Malaysian Journal of Mathematical Sciences 11(S) April : 107–118 (2017)
Special Issue: The 2nd International Conference and Workshop on Mathematical Analysis (ICWOMA 2016)

MALAYSIAN JOURNAL OF MATHEMATICAL SCIENCES

Journal homepage: http://einspem.upm.edu.my/journal

Box-Cox Transformation of Monthly Malaysian Gold Price Range

Gopal, K. *[1], Abdul Rahim, M. F.[1], and Adam, M. B.[1,2]

[1]*Institute for Mathematical Research, Universiti Putra Malaysia, Malaysia*
[2]*Department of Mathematics, Faculty of Science, Universiti Putra Malaysia, Malaysia*

E-mail: kathiresan@upm.edu.my
* *Corresponding author*

ABSTRACT

Gold has a long history in trading over the past decades until its role in trading was replaced by the introduction of banknotes and coins. Gold is undeniably one of the prime commodities in investments and as well as a hedging tool. Thus, it is important to study the fluctuation in monthly gold price in Malaysia to extract useful information that could benefit investors and the government. In this paper, Box-Cox transformation was applied on the original data to deal with the presence of heteroscedasticity, non-normality and outliers. Monthly gold price data for the period of 2002 to 2011 were obtained to study its fluctuation using range as a measure of dispersion. It is found that the Malaysian gold price range is best described by the Normal distribution with parameters estimated using Maximum Likelihood Estimation.

Keywords: Gold, Malaysian gold price range and Box-Cox transformation.

1. Introduction

Gold has a long history in trading over the past decades. In Malaysia, gold was one of the natural resources and besides trading, gold mining was also active back then. Gold was one of the important mediums of trading utilised until the development of banknotes and coins took over its role.

Sukri et al. (2015) stated that gold is a good investment that is neither short-term nor long-term and one of the privileges of gold is that it is considered valuable by the entire human race. It is also a great asset that can be converted into paper money. No other assets can match the liquidity of gold. Gold is a valuable item that is easy to handle and is a tangible value item. Worthington and Pahlavani (2007) pointed out that gold has attracted people since thousands of years ago until recent because unlike most other commodities, gold is a metal that is durable, easy to carry, as well as universally accepted and validated.

According to Zhang and Wei (2010), gold is one of the hedging tools. The liquidity of the gold value in the future is more efficient rather than keeping stock or paper money. The role of gold as a hedge means that it helps offset the significant losses incurred by the investors. In simple words, the hedge is to reduce the risk of loss on an investment or depreciation of an asset's value. Gold is not only regarded as a commodity but also classified as a financial asset. Economic instability cannot be predicted thus causing investors to take the initiative to make gold as an asset to the risks incurred from getting too high as pointed out by Bernstein (2012). In addition to hedge, gold is a store of wealth, which means safe storage.

The price of gold in Malaysia depends on the global market consensus and regulated by the Central Bank of Malaysia (BNM). In this study, we have used the trading prices (in ringgit) of Malaysia's gold bullion coin, the *Kijang Emas*. Besides the *Kijang Emas*, elements of gold such as gold jewellery, gold bar, or gold ore could also be used to study the variation in gold price as every element of gold has different values depending on the supply and demand as well as the regulation of the country, see Sukri et al. (2015).

Data of Malaysian gold price (MGP) based on the trading prices of *Kijang Emas* for the years 2002 to 2011 were obtained from the official website of BNM, (*www.bnm.gov.my*). The data were in the form of daily trading prices of per oz gold for each months throughout the 10 years.

Gold prices are often recorded on daily basis due to its natural fluctuation. This process yields time series data which are usually compiled for months or years by averaging the gold prices over the period. However, it is important to note that there will be no record(s) of gold price for certain day(s). This is a common occurrence due to trade market closure. As such, the monthly or yearly compilation would not be appropriate because the missing values in monthly data will affect the computation of monthly means.

