CHAPTER 3

METHODOLOGY

3.1 Overview

Seasonal Autoregressive Integrated Moving Average (SARIMA), Holt-Winters Exponential Smoothing (HWES), Autoregressive Integrated Moving Average (ARIMA) and Single Exponential Smoothing (SES) time series are approaches used by practitioners in many fields. SARIMA, ARIMA, SES and HWES used to forecast zakat collection in zakat institutions using past data. The accuracy of the models was compared using various error measurements such as of the Mean Square Error (MSE), Mean Absolute Error (MAE) and Mean Absolute Percentage Error (MAPE). The forecasting results improve as the MSE, MAE, and MAPE values are reduced.

3.2 Data

The data used in this research are secondary. This secondary data obtained from the official website of the zakat institutions (https://www.zakatselangor.com.my/informasi/laporan-aktiviti-zakat/) and (https://www.maiamp.gov.my/index.php/info/muat-turun/category/17-laporan-tahunan-maipk.html) for LZS and MAIPk and the data for PPZ, PZNS is from the zakat institutions. This study uses monthly data of zakat collection from January, 2010 to December, 2019 for LZS, PZNS and PPZ containing 120 observations for each zakat institutions and divided into three groups as follows:

i. The in-sample data comprising of 108 observations (January 2010 to December 2018).

ii. The model validating data for 12 observations (January 2019 to December 2019).

iii. The forecasted values of monthly data of January 2020 to December 2020.

Whereas yearly data of zakat collection from 1991 to 2019 is used for MAIPk containing 29 observations and divided into three groups as follows:

i. The in-sample data comprising of 24 observations (1991 to 2014)

ii. The model validating data for 5 observations (2015 to 2019)

iii. The forecasted values of monthly data of 2020 to 2031

3.3 Methods

Box and Jenkins (1976) popularized autoregressive integrated moving average (ARIMA) models for analyzing time series data with seasonal trends to develop a forecasting model. Seasonal Autoregressive Integrated Moving Average or better known as SARIMA method is time series forecasting method for stochastic model data with seasonal data pattern (Pongdatu and Putra, 2018). According to Li, Le, Xinli & Xiaoli (2017) and Siregar, Makmur & Saprin (2018) the SARIMA modeling consists of four steps: smoothing over time series (differential operation), model recognition, parameter estimation and model diagnosis, and model forecasting. Data transformation is frequently required during the smoothing process to make the time-series stationary. An ARIMA model was developed after the series was verified to be stationary. The model was tested for adequacy before being used for forecasting. This was accomplished by examining the residuals' autocorrelation function (ACF) and partial autocorrelation function (PACF). In addition, the Ljung-Box test was used to determine whether the model was correctly specified. A significant p-value less than 0.05 was used to acknowledge the presence of structure in the observed series that was not accounted for by the model. After determining the best model, monthly zakat collection values for the year 2019 were forecasted. To analyze and find the best-fitting model, Minitab version 19 was used.

It is not uncommon for data to show a trend pattern in the time series forecasting technique, where the pattern of data shows a tendency to increase or decrease. The Holt-Winters Exponential Smoothing method is a forecasting method that uses an exponential smoothing approach and is based on previous forecasting results. This method also includes additional parameters for dealing with seasonal data patterns. In the Holt-Winters Exponential Smoothing method, there are two main models: multiplicative and additive. The seasonal pattern was used to determine the model's determination (Pongdatu and Putra, 2018).

3.3.1 ARIMA and SARIMA Models

This method for selecting an appropriate ARIMA model for estimating and forecasting a time-series included identification, estimation, diagnostic checking and forecasting. ARIMA is a model that includes autoregressive (p) terms, differencing (d) terms, and moving average (q) operations (p, d, q). The statistical properties of a stationary time series, such as mean and variance, remain constant over time. The series is usually nonstationary because of the seasonality. Hence, Seasonal ARIMA

(SARIMA) is an extension of the ARIMA technique to a series in which a pattern repeats seasonally throughout time and it is expressed as SARIMA (p, d, q) (P, D, Q)s, which that the seasonal autoregressive (P), seasonal differencing (D) and seasonal moving average parameters (Q). For monthly data, s is the number of periods before the pattern repeats (Siregar et al., 2018).

The ARIMA model has gained popularity among researchers who believe it is superior in providing short-term forecasts (Pai and Lin, 2005). The ARIMA model projects future values based on a linear combination of historical error terms and values. ARIMA is made up of three parts:

i. AR, which stands for autoregressive.
ii. I, which stands for differencing.
iii. MA, which stands for moving average.

