Introduction to Time Series Analysis II

Chalermrat Nontapa

Outline

1. Smoothing Method
   1.1 Simple Smoothing (SS)
   1.2 Moving Average (MA)
   1.3 Exponential Smoothing (ES)
   1.4 Holt-Winters (HW)
2. Cases Study of Combined and Hybrid Forecasting
3. Interpret the results for time series data with some of my research on classical time series.

1. Smoothing Method

Smoothing Method is a time series method for forecasting univariate time series data. Time series methods work on the principle that a prediction is a weighted linear sum of past observations or lags.

The Exponential Smoothing time series method works by assigning exponentially decreasing weights for past observations. It is called so because the weight assigned to each demand observation is exponentially decreased.

1.1 Simple Smoothing (SS)

1.2 Moving Average (MA)

1.3 Exponential Smoothing (ES)

1.4 Holt-Winters (HW)

1.1 Simple Smoothing (SS)

The **Simple Smoothing (SS)** method is a popular data smoothing method because of the ease of calculation, flexibility, and good performance. It uses an average calculation for assigning the exponentially declining weights beginning with the most recent observation. This method is suitable for forecasting data **with no clear trend or seasonal pattern**.

$$\boxed{\begin{aligned} \hat{y}_t &= S_0 \quad ; \; t = 1 \\ \hat{y}_t &= S_{t-1} \quad ; \; t = 2, 3, \ldots, \mathrm{n} \end{aligned}} \qquad \ldots.(1.1)$$

Where, $S_t = \alpha y_t + (1-\alpha) y_{t-1} \quad ; \; t = 2, 3, \ldots, n \qquad \ldots.(1.2)$

Simple Exponential Smoothing (SES)

The **level** is the average value around which the demand varies over time. As you can observe in the figure below, the level is a smoothed version of the demand.

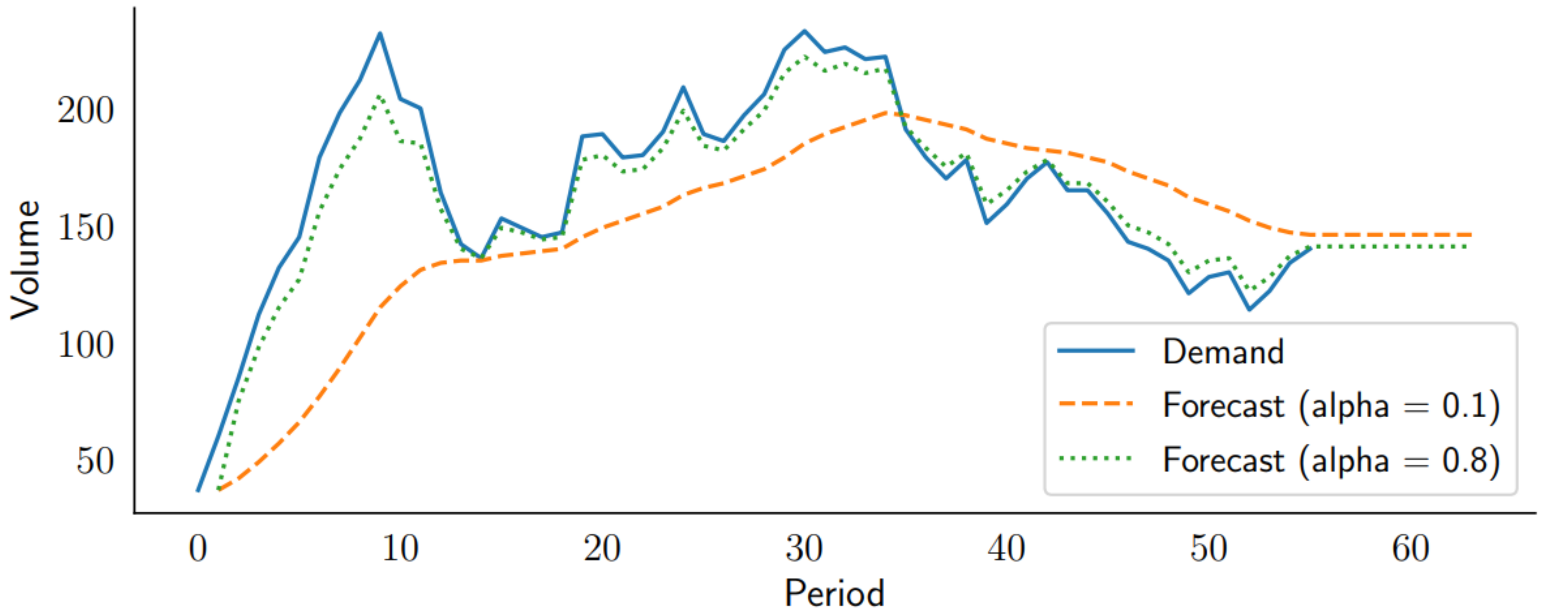


Figure 1.1: Characteristic of alpha on demand data set. [2]

Alpha

$0 \leq \alpha \leq 1$

Alpha (α) is a ratio of how much importance the model will allocate to the most recent observation compared to the importance of demand history.