In order to avoid the issue of missing values in monthly gold price, we may use the maximum or minimum gold price of a month yet these extreme values might not be sufficient to explain the monthly gold price fluctuation as only one of the extremes are taken into consideration. Eventually, the maximum minimum approach would not be able to indicate the variation present in the monthly gold price. Hence, range is found to be a better tool to depict the fluctuation in the monthly gold price. Studying the fluctuation helps to visualise the variation pattern in monthly MGP. Furthermore, the obtained monthly MGP range data were found to exhibit anomalies such as heteroscedasticity, non-normality and presence of outliers which are described in detail below.

Heteroscedasticity describes a situation in which the error term in the relationship between the independent variables and the dependent variable is the not the same across all values of the independent variables (non-constant error variance), see Lyon and Tsai (1996). Heteroscedasticity in our data is present when the size of the error term from Malaysian gold price range quantified as variance differs across values of time (months). Presence of heteroscedasticity causes several problems such as affecting the parameter estimation to be inefficient and making the standard error to be biased as mentioned in Lyon and Tsai (1996). One approach for dealing with heteroscedasticity is to transform the dependent variable i.e. Malaysian gold price range using any variance stabilizing transformations such as the Box-Cox transformation.

Normality is an important concept in statistics. A random variable is said to possess normality if it is normally distributed. Yazici and Yolacan (2007) stated that normality is vital as many advanced statistical theories rely on the observed data possessing normality such as the t-test. Normality can be identified using visual inspection with Normal Probability Quantile-Quantile (Normal Q-Q) plot and to affirm the inspection we should make use of statistical tests for normality such Shapiro-Wilk test and Jarque Bera test. The probability plot and normality tests are usually the best tools for judging normality, especially for smaller sample sizes.

Hogg et al. (2013) explained that an outlier is an observation which devi-

ates so much from the other observations as to arouse suspicions that it was generated by a different mechanism. Outliers do not fit the general trend of the data and its presence can skew or change the shape of data distribution and eventually affects the summary statistics to be inaccurate, see Hogg et al. (2013). The common approach to identify outliers is using the boxplot.

In view of the above discussion, the primary objective of this study is to eliminate the heteroscedasticity, non-normality and outliers issues present in the original monthly MGP range data by employing the Box-Cox transformation. The secondary objective is to estimate the parameters of the transformed data distribution for which the Normal distribution best describes the data using Maximum Likelihood Estimation (MLE).

2. Range as a Tool for Measuring Data Dispersion

Range is a measure of data dispersion or spread besides variance, standard deviation and interquartile range. Range is used in this study to describe the monthly MGP fluctuation. Range is defined as the difference between maximum and minimum values. Although range uses only extreme values, it still preserves the metric as the original data unlike in the variance computation.

Using range of monthly MGP not only avoided the problem of missing values in the monthly gold prices but also provided a way to encompass both the extreme values of monthly MGP. This approach would provide more insight when studying the fluctuation of monthly MGP.

From the monthly datasets comprising of daily MGP, maximum and minimum MGP for each month were obtained and the range for each month was computed and finally compiled to a dataset consisting of 120 observations corresponding to ranges of monthly MGP of 10 years in study. We define $Y = (y_1, y_2, \ldots, y_{119}, y_{120})$ as the ranges of monthly MGP in sequence from January 2002 to December 2011, where y_i denotes the range of gold price of the i^{th} month.

2.1 Box-Cox Transformation

The power transformation introduced by Tukey in 1977, see Sakia (1992) can be very effective when the relationship between independent and dependent variable is simple monotone i.e. either strictly increasing or strictly decreasing

with no inflection point. This data transformation technique is useful to reduce anomalies such as non-linearity, heteroscedasticity (stabilizing variance) and non-normality as well as skewness. Following Sakia (1992), the power transformation with the power parameter λ is defined in Equation (1).

$$y_i^{(\lambda)} = \begin{cases} y_i^{\lambda} & \lambda \neq 0 \\ \log y_i & \lambda = 0 \end{cases} \tag{1}$$

The simple power transformation has a discontinuity problem at $\lambda = 0$. This gives rise to the Box-Cox transformation as defined in Equation (2).