As shown in (3.1) (Razali et al., 2018), the ARIMA (p,d,q) model is written in intercept form:

$$Y_t = \phi_0 + \phi_1 Y_{t-1} + \phi_2 Y_{t-2} + \cdots + \phi_p Y_{t-p} + u_t - \Theta_1 u_{t-1} + \Theta_2 u_{t-2} + \cdots + \Theta_q u_{t-q} \tag{3.1}$$

where,

Y_t is an actual value, u_t are standard errors, Θ and ϕ are coefficients, p and q are integers. (Adebiyi et al., 2014).

When an ARMA model's time series must be differenced a certain number of times before becoming stationary, the model is called an ARIMA model. As previously stated, a time series that must be differenced d times to become stationary is referred to as an integrated of order d, denoted I (d). The graph (see Appendix A) where the MA and AR values are plotted with their respective significant values. We consider only 1 significant value from the AR model and likewise 1 significant value from the MA model. So, the ARMA model will be obtained from the combined values of the other two models will be of the order of ARMA (1,1). This further means that if the underlying AR and MA models are of the first-order and the time series is stationary at the first difference, the ARIMA model is denoted ARIMA (1, 1, 1). There are several methods for determining whether or not a time series is stationary. The autocorrelation function, also known as the ACF, is another way to check for stationarity. The autocorrelation function is the ratio of the covariance at a given lag to the autocorrelation function. Generally expressed as lag k, to the variance. At lag k, ρ_k denotes the ACF and is defined as follows; (Jansson and Larsson, 2020)

$$\rho_k = \frac{\gamma_k}{\gamma_0} \tag{3.2}$$

where γ_k is the covariance at lag k and γ_0 is the variance. A correlogram can be used to plot the ACF. The time series could be considered stationary if all or most of the lags are statistically insignificant, there is no specific pattern, constant variance, and the autocorrelations at various lags hover around zero in the correlogram. If the ACF correlogram resembles a white noise process, the time series is most likely stationary.

The Box-Jenkins methodology consists of four consecutive steps that should be followed when building an ARIMA model. The first step, identification, is used to determine appropriate values for p, d, and q. In the first step, the ACF and partial autocorrelation function (PACF) with their respective correlograms are used to detect patterns in p, d, and q. The PACF calculates the autocorrelation between observations in a time series separated by k lags, with the intermediate autocorrelation between lags kept constant. The second step is to estimate the parameters in the model. Step three is diagnostic checking, which evaluates the goodness of fit of the chosen ARIMA model, typically by determining whether the residuals are white noise. If the residuals are not white noise, steps one through three should be repeated with new values for p, d, and q. Yet, if the residuals are white noise, the model should be accepted and it is possible to proceed to step four. Forecasting is the fourth step, in which the model is used to forecast desired periods for the time series. (Jansson & Larsson, 2020) (Razali et al., 2018)

To test if there is joint autocorrelation for a certain number of lags, the Ljung-Box statistic has been used. The number of degrees of freedom in the Ljung-Box statistic is m, where m is the number of lags. The Ljung-Box statistic can also be used to determine whether a series is white noise for a given number of lags, so it can be used for the third step in the Box-Jenkins methodology, determining whether the residuals of the estimated ARIMA model are white noise. There is no evidence that residuals are not a white noise process if the Ljung-Box statistic is statistically insignificant. The Ljung-Box statistic is defined as

$$BL = n(n+2)\sum_{k=1}^{m}\left(\frac{\hat{p}_k^2}{n-k}\right) \tag{3.3}$$

where n denotes the size of the sample, m denotes number of lags and $\hat{p}_k$ is the autocorrelation at the kth lag. (Jansson & Larsson, 2020)

In this study, Box and Jenkins time series modelling techniques are used to model zakat collection data and forecast zakat collection over a short period. In the zakat collection time series data, a combination of seasonal and non-seasonal Autoregressive Integrated Moving Average ARIMA (p,d,q)x(P,D,Q) model is used.

The non-seasonal component of the Box and Jenkins ARIMA model is represented by three parts: Autoregressive (AR) of order p, Moving Average (MA) of order q, and the degree of Integration of order d, ARIMA (p, d, q).

