CHAPTER 2
LITERATURE REVIEW AND THEORETICAL BACKGROUND

2.1. Literature Review

2.1.1. Previous Researches

In retail, Trauzettel (2014) conducted a research about how to reduce overstock problems in a retail store using simulations by changing the days of ordering within the week. The data used was daily number of demands from the retail store. In the paper, the quantity of ordered items was unchanged. It was mentioned that to reduce overstock problems, on top of moving the delivery date to the day when demand was not too high, increasing the number of stored units per items was also necessary.

Kalaoglu et al. (2015) performed a research on retail demand forecasting in clothing industry to produce new styles and designs by looking at the future demand for certain products. Forecast methods used were simple moving average, weighted moving average, and linear trend forecasting methods. The forecasts were applied on seven styles of products which were shirts, jeans, pants, women blouses, women jeans, women pants, and skirts. There were three year of monthly data and the last two months of the data were forecasted. Correlation values for forecasted values and actual data of the last two months were calculated to see the accuracy of the forecast. Forecast error parameter of least squares method was also used.

When designing a forecasting model, family level (shirts, T-shirts, trousers, etc.) forecast and SKU level forecast can be considered (Choi et al., 2014). Family level forecast is used when there is historical data and SKU level forecast is used when there is no historical data available. As historical data are not available in SKU level, the items changed quickly, and the SKU changed too even if similar items are sold later.

Outside of fashion retail, there are many researches that use time series forecasting. Nafitri (2010) wrote a thesis about reducing production cost in a manufacturing company by determining the quantity of safety stocks based on the result from forecasting customer’s demand. The method used were naïve method, simple exponential smoothing method, Holt-Winters exponential smoothing method, and a combination of Holt-Winters and naïve methods. The combination method was calculated by averaging forecast results from Holt-Winters method and naïve method. In the thesis, it was mentioned that the combination method was the best method for forecasting customer's demand because it produced the smallest mean absolute percentage error (MAPE). The forecast error parameters used were mean absolute deviation (MAD), mean squared error (MSE), and mean absolute percentage error (MAPE). The safety stock was calculated from mean absolute deviation (MAD). Halder et al. (2015) did a research about demand forecasting by comparing several forecasting methods including naïve method, moving average method, weighted moving average method, and exponential smoothing method. The forecast error used were MAD, MSE, and MAPE. Verma et al. (2017) used time series forecasting to predict the quantity of active students in a university to decide the quantity of raw materials for the university's food cafeteria. The research used ARIMA model to forecast non-seasonal data with a trend and irregular components. The forecast errors calculated were ME (mean error), RMSE (root mean squared error), MAE (mean absolute error), MPE (mean percentage error), MAPE (mean absolute percentage error), MASE (mean absolute sealed error), and ACFI (autocorrelation of errors at lag 1).

Not all data for forecasting is simple. With time, the size of data became larger and more complex. Different forecasting models and methods were upgraded to be able to do forecasting to a complex data or condition. Bontempi et al. (2013) conducted a research about machine learning strategies for time series forecasting. Rodrigues and Figueiredo (2013) used ANN (artificial neural network) to forecast demand. Machine learning and ANN can handle complex data well, but these methods need a lot of data available to train and test the machine or network.

Nenni et al. (2013) mentioned that in fashion industry, demand forecasting had an important role acting as an input for planning activities. When poor forecast results are used many problems can happen which would affect the inventory. It is said that traditional methods, like exponential smoothing, do not work well with intermittent (when an item has several periods of zero demand), lumpy (when items are too slow-moving or too expensive), and erratic demand (when an item almost has no low demand). When traditional methods do not work well, more modern methods like ANN are needed.

2.1.2. Current Research

Current research is about demand forecasting in X clothing store to manage inventory. This research was done to help X clothing retail store reduce their overstocked and understocked items. This research used historical data on the number of items sold monthly for almost every clothing product in X clothing store. The store ordered their products every week, but the data used were monthly. The results of forecasting could then be divided into four for weekly order.

Item classification in the X clothing store is based on the original SKU that exists in the clothing store (SKU and family level of classification) based on Choi et al. (2014). They were grouped based on their first three letters which indicate the type of items sold (blouses, long pants, short pants, etc.). Both family level and SKU level were used because every item in every SKU had similarities and it could be grouped to forecast the demand quantity of those certain types of products.

Before forecasting, the patterns of the data were analyzed to see if the data are sufficient to forecast and if the data have any of time series components. The methods used for forecasting were simple average methods, moving average methods, and exponential smoothing methods, which were similar to Kalaoglu et al. (2015), Nafitri (2010), and Halder et al. (2015). These methods are suitable for X clothing store as it is a small company and it does not have the personnel required to forecast using complicated methods.