1.2 Moving Average (MA)

This method is based on the premise that if values in a time series are averaged over a sufficient period, the effect of short-term variations will be reduced. The technique for finding a moving average for a particular observation is to find the average of the n observations before and after the observation itself

$$\hat{y}_t = \frac{1}{n}\left[y_{t-1} + y_{t-2} + \ldots + y_{t-n}\right] \;\; ; \;\; t > n \qquad \ldots.(1.3)$$

Moving Average (MA)

The classical method of time series decomposition originated in the 1920s and was widely used until the 1950s. It still forms the basis of many time series decomposition methods, so it is important to understand how it works. The first step in a classical decomposition is to use a moving average method to **estimate the trend-cycle**, so we begin by discussing moving averages.

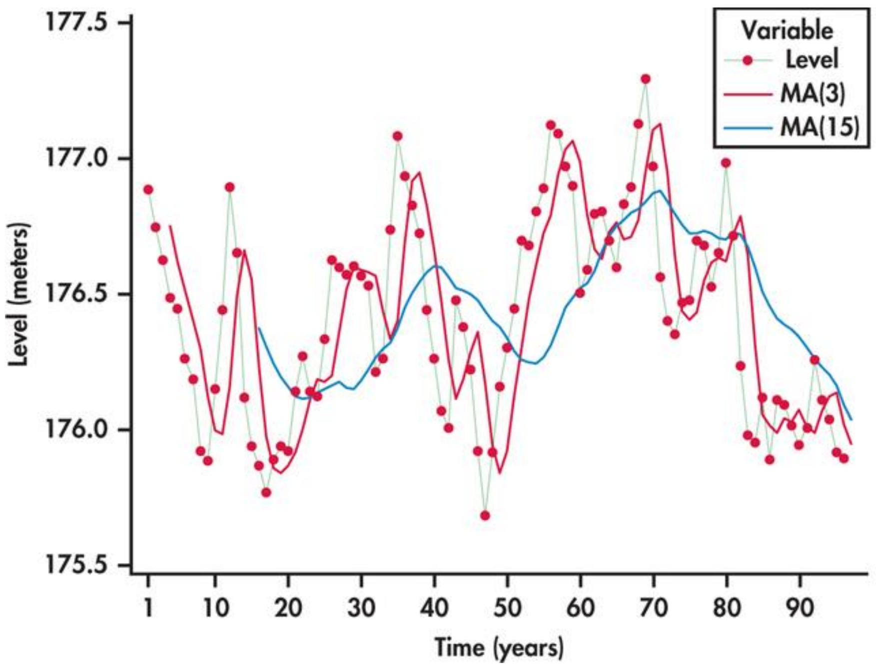


Figure 1.2: Characteristic of alpha on demand data set.

1.3 Exponential Smoothing (ES)

Exponential Smoothing was proposed in the late 1950s (Brown, 1959; Holt, 1957; Winters, 1960), and has motivated some of the most successful forecasting methods. Forecasts produced using exponential smoothing methods are weighted averages of past observations, with the weights decaying exponentially as the observations get older. In other words, the more recent the observation the higher the associated weight.

$$\hat{y}_{n+1} = wy_n + (1-w)wy_{n-1} + (1-w)^2 wy_{n-2} + \ldots + (1-w)^{n-1}\hat{y}_1 \quad \ldots.(1.3)$$

The weight $\boldsymbol{w}$ is called the smoothing constant for the exponential smoothing model and is traditionally assumed have a range of $0 \leq \boldsymbol{w} \leq 1$.

Exponential Smoothing (ES)