Assume we have N observations for a univariate time series at t, Y_1, Y_2, …, Y_t. Then, the Box-Jenkins ARIMA model for time series data is given by (Yemataw, 2014):

$$\phi(\beta)\Delta^d Y_t = \mu + \theta(\beta)\epsilon_t \tag{3.4}$$

where, $\phi(\beta) = 1 - \Phi_1\beta^1 - \Phi_2\beta^2 - \Phi_3\beta^3 - \cdots \Phi_q\beta^p$ and $\theta(\beta) = 1 - \theta_1\beta^1 - \theta_2\beta^2 - \theta_3\beta^3 - \ldots - \theta_q\beta^p$

β is the backward shift operator and $\Delta=1-\beta$, $\Phi(\beta)$ and $\theta(\beta)$ is the non-seasonal AR and MA operators, d is the orders of the integration/differencing and ϵ_t is Gaussian white noise while μ is a constant.

The Seasonal-ARIMA model can be represented in a multiplicative form as ARIMA (p, d, q)(P, D, Q)s, where (p, d, q) represents the non-seasonal component of the model, (P, D, Q)s represents the seasonal component of the model, and s denotes the number of periods per season. P stands for Seasonal Autoregressive (SAR) term, D for n number of seasonal difference (s) performed, and Q for Seasonal Moving Average (SMA) term in the seasonal component (Pongdatu and Putra, 2018). A fit from the Seasonal-ARIMA model can be written in the following general notational form:

$$(1-\beta)^d(1-\beta)^D Y_t = \mu + \frac{\theta(\beta)\theta(\beta^s)}{\phi(\beta)\Phi(\beta^s)}\epsilon_t \tag{3.5}$$

where, β is a backward shift operator and Φ and Θ are the Seasonal Moving Average (SMA) and the Seasonal Autoregressive (SAR) polynomials of order P and Q respectively.

If the data is not stationary, modifications are required to produce stationary data. One of the common ways used is the differencing method. To determine whether the series is stationary or non-stationary, viewing the plot of the series or its difference form can be used. The process of differencing can be done for some period until the data become stationary by subtracting data with previous data. According to (Tinungki, 2019) a handy notation in the differencing method is the backward shift B operator, as follows:

$$BX_t = X_{t-1} \tag{3.6}$$

Notation B has the effect of shifting data 1 period backward. And then will change the data 2 periods back, as follows:

$$BBX_t = B^2 X_t = X_{t-2} \tag{3.7}$$

If a time series is not stationary, then the data can be drawn closer to stationary by making the first differencing.

$$X'_t = X_t - X_{t-1}$$

$$X'_t = X_t - BX_t = (1-B)_t X_t \tag{3.8}$$

The first differencing is denoted by (1-B). The purpose of the differencing calculation is to achieve the stationarity, and in general, if there is a difference of the-d or do to reach stationarity, it is written as follows:

$$(1-B)^d X_t \tag{3.9}$$

Plot the series in Minitab, to find out whether the series is stationary or not. It will be checked by viewing the plot, ACF graph, and PACF graph. To determine the mean value of zakat collection the following equation is used:

$$\bar{x}_t = \frac{\sum_{t=1}^{N} x_t}{N} \tag{3.10}$$

Based on Minitab 19, the time series plot in figure 3.1 shows that the series is not constant. After viewing the series data distribution plot, the ACF illustrates that the series is not stationary because many time lags go off from the significant limits, and the information slowly decays. The ACF values that are not significantly close to zero confirmed the ACP plot. PACF diagram describes that the data is interrupted after lag 6. So, the information is not stationary mean.

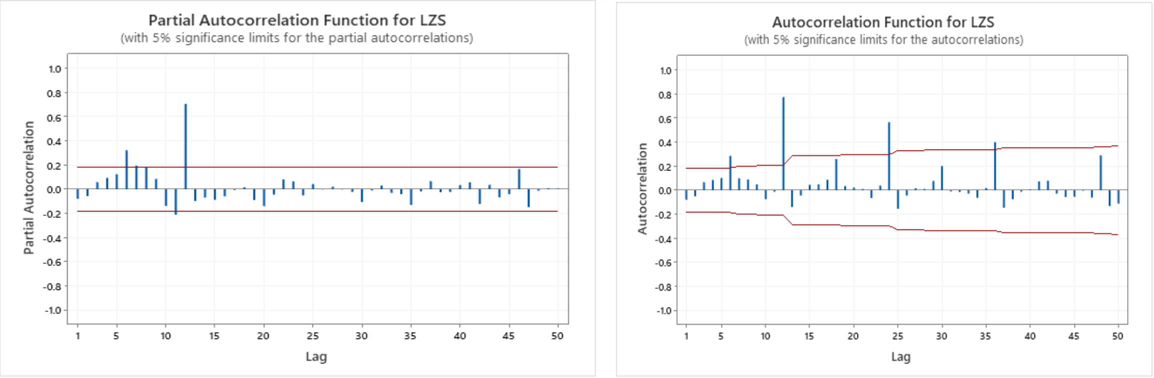

Figure 3.1: Plot time series ACF and PACF function.