2.2. Theoretical Background

2.2.1 Inventory Management

According to American Production and Inventory Society (APICS), inventory management is concerned with inventory planning and inventory control. The goal of inventory management is to have low inventory management cost but still can give maximum customer satisfaction. To manage inventory, it must be based on the type of products, the customers, and the methods to have the product manufactured or purchased (Toomey, 2000). The level of inventory will be affected by the products characteristics, the quantity of the product needed based on customers demand, and the costs to manage the inventory (Wild, 2002).

2.2.2. Inventory Management in Fashion Retail

Fashion retails are fast-moving because the fashion trend can change easily and customers do not want to buy outdated fashion products (Berman and Evans, 2013). In fashion retails, the concerns in inventory management are directed to fashion style and design, brands, fashion trend at the time, season or the weather at the time, and also promotions (Ray, 2010).

For example: apparel products for adults and for children are different (in design and size); style and design for men and women are also different. There are also different styles and designs for different occasions and seasons, that is why there are many types of apparels from informal wear, formal wear, sportswear, traditional apparel, winter wear, including shoes, flip-flop, and others.

To manage inventory in fashion retails, those things mentioned above are needed to be considered, since customers might want certain brands, styles, fabric materials, colors that are in trend or suitable for the seasons. Apparel products could last a long time in storage room assuming they are kept properly in a good environment, but it would cost a fashion retailer extra money to keep them for a long time. It takes up space in storage room for new products to be stored and it does not make profits for the retailer (Berman and Evans, 2013). That is why when determining what to order and its quantity to suppliers, a fashion retail store management needs think about how long the product styles or designs would last, retailers also need to think about the following: if a certain product is what the customer want or need at the moment; how fast can the products sell until it is considered outdated or until the next season comes; is the cost reasonable enough so the selling price could be reasonable too; and other things related to consumer behaviors and fashion conditions in the country or even globally.

2.2.3. Forecasting in General

Based on the data used, there are three types of forecasting techniques that can be used (Bodily, 2008). The first one is qualitative technique. This technique does not rely on historical data, it can be used even when historical data are not available. The reason for unavailability of data is usually because there is really no recorded data or the cost for gathering data is too expensive and is not worth doing. So, qualitative technique uses any relevant and similar information to the problem and also it uses the experts' opinion.

The second is causal procedure techniques. The goal of this technique is to forecast by analyzing the independent and dependent variables. In research, Independent variables are the things that stand alone but affect other things (dependent variables) and dependent variables are the things that are affected by independent variables (Kaur, 2013). This method uses independent variables (price, method of advertising, frequency of advertising, etc.) to forecast dependent variables (sales level, demand level, and other things).

The last technique is time series method. Time series method uses historical data that are collected over time. Time series forecasting consists of identifying patterns in data and project the patterns into the future by considering the fluctuations that had happened in past data (Hanke and Wichern, 2009).

2.2.4. Time Series Data

Hanke and Wichern (2009) defined time series data as data collected annually every period with constant time gap (every hour, every day, every week, every month, etc.). In time series data, there are several components that might be present:

a. Trend

   Time series data consist of trend components when the data show an increase or decrease over several period of time. A linear line needs to be drawn to see the increase or decrease in trend.

b. Seasonal

   Data with seasonal component have a pattern that repeats periodically (every Friday, every July, every four months, etc.). The length of the time period before the pattern repeats again is called seasonal length. It is a pattern that is always present every year. Seasonal patterns can be caused by holidays, season change, and other things that happen every year like the start and end of school term.

c. Cyclical

   Cyclical component is similar to seasonal component, but the pattern lasts for more than a year and the pattern does not repeat at constant time intervals.

d. Irregular

   Irregular component is a random fluctuation in time series data. The data will not follow a certain pattern.

Aside from having one or more of those components, time series data can be stationary (have constant level or mean) and non-stationary. Having a constant level or mean means that the data does not have trend component. Depending on its stationarity and the components a data has, the best forecasting method to be used will be different.

2.2.5. Time Series Analysis

To know what components or patterns the data have, a time series plot needs to be made (Hanke and Wichern, 2009). Time series plot is a line chart with time indication on the x-axis. The forecaster can subjectively decide what pattern or components that exist in the data by observing the line chart. Aside from the patterns and components, outliers can also be seen from the time series plot.