Simple Exponential Smoothing

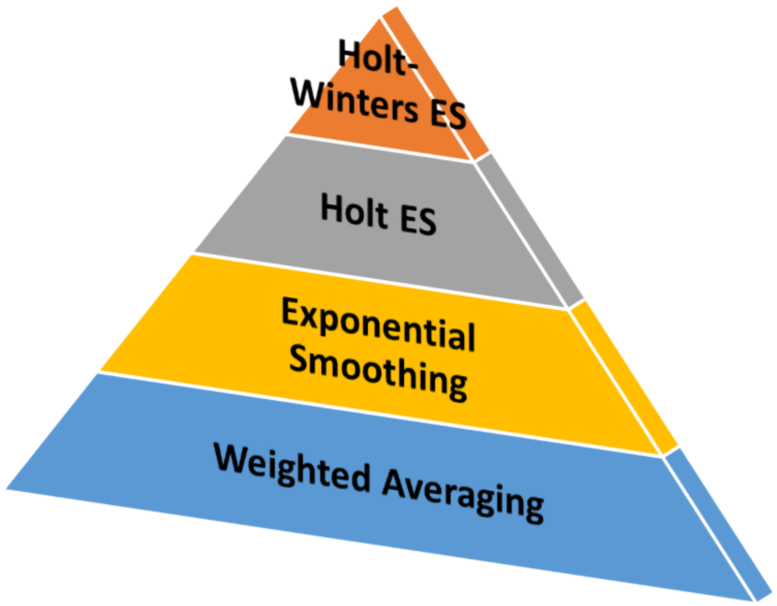
$$\begin{aligned} s_t &= \alpha x_t + (1-\alpha) s_{t-1} \\ &= \alpha x_t + \alpha(1-\alpha) x_{t-1} + (1-\alpha)^2 s_{t-2} \end{aligned}$$

Double Exponential Smoothing

$$\begin{aligned} s_t &= \alpha x_t + (1-\alpha)(s_{t-1} + b_{t-1}) \\ b_t &= \beta(s_t - s_{t-1}) + (1-\beta) b_{t-1} \end{aligned}$$

Tripple Exponential Smoothing

$$\begin{aligned} s_0 &= x_0 \\ s_t &= \alpha(x_t - c_{t-L}) + (1-\alpha)(s_{t-1} + b_{t-1}) \\ b_t &= \beta(s_t - s_{t-1}) + (1-\beta) b_{t-1} \\ c_t &= \gamma(x_t - s_{t-1} - b_{t-1}) + (1-\gamma) c_{t-L} \\ F_{t+m} &= s_t + m b_t + c_{t-L+1+(m-1) \bmod L} \end{aligned}$$



Choosing Alpha, Beta & Gamma with AutoES

Typically, these parameters can be found by back fitting the model to historical data and optimizing for various metrics e.g. **minimizing least squares, AIC, maximizing the likelihood** etc. Thankfully, there are many software packages available to do this quickly and easily.

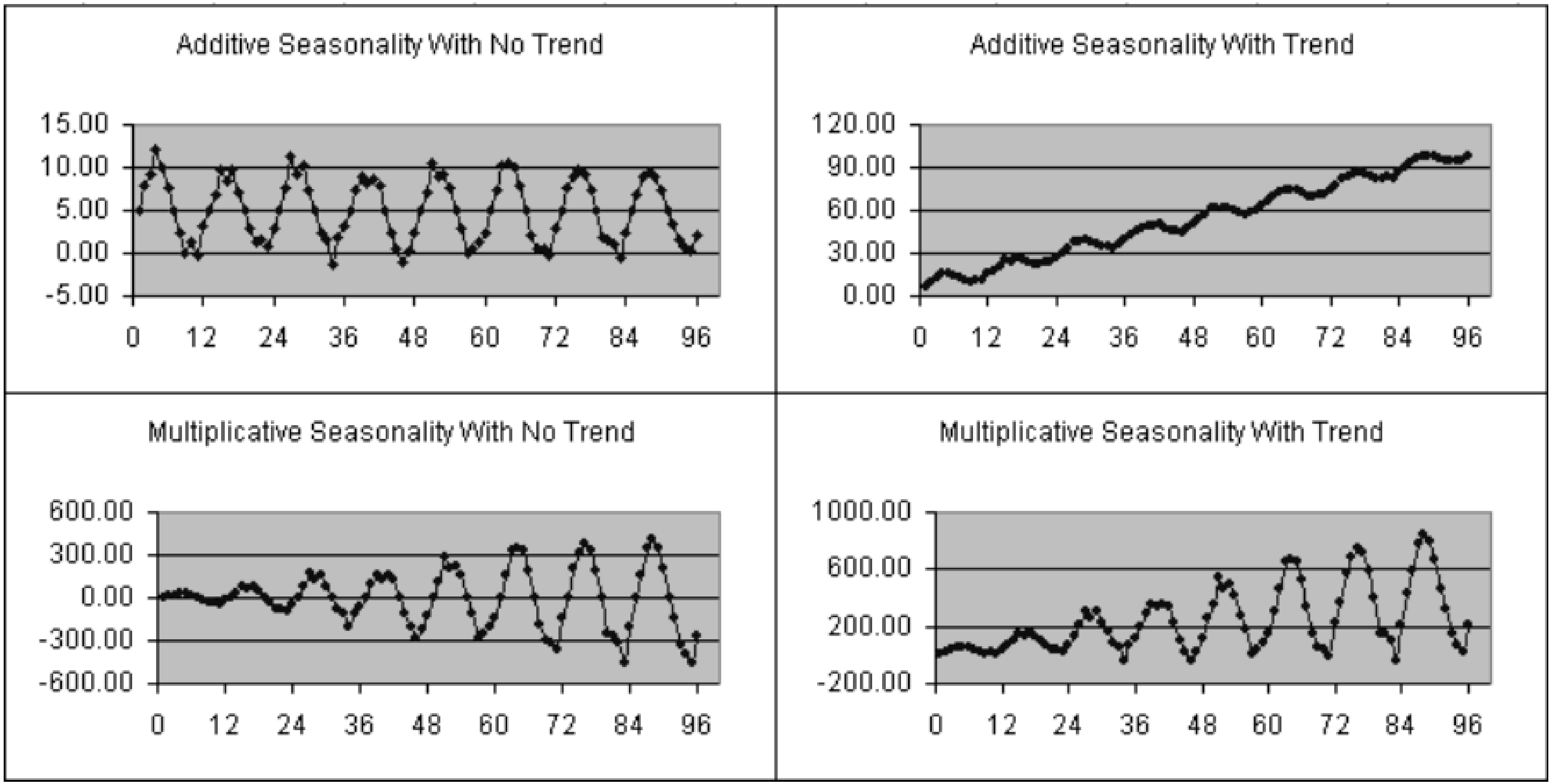
1.4 Holt – Winters (HW)