$$y_i^{(\lambda)} = \begin{cases} \dfrac{y_i^{\lambda} - 1}{\lambda} & \lambda \neq 0 \\ \log y_i & \lambda = 0 \end{cases} \tag{2}$$

When transforming the dependent variable and trying to find the best value of λ in the Box-Cox transformation, there is an additional problem. After transforming the dependent variable, the scores are no longer in their original metric, see Sakia (1992). Consequently the residual sum of squares no longer has the same statistical meaning as it did prior to transformation. As a result, we cannot find the best λ by comparing the residual variance or residual sum of squares for several competing values of λ. Sakia (1992) further stated that a solution to this problem is introduced by Box and Cox through the Standardized Box-Cox Transformation which incorporates the geometric mean of the dependent variable say Y, denoted as $\bar{g}_Y$ to simplify the derivation of a maximum-likelihood method as given in Equation (3).

$$y_i^{(\lambda)} = \begin{cases} \dfrac{y_i^{\lambda} - 1}{\lambda \bar{g}_Y^{\lambda - 1}} & \lambda \neq 0 \\ \bar{g}_Y \log y_i & \lambda = 0 \end{cases} \tag{3}$$

where $\bar{g}_Y = (\prod_{i=1}^{n} y_i)^{\frac{1}{n}}$ and it follows that $\log \bar{g}_Y = \dfrac{1}{n} \sum_{i=1}^{n} \log y_i$.

The power parameter λ can be specified or estimated using Maximum Likelihood Estimation to obtain an optimal λ that minimizes the sum of squares of prediction (SSE) of the transformed distribution or maximizes the log-likelihood function from a range of λ values usually in the interval [-2,2] as

given in Sakia (1992). In this study, we opt for the latter.

It is important to verify the presence of heteroscedasticity on the transformed data by visual inspection and drawing conclusion using the statistical tests available for this purpose. In this paper, we employ the studentized Breusch-Pagan and Non-constant Variance Score tests, see Lyon and Tsai (1996) with the null hypothesis tested against the alternative hypothesis as stated below:

H_0: *Homoscedasticity (no heteroscedasticity) is present in the MGP range.*
H_1: *Heteroscedasticity is present in the MGP range.*

Yazici and Yolacan (2007) asserted that normality tests are supplementary to the graphical assessment of normality. The tests usually compare the scores in the sample to a normally distributed set of scores with the same mean and standard deviation. If the test is significant, the distribution is non-normal. For small sample sizes, normality tests have little power to reject the null hypothesis and therefore small samples most often pass normality tests whereas for large sample sizes, significant results would be derived even in the case of a small deviation from normality although this small deviation will not affect the results of a parametric test, see Yazici and Yolacan (2007). We used the popular omnibus test, Shapiro-Wilk and the robust version of Jarque Bera Test i.e. Robust Jarque Bera tests in this paper. The null hypothesis and alternative hypothesis for these tests are as below:

H_0: *MGP range is normally distributed.*
H_1: *MGP range is not normally distributed.*

Parameter estimation is an essential step in obtaining the sample estimates of the population parameters concerned in a probability distribution. The method of maximum likelihood corresponds to many well-known estimation methods in statistics. It seems reasonable that a good estimate of the unknown parameter θ would be the value of θ that maximizes the probability that is, the likelihood of getting the data we observed, see Hogg et al. (2013).

If $Y \sim N(\mu, \sigma^2)$, with unknown μ and σ^2, then the log-likelihood function of the Normal distribution that we wish to maximise is given by:

$$\ell(\mu, \sigma^2; y_i) = \left(\frac{-n}{2}\right) \log 2\pi - \frac{n}{2} \log \sigma^2 - \frac{1}{2\sigma^2} \sum_{i=1}^{n} (y_i - \mu)^2$$

