CHAPTER 3

RESEARCH METHODOLOGY

3.1 Assumptions of ARIMA Model

In developing the time series model, several assumptions have been made before the Box-Jenkins ARIMA model can be employed. These assumptions are as follows.

1. The Box-Jenkins model assumes that the time series is stationary. A stationary process has the property that the mean, variance and autocorrelation structure do not change over time. This is a form of weak stationarity which necessitates a time series without trend, constant variance over time, a constant autocorrelation structure and no periodic fluctuations (seasonality). If the series is non-stationary, it is recommended that the series be differenced once or twice to achieve stationarity.
2. The model assumes that there exists a linear relationship between the forecast (dependent) variable and the explanatory (independent) variables. If this assumed form is incorrect, then the forecasts may be inaccurate and often a more appropriate model form (e.g., a non-linear model) can be found and the model needs to be re-estimated.

3. It is assumed that all data points behave consistently and thus there are no extraordinary outliers. The presence of outliers may affect the model strongly and misleadingly.
4. If shocks are present in the time series, they are assumed to be randomly distributed with a mean of 0 and a constant variance.
5. It is also assumed that the residuals in a time series model is randomly distributed, exhibit a normal distribution, have non-significant autocorrelations and partial autocorrelations, and have a mean of 0 and homogeneity of variance over time.

3.2 AR and MA Models

In general, the pth order autoregressive (AR) model is defined as follows

$$y_t = \delta + e_t + \Phi_1 y_{t-1} + \Phi_2 y_{t-2} + \ldots + \Phi_p y_{t-p}$$

where y_t is the time series, e_t is white noise, $\Phi_1, \ldots, \Phi_p$ are the parameters of the model and

$$\delta = (1 - \sum_{i=1}^{p} \Phi_i)\, \mu$$

with μ denoting the process mean.

An autoregressive model is simply a linear regression of the current value of the series against one or more prior values of the series. The value of ***p*** is called the order of the AR model. The parameters of the AR model can be estimated using

standard linear least squares or maximum likelihood method. They also have a straightforward interpretation.

Another common approach for modeling univariate time series models is using the moving average (MA) model. The qth order MA model relates observation y_t to weighted random disturbances (shocks) going back q periods is given as follows

$$y_t = \mu + e_t + \theta_1 e_{t-1} + \theta_2 e_{t-2} + \ldots + \theta_q e_{t-q}$$

where y_t is the time series, μ is the mean of the series, e_{t-i} are white noise, and $\theta_1, \ldots, \theta_q$ are the parameters of the model. The value of q is called the order of the MA model.

In short, a moving average model is conceptually a linear regression of the current value of the series against the white noise or random shocks of one or more prior values of the series. The random shocks at each point are assumed to come from the same distribution, typically a normal distribution, with mean at zero and constant variance. Fitting the MA estimates is more complicated than the AR models because the error terms are not observable. This means that iterative non-linear fitting procedures need to be used in place of the ordinary linear least squares. MA models also have a less obvious interpretation than AR models. Sometimes the ACF and PACF will suggest that an MA model would be

a better model choice and sometimes both AR and MA terms should be used in conjunction.

3.3 ARMA Models

The Box-Jenkins ARMA model is a combination of the AR and MA models as follows

$$y_t = \delta + e_t + \Phi_1 y_{t-1} + \Phi_2 y_{t-2} + \ldots + \Phi_p y_{t-p} + \theta_1 e_{t-1} + \theta_2 e_{t-2} + \ldots + \theta_q e_{t-q}$$

where the terms in the equation have the same meaning as given for the AR and MA model.

The Box-Jenkins model assumes that the time series is stationary. Box and Jenkins recommend differencing non-stationary series one or more times to achieve stationarity. If the effect of differencing is integrated in the model, the ARMA model becomes more general and thus produces an ARIMA model, with the "I" standing for "Integrated". This model can be further extended to include seasonal autoregressive and seasonal moving average terms. Understandably this inclusion complicates the notation and mathematics of the model. However, the underlying concepts are still the same. The most general Box-Jenkins model includes difference operators, autoregressive terms, moving average terms, seasonal difference operators, seasonal autoregressive terms, and seasonal

moving average terms. Nevertheless, only necessary terms should be included in the model.

3.4 Statistical tests

A number of statistical tests are employed to test the validity of the model. Among them are the adjusted R square test, the Akaike Information Criterion and the residual autocorrelation test.

3.4.1 Adjusted R^2

After estimating the coefficients of a model, we can measure the errors of fit by subtracting the estimated value ($\hat{Y}$) from the actual value (Y) for example $e_i = (Y_i - \hat{Y}_i)$ for the *i*th observation. Another measure of the goodness of fit for a model is how good it can explain the variation in Y. For this purpose, the variation of Y from its mean $(Y_i - \overline{Y})$ is partitioned into two parts:

$$(Y_i - \overline{Y}) = (Y_i - \hat{Y}_i) + (\hat{Y}_i - \overline{Y})$$

Squaring both sides of the above equation and summing up over all data

$$\sum (Y_i - \overline{Y})^2 = \sum (Y_i - \hat{Y}_i)^2 + \sum (\hat{Y}_i - \overline{Y})^2 + 2\sum (Y_i - \hat{Y}_i)(\hat{Y}_i - \overline{Y})$$

The last term on the right of the equation is zero (Makridakis 1998) and thus we have

$$\sum(Y_i - \bar{Y})^2 = \sum(Y_i - \hat{Y}_i)^2 + \sum(\hat{Y}_i - \bar{Y})^2$$

Or in simplified notation

$$SST = SSE + SSR$$

where SST is the total sum of squares, SSE is the sum of squared errors and SSR is the sum of squares from regression. SSE is regarded as the unexplained variation in Y while SSR is the part of the deviation that is captured (explained) by the model.