The Holt-Winters method is a well – known and effective approach for forecasting time series, particularly when **<u>Trend (T) and Seasonality (S)</u>** exist. The proper settings of the initial values for level, trend, and seasonality play an important role in this method and lead to better forecasting results.

1.4.1 The Exponential Smoothing by **Holt's** is suitable for forecasting data with trend and no seasonal pattern.

1.4.2 **Winter's Exponential Smoothing** as named after its two contributors: Charles Holt and Peter Winter's is one of the oldest time series analysis techniques which takes into account **the trend and seasonality** while doing the forecasting.

Type of Trend (T) and Seasonal (S)



Additive Model

$$Y_t = \boxed{L_t} + \boxed{S_t} + \varepsilon_t$$

level or linear trend component
can in turn be represented by $\beta_0 + \beta_1 t$

the seasonal adjustment

$S_t = S_{t+m} = S_{t+2m} = \ldots$ for $t = 1, \ldots, m - 1$

length of the season (period) of the cycles

Multiplicative Model

$$Y_t = \boxed{(L_t)}\boxed{(S_t)} + \varepsilon_t$$

level or linear trend component
can in turn be represented by $\beta_0 + \beta_1 t$

the seasonal adjustment

$S_t = S_{t+m} = S_{t+2m} = \ldots$ for $t = 1, \ldots, m - 1$

length of the season (period) of the cycles

Additive & Multiplicative Holt-Winter’s Method

Additive:

$$\hat{Y}_{t+h}(t) = \underbrace{\boxed{L_t}}_{\text{Level}} + \underbrace{\boxed{B_t}}_{\text{Trend}} h + \underbrace{\boxed{S}}_{\text{Seasonal}}{}_{t+h-m}$$

Multiplicative:

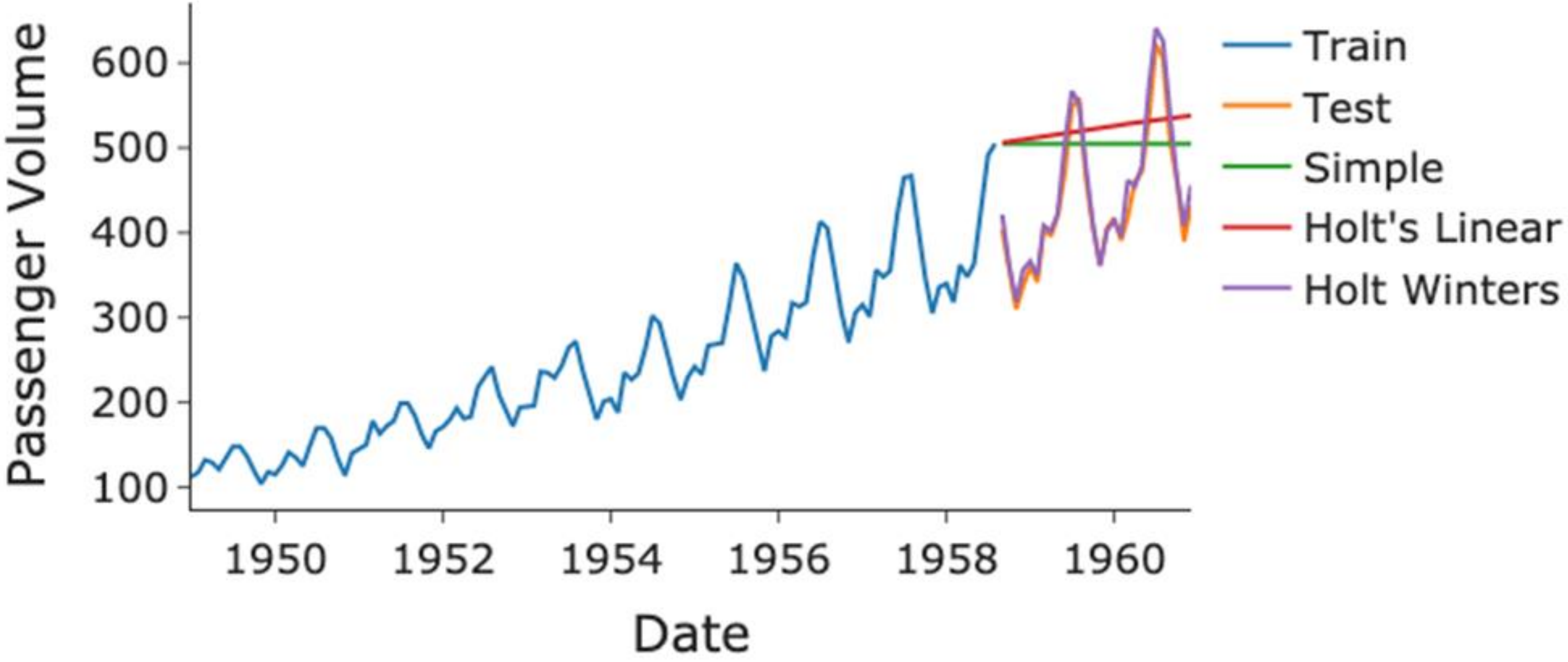
$$\hat{Y}_{t+h}(t) = (\overbrace{\boxed{L_t}}^{\text{Level}} + \overbrace{\boxed{B_t}}^{\text{Trend}} h)\overbrace{\boxed{S}}^{\text{Seasonal}}{}_{t+h-m}$$

Additive & Multiplicative Holt-Winter's Model

Suppose the time series is denoted by $y_1, \dots, y_n$ with m seasonal period.

	Additive	Multiplicative
Est. of level	$L_t = \alpha(Y_t - S_{t-m}) + (1-\alpha)(L_{t-1} + B_{t-1})$	$L_t = \alpha\left(\frac{Y_t}{S_{t-m}}\right) + (1-\alpha)(L_{t-1} + B_{t-1})$
Est. of trend	$B_t = \gamma(L_t - L_{t-1}) + (1-\gamma)B_{t-1}$	$B_t = \gamma(L_t - L_{t-1}) + (1-\gamma)B_{t-1}$
Est. of seasonal	$S_t = \delta(Y_t - L_t) + (1-\delta)S_{t-m}$	$S_t = \delta\left(\frac{Y_t}{L_t}\right) + (1-\delta)S_{t-m}$
Forecast	$\hat{Y}_{t+h}(t) = L_t + B_t h + S_{t+h-m}$	$\hat{Y}_{t+h}(t) = (L_t + B_t h)S_{t+h-m}$

Example Holt-Winter’s Model



2. Cases Study of Hybrid Forecasting and Combined

2.1 Relationship between the linear and nonlinear is an Additive Model.

1) Zhang (2003)

Model: ARIMA + ANN

$$y_t = L_t + N_t + \varepsilon_t$$

Data: Yearly; Sunspot, Canadian lynx, Weekly; BP/USD exchange rate

Accuracy: MSE, MAD testing with month, 6 months, and 12 months

2) Aladag (2009)