Solving the partial derivatives of the above equation with respect to μ and σ^2 yields the MLE estimates:

$$\hat{\mu} = \frac{1}{n}\sum_{i=1}^{n} y_i = \bar{Y}$$

$$\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(y_i - \bar{Y})^2$$

3. Application of Monthly Malaysian Gold Price (MGP) Range Data

Figure 1 (top left) shows the scatterplot of monthly MGP range for the period of 120 months depicting the non-constant fluctuation pattern or inconsistent variation in the monthly MGP range indicating presence of heteroscedasticity; Figure 1 (top right) is the Normal Q-Q plot that clearly shows the depart from normality for monthly MGP range indicating uneven or asymmetrical distribution of the fluctuations; Figure 1 (bottom left) is the histogram showing that the distribution of monthly MGP range is rather skewed to the right and Figure 1 (bottom right) is the boxplot displaying the existence of outliers which are extreme fluctuations of monthly MGP range.

Based on Figure 1, we can infer that a data transformation is essential in order to describe the variation in the monthly MGP range. Figure 2 is the plot of Sum of Squared Errors of Prediction (SSE) against the possible values of λ in the interval [-2,2]. It can be observed that $\lambda = -0.1$ is the optimal value that has the lowest SSE of 503954.2, however the SSE of $\lambda = -0.1$ is pretty close to that of $\lambda = 0$ with SSE = 506822.1 which yields a difference of 0.5% only.

Thus, it makes more practical sense to pick $\lambda = 0$ i.e. the log transformation. As suggested by Hogg et al. (2013), strict minimization of SSE may not often carry more meaning than practical or scientific explanations or even standard conventions. Hence, the Box-Cox transformation is performed using Equation (3) with $\lambda = 0$. We denote the transformed monthly MGP range as $X = (x_1, x_2, \ldots, x_{119}, x_{120})$ where $x_i = \bar{g_Y} \log y_i$.

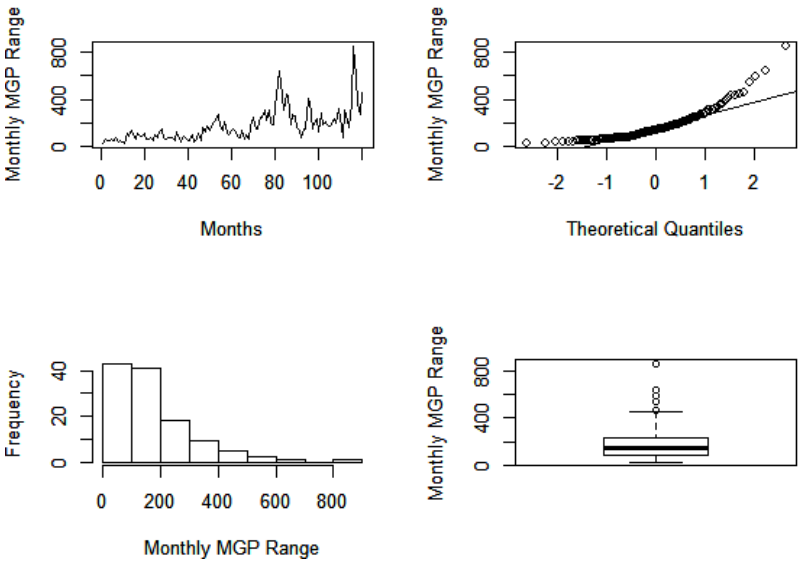

Figure 1: Top Left: Scatterplot of monthly MGP range from January 2002 to December 2012; Top Right: Normal Q-Q plot of monthly MGP range; Bottom Left: Histogram of monthly MGP range; Bottom Right: Boxplot of monthly MGP range.