Outliers are values that are abnormally high or low compared to the other values (Hanke and Wichern, 2009). It is usually caused by data input error or some marketing strategy that attracts more customers, so more items were sold. Outliers can be identified by visualization from the time series plot or by calculating the upper and lower bounds of the data. When outliers occur in data, those values need to be removed before being forecasted (Ray, 2010)

However, in time series data it is not recommended to remove outliers. If the reason for outliers are not error when inputting data, forecaster can study it in accordance with what had happened at the time. For example, it can be used to predict the increase in demand when there is a certain type of promotion or to know what kind of condition that resulted in abnormally low sales so that it would not happen again. It is also not recommended because removing those outliers would mess the component or pattern in time series data. That is why, there is a method of winsorization (Reifman and Keyton, 2010). To winsorize, is to replace the outliers with the largest or smallest value that are not outliers. Abnormally high outliers would be replaced by the largest value that are not outliers and abnormally low outliers would be replaced by the smallest value that are not outliers. It can be done manually by the forecaster.

2.2.6. Time Series Forecasting

There are three basic forecasting methods namely naïve method, averaging methods, and exponential smoothing methods (Hanke and Wichern, 2009).

a. Naïve Method

Naïve method uses the most recent observation to forecast the future. The simplest forecasting in naïve method is taking the value in this period to forecast the value in the next period. The equation is

$$\hat{Y}_{t+1} = Y_t \tag{2.1}$$

where $\hat{Y}_{t+1}$ is the forecast for the next period and Y_t is the actual data at time t.

If a set of data have a trend, naïve method considers the difference (Equation 2.2) or the change rate (Equation 2.3) between this period and last period.

$$\hat{Y}_{t+1} = Y_t + (Y_t - Y_{t-1}) \tag{2.2}$$

$$\hat{Y}_{t+1} = Y_t \frac{Y_t}{Y_{t-1}} \tag{2.3}$$

If a set of data have seasonal components, the simplest way is to take the value from this seasonal period to forecast the next period. For example, if the data have four months seasonal length (quarterly data), the equation below is used.

$$\hat{Y}_{t+1} = Y_{t-3} \tag{2.4}$$

When considering seasonal and trend components, naïve method would add the value from this seasonal period with the average value of difference from current seasonal period. Here is the equation to forecast quarterly data.

$$\hat{Y}_{t+1} = Y_{t-3} + \frac{(Y_t - Y_{t-1}) + \cdots + (Y_{t-3} - Y_{t-4})}{4} \tag{2.5}$$

b. Averaging Methods

i. Simple Average

In simple average method, value of next period is calculated by averaging the values from the beginning until the most recent data. Simple average method is for non-seasonal data. The minimum data requirement is thirty data points.

$$\hat{Y}_{t+1} = \frac{1}{t}\sum_{i=1}^{t} Y_i \tag{2.6}$$

When only the most recent observation is used to forecast the next second period, below equation is used.

$$\hat{Y}_{t+2} = \frac{t\hat{Y}_{t+1} + Y_{t+1}}{t+1} \tag{2.7}$$

ii. Moving Average

Moving average method cares only to recent data which are used for averaging and the number of periods used can be changed depending on the forecaster. Smaller number of periods is better for adapting to sudden changes in the data. Larger number of periods is better when changes are moving slowly. Moving average method cannot handle trend and seasonality very well even though it is better than simple average method. A moving average with k number of period computed, MA(k), has the following equation.

$$\hat{Y}_{t+1} = \frac{Y_t + Y_{t-1} + \cdots + Y_{t-k+1}}{k} \tag{2.8}$$

Moving average also has a smoothing effect. The larger number of periods to be averaged, the larger the smoothing effect and can eliminates or smoothens seasonal components.

c. Exponential Smoothing Methods

i. Simple Exponential Smoothing (SES)

SES is appropriate for data with unpredictable increasing or decreasing trend (no trend and no seasonal component). SES has a weighting system on recent and past data, symbolized by α. The value of α is 0 < α < 1 and can be adjusted to reach minimum forecast error. The larger the value of α, the larger the weight on recent values or data, and vice versa. The larger the weight means the calculation would put more importance to either past or recent data depending on how large the value of α is. The formula of SES is

$$\hat{Y}_{t+1} = \alpha Y_t + (1-\alpha)\hat{Y}_t \tag{2.9}$$

where $\hat{Y}_t$ is the value of the old forecast.

At the beginning of calculating SES, $\hat{Y}_t$ can be the same value as Y_t or it can be calculated by averaging the first several values (usually the first five or six values).

ii. Double Exponential Smoothing

Double exponential smoothing is also known as Holt's Method (Holt's Linear Exponential Smoothing), which is exponential smoothing adjusted to consider trend component of data. Double exponential smoothing is best for data with linear trend. Double exponential smoothing estimates the smoothed data (level as intercept) and trend (as slope) of a data and uses them to forecast the next value.