Model: ARIMA + ERNN

Data: Yearly; Canadian lynx

Accuracy: MSE testing with 14 observations

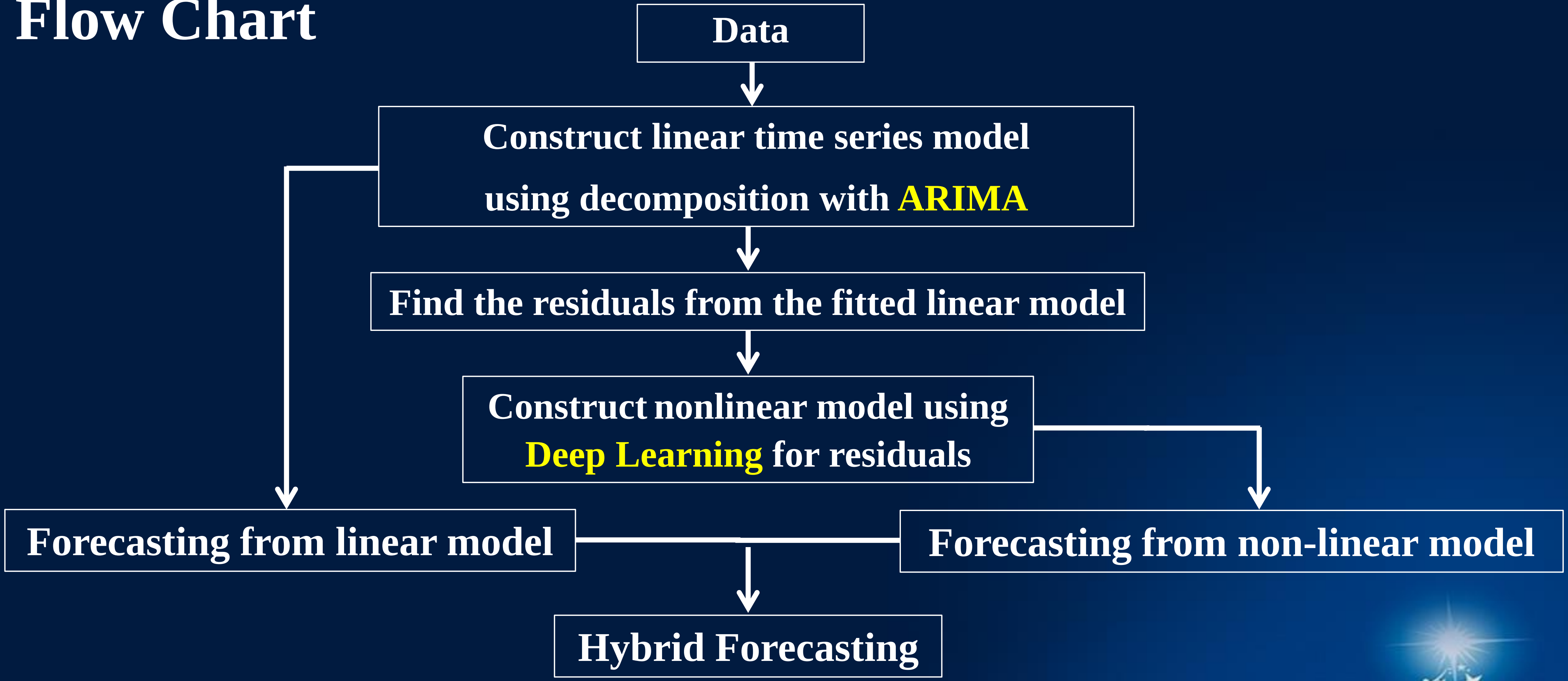
Problem Formulation

- Box-Jenskins method is the one of classical time series forecasting method, which estimates linear trend component.

- This approach is not appropriate because it has many errors in time series forecasting. Since, the mathematical models cannot explain behavior of nature of linear trend component.

- One way to describe the linear trend component in a time series data, we will fit the linear trend component using higher efficiency model.

Flow Chart



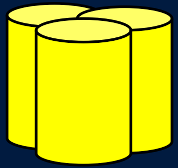
The proposed method is applied in order to combine the linear and nonlinear models together. This method includes three steps as follows.

Modeling linear part

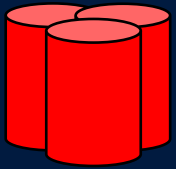
Cleansing Data

Stationary Property

Raw Data

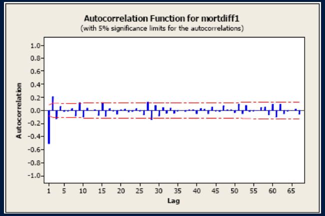


Training data

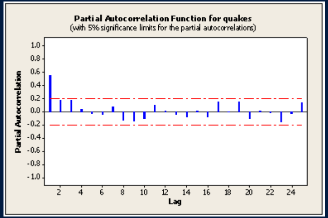


Model Learning

ACF



PACF

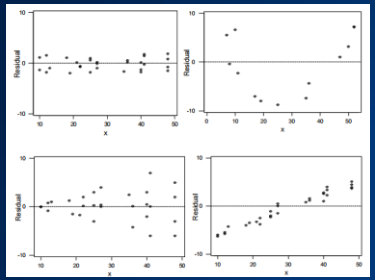


Possible Model

Approximate Parameter

Test Parameter

Residual Analysis



Forecasting

Linear Component

Modeling non-linear part

Preparing Data

$$e_t = f(e_{t-1}, e_{t-2}, \dots, e_{t-p})$$

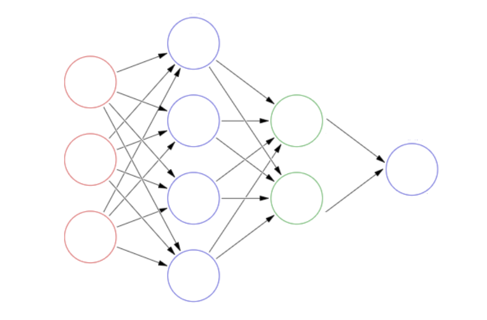
Find error

$$e_t = y_t - \hat{L}_t$$

Linear Component

Model Learning

Construct Network



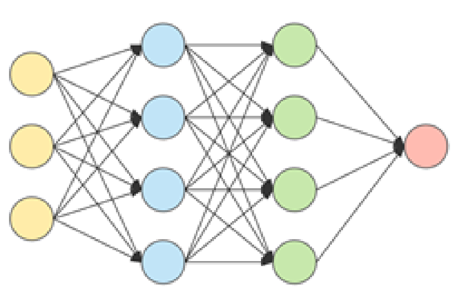
Loss Function $\ell(\hat{\boldsymbol{y}}, \boldsymbol{y}_i)$