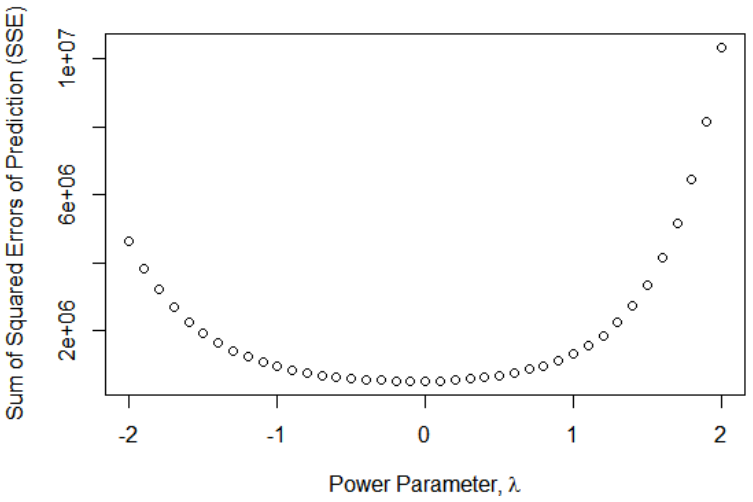

Figure 2: Plot of Sum of Squared Errors of Prediction (SSE) against the possible values of λ.

Figure 3 (top left) is the scatterplot of transformed monthly MGP range for the period of 120 months depicting the absence of heteroscedasticity or presence of homoscedasticity with approximately constant variance pattern for the

monthly MGP range; Figure 3 (top right) shows the Normal Q-Q plot suggesting normality for the transformed monthly MGP range; Figure 3 (bottom left) is the histogram depicting that the distribution of transformed monthly MGP range is almost symmetrical indicating that the fluctuations are more evenly distributed as compared the original data and Figure 3 (bottom right) is the boxplot showing outliers or extreme fluctuations are not present anymore.

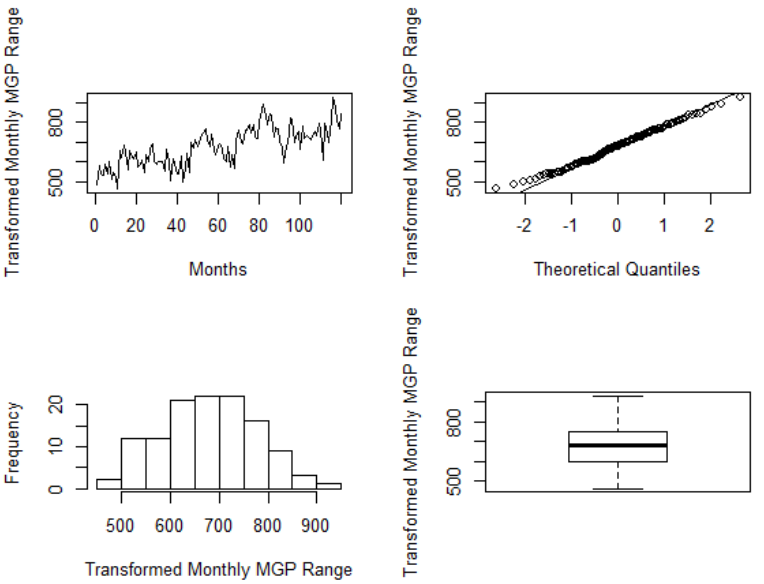

Figure 3: Top Left: Scatterplot of transformed monthly MGP range from January 2002 to December 2012; Top Right: Normal Q-Q plot of transformed monthly MGP range; Bottom Left: Histogram of transformed monthly MGP range; Bottom Right: Boxplot of transformed monthly MGP range.

Table 1 displays the p-values obtained for the normality and heteroscedasticity tests respectively. At $\alpha = 0.05$ level of significance, we can conclude that the transformed monthly MGP range is normally distributed based on the p-values of the Shapiro-Wilk and Robust Jarque Bera tests. On the other hand, based on the p-values for the studentized Breusch-Pagan and Non-constant Variance Score tests, we can affirm that the transformed monthly MGP range is free from heteroscedasticity effect at the same significance level.

Table 1: The p-values of statistical tests conducted in this study.

Test	p-value
Shapiro-Wilk	0.6592*
Robust Jarque Bera	0.4751*
studentized Breusch-Pagan	0.0989*
Non-constant Variance Score	0.1109*

*significant at $\alpha = 5\%$ level

Next, we proceed to estimate the parameters of the transformed monthly MGP range which is normally distributed as denoted by $X \sim N(\mu, \sigma^2)$ with unknown mean μ and variance σ^2. The MLE estimator for the population mean, $\hat{\mu} = \bar{X} = \frac{1}{n}\sum_{i=1}^{n} x_i$ which is the sample mean. As for the population variance, the MLE estimator is the unadjusted sample variance, $\hat{\sigma}^2 = \frac{1}{n}\sum_{i=1}^{n}(x_i - \bar{X})^2$.