R^2 is defined as the ratio of SSR over SST.

$$R^2 = \frac{\sum(\hat{Y}_i - \bar{Y})^2}{\sum(Y_i - \bar{Y})^2} = \frac{SSR}{SST}$$

where $\hat{Y}$ is the estimated value of Y derived from the model and $\bar{Y}$ is the mean of the samples.

It is observed that the value of R^2 increases with the number of variables used in the model. Therefore, maximizing the value of R^2 would not be a good approach

to finding the best model since it entails selecting the model with the maximum number of variables. This selection criterion goes against the principle of parsimony which encourages the use of the fewest variables possible. The adjusted R^2 corrects this anomaly by taking into account the number of variables in a model.

$$\text{Adjusted } R^2 = 1-(1-R^2)\frac{N-1}{N-k-1}$$

where N is the number of observations and k is the number of explanatory variables in the model. The higher the adjusted value of R^2 is (closer to 1) the better the model.

3.4.2 The Akaike Information Criterion (AIC)

The AIC provides a measure of the goodness of fit of a model which takes into account the number of terms in the model. The AIC is given as (Makridakis et. al. 1998)

$$AIC_{p,q} = -2 \log L + 2m$$

where L denotes the likelihood function and m = p + q + 1.

The log likelihood function is the maximized (Gaussian) log likelihood and m represents number of independently adjusted parameter in the candidate model.

The AIC is commonly used with ARIMA models and its value is minimized to find the appropriate model order (i.e. the values of p and q). Since not all computer programs produce the AIC or the likelihood L, it is not always possible to find the AIC for a given model. A useful approximation to the AIC is obtained via the approximation

$$-2 \log L \approx n (1+ \log (2\pi)) + n \log \sigma^2$$

where σ^2 is the variance of the residuals and n is the number of observations in the series. All computer programs will produce the value of σ^2 so the AIC can be found approximately using the formula

$$AIC \approx n (1 + \log (2\pi)) + n \log \sigma^2 + 2m$$

Sometimes the first term above is omitted because it is the same value for all models. It should be noted that AIC is not very meaningful by itself. It is only useful in comparison to the AIC value for another model fitted to the same data set (Makridakis et. al. 1998).

3.4.3 Residual Autocorrelation Statistics

The error terms after the model is fit (the residuals) should be independent and follow the last assumption of the model. Therefore, the autocorrelation function of the residuals should be close to 0 for displacement (lag) greater than or equal to one. But since we have finite samples, each of the sample autocorrelation will not be exactly zero. It has been shown that the autocorrelation coefficients of white

noise data have a normal sampling distribution with mean zero and standard deviation $1/\sqrt{N}$ where N is the number of observations in the series. Therefore, it can be inferred that 95% of all sample autocorrelation coefficients must lie within the range specified by the mean plus or minus 1.96 standard deviation.

3.5 Kalman filter

It is not necessary to know the details of the mathematical formulation of the Kalman filter to be able to apply it effectively. However, literature on Kalman filter and its applications is abundant since it is widely used in many fields. Maybeck (1979) provides a friendly introduction to the general idea of the Kalman filter. Brockwell and Davis (1996) show that any ARIMA model can be written in state space form which can be executed using Kalman filter algorithm.

Basically, the Kalman filter addresses the general problem of trying to estimate the state x of a discrete-time controlled process that is governed by the linear stochastic difference equation (Welch and Bishop 2001)

$$x_k = Ax_{k-1} + Bu_k + w_{k-1},$$

where x_k is the (nx1) state vector of the process at time k; A is the (nxn) state transition matrix of the process from state k-1 to k and w_k is the associated process noise which is assumed to be white. Matrix B relates the optional control input u to the state x.

Observations on this variable are made through measurement z that is modeled in the form

$$z_k = Hx_k + v_k .$$

where z_k is the actual measurement of x at time k; H is the matrix that relates the state to the measurement and v_k is the associated measurement error. The random variables w_k and v_k are assumed to be independent (of each other), white, and with normal probability distributions

$$p\,(w) \sim N\,(0\,,\,Q)\,,$$

$$p\,(v) \sim N\,(0\,,\,R).$$

The process noise covariance Q and measurement noise covariance R matrices are assumed to be constant although in practice they might change with each time step or measurement. Similarly, A and H might also change with each time step, but here we assume they are constant.

In order to implement Kalman filtering, matrices *A, B and H* are assumed to be known. Also the statistics of the noises are assumed known. What a Kalman filter does is make an optimal estimate of x by recursively updating two sets of equations namely the time update equations and the measurement update equations. The time update equations are

$$\hat{x}^{-}_{k} = A\hat{x}_{k-1} + Bu_k$$

$$\bar{P}_k = AP_{k-1}A^T + Bu_k + Q$$

The measurement update equations are

$$K_k = \bar{P}_k H^T (H\bar{P}_k H^T + R)^{-1}$$

$$\hat{x}_k = \bar{\hat{x}}_k + K_k (z_k - H\bar{\hat{x}}_k)$$

$$P_k = (I - K_k H)\bar{P}_k$$

Since Kalman filter is an optimal predictor, it is hypothesized that its performance would be better than those of the ARIMA models. So, based on the above literature on Kalman Filter, the hypothesis of this study is as follows:

H1: Kalman Filtering Method offers a more superior best least square estimate as compared to ARIMA models.