------------|---------|---------|
| Year            | Air Asia 2009 | Air Asia 2012 |
| Pearson         | 0.2145  | 0.0173  |
| $R^2$           | 0.046   | 0.0003  |
| $t$-distribution| 2.2799  | 2.2605  |
| $p$-value       | 0.025   | 0.025   |
The other important assumption is stationary. It is essentially for time series to be stationary so that we can develop models and forecasting [7]. A stationary time series can be defined as a finite variance process which the mean and variance are constant in time. The correlation between observations from different points in time is only lag dependent. Figure 3 shows the stationary plot for both data sets.

The Geometric Brownian Motion (GBM) also known as lognormal growth process has widely accepted as a valid model for the growth in the price of stock over time [8]. GBM is a mathematical approach for stock market modeling. GBM is a stochastic process with assumption that returns, profits or losses on the stock are independent and normally distributed [4].

According to [9], the GBM model implies that

\[ L(n) = a + L(n-1) + \varepsilon(n) \]  

where \( L(n) = \log(Sd(n)) \), \( \varepsilon(n) \), \( n \geq 1 \), is a sequence of independent and identically distributed with mean 0 and variance \( \sigma^2 / n \) and \( a \) is equal to \( \mu / N \). In GBM, \( \mu \) is the mean (drift) parameter and \( \sigma \) is the associated volatility parameter.

Equation (1) consider fitting a more general equation for \( L(n) \), the linear regression equation

\[ L(n) = a + bL(n-1) + \varepsilon(n) \]  

where \( b \) is a constant value need to be estimated. Equation (2) is the classical linear regression model and the techniques to estimate the parameters \( a \), \( b \) and \( \sigma \) is known. Equation (2) is called the autoregressive model of order 1 (AR 1) since the log price at time \( n \) in terms of the log price one time period earlier.

The parameter \( a \) and \( b \) of the autoregressive model given in Equation (2) are estimated from historical data. Let \( L(0), L(1), \ldots, L(r) \) are the logarithms of the end-of-day price for \( r \) successive days. When \( a \) and \( b \) are known, the predicted value of \( L(i) \) based on prior log price is

\[ L(i) = a + bL(i-1) \]  

The log return from day \( i \) to day \( i+1 \) is follow

\[ R_i = Sd(n+1) / Sd(n) = L(n+1) / L(n) \]  

3. Mean Absolute Percentage Error (MAPE)

The mean absolute percentage error (MAPE) is the forecasting method that widely used which attempt to consider the effect of the magnitude of the actual values [10]. The formula for MAPE is as follows:

\[ MAPE = \frac{1}{n} \sum \left| \frac{Y_i - F_i}{Y_i} \right| \]  

where \( Y_i \) is the actual value and \( F_i \) is the forecast value. The lower the value of MAPE, the more accurate the forecast. By using this formula and applying the Lewis’s scale provide some framework to judge the model as below:

2. Geometric Brownian Motion (GBM)

The Geometric Brownian motion (GBM) also known as lognormal growth process has widely accepted as a valid model for the growth in the price of
TABLE 3. A Scale of Judgment of Forecast Accuracy (Lewis)

<table>
<thead>
<tr>
<th>MAPE</th>
<th>Judgment of Forecast Accuracy</th>
</tr>
</thead>
<tbody>
<tr>
<td>Less than 10%</td>
<td>Highly accurate</td>
</tr>
<tr>
<td>11% to 20%</td>
<td>Good forecast</td>
</tr>
<tr>
<td>21% to 50%</td>
<td>Reasonable forecast</td>
</tr>
<tr>
<td>More than 51%</td>
<td>Inaccurate forecast</td>
</tr>
</tbody>
</table>

RESULTS AND DISCUSSION

The first set of data in this study contained information for number of passengers from January 2009 until March 2009. The next 30 observations (April 2009) were used as the estimated values for number of passengers for this period. Another set of data contained information for number of passengers from January 2012 until May 2012. The next 29 observations (Jun 2012) were used as the estimated values for number of passengers for this period. Figure 3 and Figure 4 shows the comparison between forecasting number of passengers using GBM with the actual number of passengers for 1 month data.

The forecast value for number of passengers in Figure 3 shows that the values are not very close to the actual number of passengers. Meanwhile, the forecast value for number of passengers in Figure 4 shows that the values are close to the actual number of passengers almost for the whole month.

TABLE 4. MAPE Values

<table>
<thead>
<tr>
<th>Time</th>
<th>MAPE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Air Asia 2009</td>
<td>11.05%</td>
</tr>
<tr>
<td>Air Asia 2012</td>
<td>4.5%</td>
</tr>
</tbody>
</table>

If we look at the MAPE value in Table 3 above, it indicates that the MAPE value for number of Air Asia passengers in 2012 is smaller than number of Air Asia passengers in 2009. The MAPE value for Air Asia 2012 is highly accurate than Air Asia 2009 since the error is less than 10%. The forecast value was highly accurate with the 5 months daily data rather than 3 months daily data.

FIGURE 3. Comparison between the forecast number of passengers and actual number of passengers for Air Asia company in year of 2009.

FIGURE 4. Comparison between the forecast number of passengers and actual number of passengers for Air Asia company in year of 2012.
CONCLUSION

This study shows that data for the number of Air Asia passengers can be represented as GBM process. By using an appropriate transformation, these data set successfully satisfied the GBM assumptions like stationary, normally distributed and even independent time series. The number of passengers for the next month can also be forecast using GBM.

ACKNOWLEDGMENTS

This research is sponsored by the Ministry of Higher Education, Government of Malaysia, through Fundamental Research Grant Schemes vote number 4F014 and Research University Grant vote number Q.J1300.2526.02H18. The author gratefully acknowledges the Ministry of Higher Education for the sponsorships and Universiti Teknologi Malaysia for the opportunity to conduct this research. The author also acknowledges the Research and Planning Department of Malaysia Airport Holdings Berhad (MAHB) for the data provided in this research. Special thanks go to the anonymous referees for their constructive comments and suggestions.

REFERENCES