Optimal parameter

$$\boldsymbol{\theta}^* = \arg\min_{\boldsymbol{\theta}} \sum_{i=1}^{N} \ell(f_{\boldsymbol{\theta}}(\boldsymbol{x}_i), \boldsymbol{y}_i)$$

$$\boldsymbol{\theta}^{(t+1)} = \boldsymbol{\theta}^{(t)} - \eta_t \nabla \ell(f_{\boldsymbol{\theta}}(\boldsymbol{x}_i), \boldsymbol{y}_i)$$

Optimal Network



Nonlinear Component

Forecasting

Hybrid Model

$$\hat{y}_t = \hat{L}_t + \hat{N}_t$$

Forecasting

Performance

2.2 Extend the input and non-linear model

3) Khashei and Bijari, (2011)

Model: ARIMA, GLAR + FNN

$$y_t = f\left(\hat{N}_t^1, \hat{L}_t, \hat{N}_t^2\right) = f\left(e_{t-1}, \ldots, e_{t-n}, \hat{L}_t, z_{t-1}, \ldots, z_{t-m}\right)$$

Data: Yearly; Canadian lynx

Accuracy: MSE testing with 14 observations

Nonlinear model

$$N_t = f\left(N_t^1, N_t^2\right)$$

$$N_t^1 = f^1\left(e_{t-1}, \ldots, e_{t-n}\right)$$

$$N_t^2 = f^2\left(z_{t-1}, \ldots, z_{t-m}\right)$$

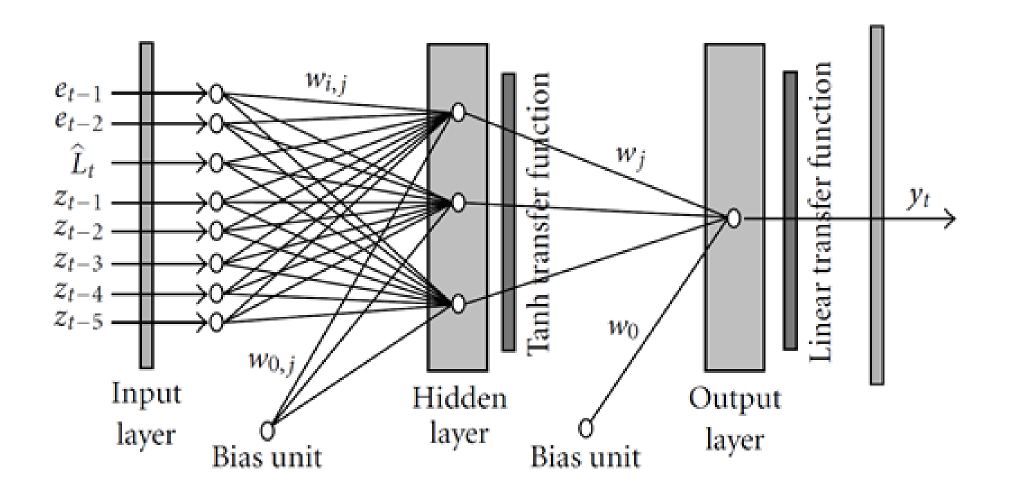


Figure 2.1: Model concept of Khashei and Bijari.

2.3 Linear Term can be modified by **Weighting Model.**

4) Khairalla et al., (2017)

Model: Weighting Linear model-ANN

$$L_t = c_1 L_{t,ARIMA} + c_2 L_{t,EXP} + c_3 L_{t,ANN} \;;\quad \sum_{i=1}^{3} c_i = 1$$

Data: Daily; exchange rate of Sudanese market 5 datasets

Accuracy: RMSE, MAE, MAPE, S.D. testing with P_{20}, P_{50}, P_{75}, and P_{90}.