There are two types of stationaries: stationary invariance and stationary in the mean. If it is not stationary to the variant, the Box-Cox transformation is carried out, whereas if it is not stationary to the mean, the differencing is carried out.

i. Testing stationery invariance, a time-series data, which is stationary from time to time, have a constant and unchanged data fluctuation. If the data is not stationary invariance, the transformation box-cox can be used on the mini-tab. The result is as follows:

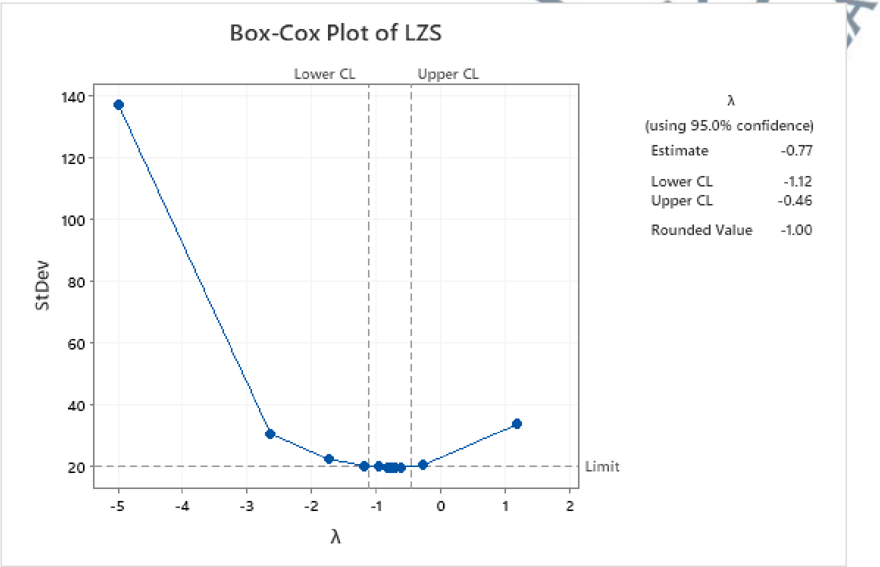

Figure 3.1: Box cox of real data.

Based on the results of box-cox testing shown in figure 3.1, it can be seen that the is not stationary to the variance yet since the rounded Value ≠ 1. Thus, it needs to be transformed according to the Rounded Value stationer.

ii. The stationary test in the mean is the fluctuation of the data around a constant mean value, a way of doing stationary testing when the data is not stationary in the mean is using the differencing method to produce stationary data.

Differencing is carried out by reducing the value in a period with the amount in the previous period.

3.3.2 Single Exponential Smoothing Method

Exponential smoothing is a powerful moving average forecasting method that is still simple to use and is commonly used for short-term forecasting. This method makes use of a few previous data entries (Khairina et al., 2019). It is simple to implement and compute because it does not require keeping track of previous input data. It fades the effect of unusual data uniformly (Karmaker, 2017). The equation of SES is as follow:

$$M_t = M_{t-1} + \alpha(M_{t-1} - N_{t-1}) \tag{3.11}$$

where,

M_t is the forecast value for period

M_{t-1} is the forecast value for the previous period $t - 1$

N_{t-1} is the actual zakat collection value for the previous period $t - 1$

α is smoothing constant $(0 \leq \alpha \leq 1)$

The value of the exponential smoothing constant determines the accuracy of this technique. Constant values should be chosen so that forecasts are more accurate. Forecasting error, which is the difference between actual and forecasted zakat collection, is used to assess forecasting accuracy. The goal is to find the constant that reduces forecast error to the smallest amount possible.