To calculate smoothed current level estimation, the formula is

$$L_t = \alpha Y_t + (1-\alpha)(L_{t-1} + T_{t-1}) \tag{2.10}$$

To calculate smoothed current trend estimation, the formula is

$$T_t = \beta(L_t - L_{t-1}) + (1-\beta)T_{t-1} \tag{2.11}$$

where α is the smoothing constant for level (0 < α < 1) and β is the smoothing constant for trend (0 < β < 1). The initial value of level (L_{t-1}) is the same as the actual value of data (Y_t) and the initial value of trend (T_{t-1}) is zero.

To forecast the value for p periods ahead is

$$\hat{Y}_{t+p} = L_t + pT_t \tag{2.12}$$

where and p is the number of periods to be forecasted into the future.

The larger the smoothing constant, the more the value follow existing rapid changes, and vice versa. Also, the smaller the smoothing constant is, the smoother the pattern values are.

iii. Triple Exponential Smoothing (ETS)

ETS is also known as Holt-Winter's Method, which is exponential smoothing method adjusted to consider trend and seasonal components in a data. Before calculating value for the next period, ETS estimates the data's smoothed level, trend, and seasonality first. To estimate the smoothed level, the formula is

$$L_t = \alpha \frac{Y_t}{S_{t-s}} + (1-\alpha)(L_{t-1} + T_{t-1}) \tag{2.13}$$

To estimate the smoothed trend, the formula is

$$T_t = \beta(L_t - L_{t-1}) + (1-\beta)T_{t-1} \tag{2.14}$$

To estimate the smoothed seasonality, the formula is

$$S = \gamma \frac{Y_t}{L_t} + (1-\gamma)S_{t-s} \tag{2.15}$$

To forecast the future value, the formula is

$$\hat{Y}_{t+p} = (L_t + pT_t)S_{t-s+p} \tag{2.16}$$

where s is the seasonal length and γ is the smoothing constant for estimating seasonality (0 < γ < 1). All smoothing constant can be changed to reach the minimum forecast error.

2.2.7. Forecast Error

Forecast error needs to be calculated to see if the forecast method used is suitable for the data or not, as a way of evaluation. The lower the forecast error, the more reliable a forecast result is. The most common way to calculate a forecast error is by calculating its mean squared error (MSE). MSE is useful to see how big the errors are (Hanke and Wichern, 2009).

$$MSE = \frac{1}{n}\sum_{t=1}^{n}(Y_t - \hat{Y}_t)^2 \qquad (2.17)$$

Since the errors are squared in MSE, the result might be too big if the errors are large so there is a forecast error evaluation called root mean squared error (RMSE). RMSE is the square root of MSE. It makes MSE easier to interpret (Hanke and Wichern, 2009).

Aside from MSE, there is mean percentage error (MPE). MPE is calculated to see if the forecast results are biased or not. Biased meaning the forecasted values are consistently forecasting values that are too low or too high from actual values. MPE cannot be calculated when the actual value of Y_t is zero. If there are zero values, it would be excluded from the calculation of MPE.

$$MPE = \frac{1}{n}\sum_{t=1}^{n}\frac{(Y_t - \hat{Y}_t)}{Y_t} \qquad (2.18)$$

The result of MPE is in percentage and can be in positive or negative number. The forecasted result is considered unbiased if MPE is close to zero. Forecasted result is considered to be consistently overestimating when MPE is a negatively large percentage and forecasted result is considered to be consistently underestimating when MPE is a positively large percentage (Hanke and Wichern, 2009).

2.2.8. Inventory Turnover Rate

Inventory turnover rate represents the number of times the number of average inventories in a certain period of time is sold, usually in a year (Berman and Evans, 2013). There are two methods to calculate inventory turnover rate. The first method is to calculate it by using costs, so unit of measurement is in currency. The second method is to calculate it using the quantity of inventory.

Inventory turnover (currency) is calculated by dividing cost of goods with average inventory investments from a certain period of time. While inventory turnover in unit is calculated by dividing quantity of demand with average number of inventories on hand from a certain period of time. For example, if turnover rate is calculated yearly, the average cost of investment and average number of inventories are calculated by taking the cost or number of inventories at the beginning of the year, add them with the cost or number of inventories at the end of the year, and divide it by two.

The ideal number of inventory turnover rate depends on the retailer. When inventory turnover rate wants to be Increased, it can be done by reducing the number of items with slow turnover.