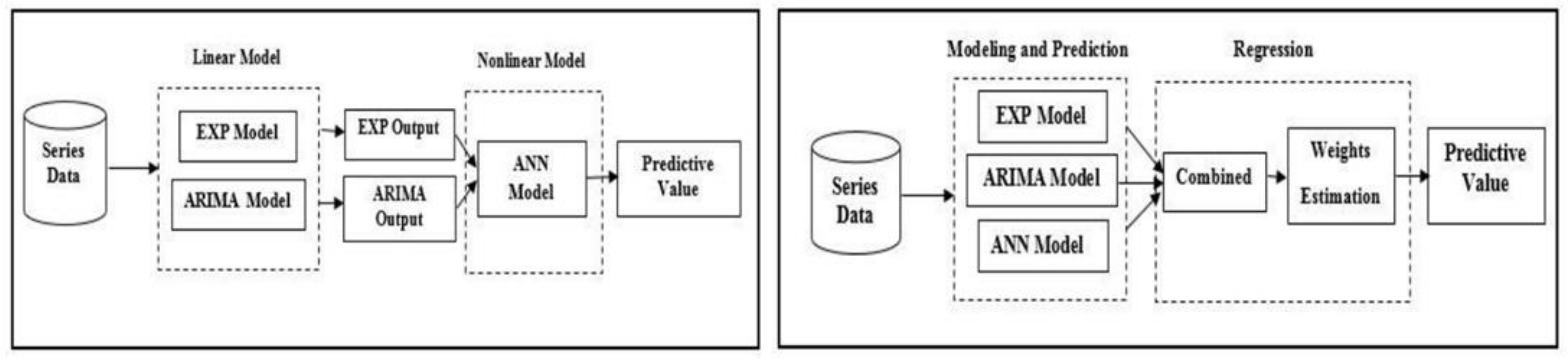


Figure 2.17: Model concept of Khairalla.

2.4 Linear Term can be modified by the Decomposition Method.

5) Suhartono and Rahayu, (2017)

Model: ARIMAX-ANN

$$L_t = \beta_1 t + \sum_{i=1}^{L} \alpha_i S_{i,t} + \sum_{j=1}^{M} \gamma_j V_{j,t} + \frac{\theta_q(B)\Theta_Q(B^L)}{\phi_p(B)\Phi_P(B^L)(1-B)^d(1-B^L)^D}\varepsilon_t$$

Data: Monthly; currency inflow and outflow at Bank Indonesia in West Java

Accuracy: RMSE testing with 12 months

Four types of simulation data are used in this research

Scenario 1: trend, seasonal homogenous, calendar variation, and linear noise patterns.

Scenario 2: trend, seasonal homogenous, calendar variation, and nonlinear noise patterns.

Scenario 3: trend, seasonal heterogenous, calendar variation, and linear noise patterns.

Scenario 3: trend, seasonal heterogenous, calendar variation, and nonlinear noise patterns.

2.5 Compare the two types of switching hybrid model

6) Khashei et al., (2018)

Model: ARIMA-MLP & MLP-ARIMA

Data: Monthly; DJIAI opening index, SZII closing index
Daily; Nikkei 225 index

Accuracy: MSE, RMSE, MAE, MAPE testing with 20%-25% of samples.

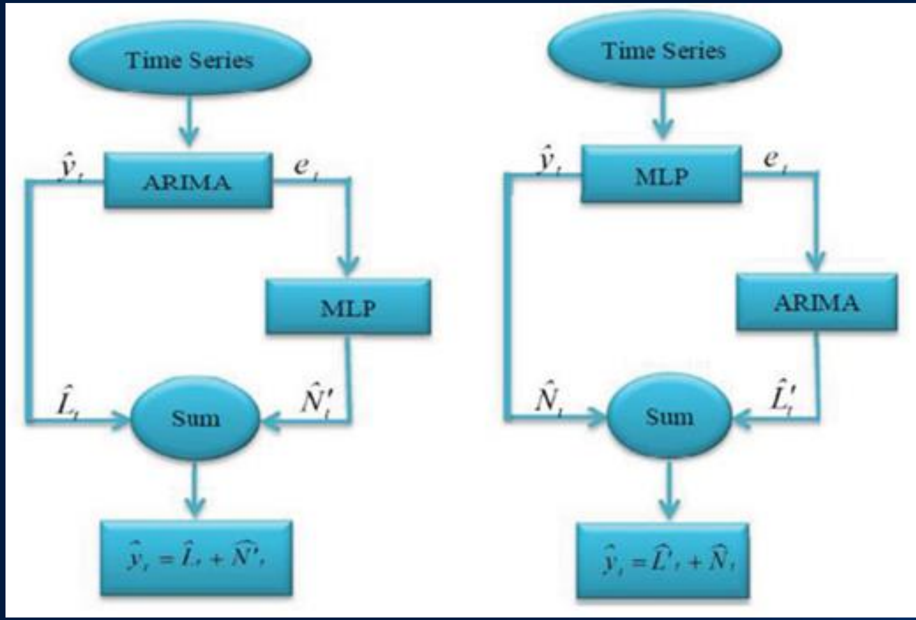


Figure 2.2: Model concept of Khashei.