3.3.3 Holt-Winters Exponential Smoothing Method

Winters (1960) extended Holt 's method to capture direct seasonality. The method of the Holt-Winters is based upon three smoothing equations one for level L_t, one for trend b_t and one for the seasonal component represented by S_t. This is similar to Holt 's method, with one more equation to handle with some seasonality. There are two different types of Holt-Winters, depending on whether the seasonality is represented in an additive or Multiplicative approach (Makridakis & Wheelwright, 1997). When seasonal variations change proportionally to the level of the series, the Multiplicative Holt-Winters (MHW) method is preferred and when seasonal variations are roughly constant throughout the series, the Additive Holt-Winters (AHW) method is preferred (Suppalakpanya et al., 2019a)

3.3.3.1 Additive Holt-Winters (AHW) Method

The Holt Winters methods include estimates of the seasonal factors for periods (denoted by S). The parameters p, states the number of seasonal periods in a year. For example, p = 12 would correspond to monthly seasonal adjustments and p = 4 would correspond to quarterly seasonal adjustments. The following are the basic formulas for additive model (Lidiema, 2017),

$$L_t = \alpha(Y_t - S_{t-s}) + (1-\alpha)(L_{t-1} + b_{t-1}) \tag{3.12}$$

$$b_t = \beta(L_t - L_{t-1}) + (1-\beta)b_{t-1} \tag{3.13}$$

$$S_t = \gamma(Y_t - L_{t-1} - b_{t-1}) + (1-\gamma)S_{t-s} \tag{3.14}$$

$$F_{t+m} = L_t + b_t m + S_{t-s+m} \tag{3.15}$$

Where s is the length of seasonality in months, L_t denotes the level of the series period ahead and the parameters (α, β and γ) generally constrained to lie b_t represents growth, S_t is the seasonal element, F_{t+m} is the forecast for m between 0 and 1.

3.3.3.2 Multiplicative Holt-Winters (MHW) Method

$$L_t = \alpha \frac{Y_t}{S_{t-s}} + (1-\alpha)(L_{t-1} + b_{t-1}) \tag{3.16}$$

$$b_t = \beta(L_t - L_{t-1}) + (1-\beta)b_{t-1} \tag{3.17}$$

$$S_t = \gamma \frac{Y_t}{L_t} + (1-\gamma)S_{t-s} \tag{3.18}$$

$$F_{t+m} = (L_t + b_t m)S_{t-s+m} \tag{3.19}$$

Y_t is the observed value at time *t*. L_t is the stable part at time *t*. S_t is the seasonal part at time *t*. b_t is the trend part at time *t*. F_{t+m} is the forecast value at time m. *m* is the predicted the number of periods. s is the season length. α , β, γ $\in [0,1]$ is the smoothing parameter, β, γ takes the principle that the root mean square error (RMSE) between the predicted and measured values is minimal.

3.3.4 Error Analysis

Forecast accuracy of forecast error measures can be critical when dealing with practical problems (Yokuma and Armstrong, 1995). Forecast error measures are commonly used to assess the quality of forecasting methods. Error measures can also detect the best forecasting mechanism in the case of multiple objects (Shcherbakov et al., 2013). As a result, it is critical to find the best smoothing parameter values as has been illustrated later in this study. When comparing Seasonal ARIMA with HWES approaches, forecasting was conducted over 12 months. The performance of each method can be calculated by using mean absolute error (MAE), mean absolute percentage error (MAPE), mean square error (MSE), and root mean square error (RMSE) to define the most efficient model for forecasting the collection of zakat. The main error measures that have been used in this study are listed below following Ahmad and Nor (2020).

$$\text{MAE: } \frac{\sum_{j=1}^{m}\left|y_j-\hat{y}_j\right|}{m} \tag{3.20}$$

$$\text{MAPE: } \frac{\frac{\sum_{j=1}^{m}\left|y_j-\hat{y}_j\right|}{y_j}}{m} \times 100\% \tag{3.21}$$

$$\text{MSE: } \frac{\sum_{j=1}^{m}\left|y_j-\hat{y}_j\right|^2}{m} \tag{3.22}$$

$$\text{RMSE: } \sqrt{\frac{\sum_{j=1}^{m}\left|y_j-\hat{y}_j\right|^2}{m}} \tag{3.23}$$

where,

$e_j = y_j - \hat{y}_j$

Here y_j is the actual load, while $\hat{y}_j$ is the predicted load. MAE is a measure of all the indication data based on the assumption that they are all equally weighted (Yaffee and McGee, 2012). If the forecasting method is good, the comparison error with previous data is close to zero. MAPE is a relative measure that is equivalent to MAE. It refers to the accuracy measurement used in quantitative method forecasting. According to Ahmad and Nor (2020), if the MAPE is less than 10%, it is considered accurate forecasting, 20%-30% good forecasting, 20%-50% acceptable forecasting, and more than 50% as inaccurate forecasting. MSE is a method of determining overall accuracy by providing a degree indication, but with more weight. If the weight is large and especially undesirable, RMSE squared root to give the smaller value.