The sample mean, $\bar{X}$ is an unbiased estimator of μ since $E[\bar{X}] = \mu$, however the unadjusted sample variance, $\hat{\sigma}^2$ is a biased estimator of σ^2 as $E[\hat{\sigma}^2] \neq \sigma^2$. Hence, we shall employ the adjusted sample variance, S^2, which is an unbiased estimator of σ^2 since $E[S^2] = \sigma^2$. It follows that $S^2 = \frac{n}{n-1}\hat{\sigma}^2 = \frac{1}{n-1}\sum_{i=1}^{n}(x_i - \bar{X})^2$.

The sample mean and sample variance were found to be 678.74 and 9493.25 respectively.The 95% confidence interval (95% CI) for the parameters are given as below:

95% CI for μ: (661.31,696.17)
95% CI for σ^2: (7477.26,12455.37).

Therefore, we can be 95% confident in claiming that the true mean and variance of the transformed monthly MGP range lie within the 95% CI given above. Inference on the original monthly MGP range (Y) can be obtained by expressing the transformed monthly MGP range (X) in its original form i.e. $Y = \exp(\frac{X}{\bar{g_Y}})$.

4. Conclusions

The use of range as a measure of dispersion depicts the variation in monthly gold price fluctuations i.e. showing how high or low the gold price can vary during a month in Malaysia. The original Malaysian gold price range data exhibited heteroscedasticity (inconsistent variation or non-constant fluctuations); non-normality (uneven distribution of fluctuation) and outliers (extreme fluctuations). Thereby, the application of Box-Cox transformation aided in obtaining the transformed monthly MGP range data which eliminated the aforementioned issues, denoted by $X = \bar{g}_Y \log Y$ where Y is the original monthly MGP range.

Further, the Normal distribution with estimated parameters from MLE ($\hat{\mu}$ =678.74, $\hat{\sigma}^2$ =9493.25) was found to best describe the transformed monthly MGP range data. The transformed data can be utilised for further analysis such as time series modelling and forecasting. Inferences can be drawn for the original monthly MGP range by expressing the transformed monthly MGP range in its original unit. In a nutshell, the transformed monthly MGP range data provides a way to describe the fluctuation in monthly MGP which could benefit investors to plan their investment strategies and the government to monitor the gold price mechanism.

Acknowledgment

We acknowledge the support from Research Acculturation Collaborative Effort (RACE) research grant by the Ministry of Higher Education, Malaysia (MOHE).

References

Bernstein, P. L. (2012). *The power of gold: the history of an obsession.* John Wiley & Sons, New York, 2nd edition.

Hogg, R. V., McKean, J., and Craig, A. T. (2013). *The power of gold: the history of an obsession.* Pearson, New Jersey, 7th edition.

Lyon, J. D. and Tsai, C. L. (1996). A comparison of tests for heteroscedasticity. *The Statistician*, **45**(3):337–349.

Sakia, R. M. (1992). The box-cox transformation technique: a review. *The Statistician*, **41**(2):169–178.

Sukri, M. K. A., Mohd Zain, N. H., and Zainal Abidin, N. S. (2015). The relationship between selected macroeconomic factors and gold price in malaysia. *International Journal of Business, Economics and Law*, **8**(1):88–96.

Worthington, A. C. and Pahlavani, M. (2007). The relationship between selected macroeconomic factors and gold price in malaysia. *Applied Financial Economics Letters*, **3**(4):88–96.

Yazici, B. and Yolacan, S. (2007). A comparison of various tests of normality. *Journal of Statistical Computation and Simulation*, **77**(2):175–183.

Zhang, Y. J. and Wei, Y. M. (2010). The crude oil market and the gold market: Evidence for cointegration, causality and price discovery. *Resources Policy*, **35**(3):168–177.