## BAB III

## METODOLOGI PENELITIAN

### 3.1 Prosedur Penelitian

### 3.1.1 Langkah-langkah Analisis Variasi Kalender

1. Plot data jumlah penumpang kereta api Malabar
2. Menentukan variabel *dummy*
3. Melakukan analisis regresi linear berganda

   Langkah ini digunakan untuk memperoleh koefisien regresi *dummy*, dilakukan dengan program Eviews 9.0
4. Melakukan uji asumsi regresi linear berganda dan uji signifikansi koefisien regresi

   Uji signifikansi koefisien regresi yaitu uji F dan uji T dilakukan untuk mengetahui pengaruh variabel-variabel bebas terhadap variabel terikat.
5. Model persamaan efek variasi kalender

### 3.1.2 Langkah-langkah Analisis ARIMAX

1. Melakukan identifikasi model, yaitu dengan mengidentifikasi plot data, fak dan fakp. Jika data belum stasioner, maka dilakukan transformasi data dengan cara melakukan differencing terhadap data.
2. Menganalisis model yang didapat dengan variabel *dummy* yang sudah ditentukan.
3. Melakukan pengujian signifikansi dan pemodelan data runtun waktu.
4. Melakukan pemeriksaan diagnostik sehingga data dari model yang didapat dan residualnya mencapai kondisi *white noise* dan berdistribusi normal.
5. Melakukan evaluasi model dengan kriteria *Akaike's Information Criterion* (AIC).
6. Melakukan peramalan

### 3.1.3 Langkah-langkah Analisis Model ARIMAX Dua Level

Prosedur pembangunan model ARIMAX dapat dijelaskan sebagai berikut :

1. Penentuan variabel *dummy* untuk periode variasi kalender.
2. Hapus efek variasi kalender dari respons dengan tepat

$$Y_t = \beta_0 + \sum_j \alpha_j D_{j,t} + \sum_j \gamma_j D_{j,t-1} + N_t \quad (1)$$

untuk model dengan tren stokastik dan musiman, atau dengan tepat

$$Y_t = \delta_1 t + \beta_1 M_{1,t} + \beta_2 M_{2,t} + \cdots + \beta_s M_{s,t} + \sum_j \alpha_j D_{j,t} + \sum_j \gamma_j D_{j,t-1} + N_t \quad (2)$$

secara bersamaan untuk model dengan tren deterministik dan musiman, untuk mendapatkan *error* $N_t$.

3. Temukan model ARIMA $N_t$ terbaik menggunakan prosedur Box-Jenkins.
4. Urutkan model ARIMA yang ditemukan dari langkah 3 digunakan untuk memodelkan data nyata dan variabel *dummy* dari efek variasi kalender sebagai variabel input secara bersamaan sebagai persamaan (1) dan (2) untuk model dengan tren stokastik dan deterministik dan musiman. Model ini adalah model variasi kalender tingkat pertama berdasarkan metode ARIMAX.
5. Uji signifikansi parameter dan lakukan pemeriksaan diagnostik hingga prosesnya stasioner dan $\varepsilon_t$ muncul sebagai proses *white noise*.
6. Perkirakan model tingkat kedua untuk memprediksi efek variasi kalender dalam setiap jumlah hari sebelum perayaan idul fitri. Model tingkat kedua ini mengandung dua efek evaluasi penjualan, yaitu efek penjualan pada satu bulan sebelum dan selama bulan perayaan Idul Fitri. Pada langkah ini, bentuk fungsional linier dari model adalah sebagai berikut :
   a. Model untuk mengevaluasi dan memperkirakan efek penjualan selama bulan perayaan Idul Fitri, yaitu :

$$\hat{\alpha}_j = v_0 + v_{1j} \quad (3a)$$

Dimana j adalah jumlah hari sebelum perayaan Idul Fitri pada bulan tertentu. Dalam model ini, variabel respon adalah estimasi koefisien regresi dalam persamaan (1) atau (2) yaitu $\hat{\alpha}_0, \hat{\alpha}_2, \ldots, \hat{\alpha}_{29}$.

b. Model untuk mengevaluasi dan memperkirakan efek penjualan pada satu bulan sebelum perayaan Idul Fitri, yaitu :

$$\hat{\gamma}_j = \omega_0 + \omega_{1j} \qquad (3b)$$

Dimana j adalah jumlah hari sebelum perayaan Idul Fitri pada bulan tertentu. Dalam model ini, variabel respon adalah estimasi koefisien regresi dalam persamaan (1) atau (2) yaitu $\hat{\gamma}_0, \hat{\gamma}_2, \dots, \hat{\gamma}_{29}$.

## 3.2 Pengumpulan Data

Data yang digunakan dalam penelitian ini adalah data sekunder yang diperoleh dari Kantor Pusat DAOP 2 PT Kereta Api Indonesia. Data dibagi menjadi dua bagian yaitu data *in* sampel sebanyak 72 data dari bulan Januari 2013 hingga Desember 2018 dan data *out* sampel sebanyak 8 data terakhir dari bulan Januari 2019 hingga bulan Desember 2019 yang akan digunakan dalam perhitungan kebaikan model.

Variabel data ini difokuskan pada jumlah penumpang kereta api Malabar jurusan Bandung-Malang selama bulan Januari 2013 hingga Desember 2019.

## 3.3 Teknik Analisis Data

### 3.3.1 Model ARIMAX Variasi Kalender Trend Deterministik

Model ARIMAX variasi kalender *trend* deterministik adalah model ARIMA yang diberi tambahan dugaan sebagai efek dari variasi kalender. Variabel yang mewakili efek variasi kalender adalah variabel *dummy* yang diperoleh.

Data dengan variasi kalender *trend* deterministik dapat dimodelkan dengan menggunakan analisis regresi linear berganda. Model regresi linier berganda untuk data yang memiliki variasi kalender *trend* deterministik dapat dituliskan sebagai berikut (Bell & Hillmer, 1983) :

$$Y_t = \gamma t + \beta_1 D_{1,t} + \cdots + \beta_k D_{k,t} + N_t$$

dengan:

t : waktu dari pengamatan pertama hingga ke-*n*. Variabel *t* menunjukkan apabila terdapat *trend* linier dalam data.

$D_{k,t}$ : variabel *dummy* untuk efek variasi kalender ke-,

$N_t$ : sisaan model ARIMAX

### 3.3.2 Model ARIMAX Dua Level dengan Efek Variasi Kalender

Model ARIMA adalah model umum untuk peramalan data. Model ARIMA dapat dinyatakan sebagai berikut (Suhartono, Lee, & Prastyo, 2015) : $\Theta$

$$\emptyset_p(B)\Phi_p(B^s)(1-B)^d(1-B^s)^D Y_t = \theta_q(B)\ \Theta_Q(B^S)\varepsilon_t \qquad (4)$$

dimana:

$\emptyset_p(B) \quad = 1 - \emptyset_1 B - \emptyset_2 B^2 - \cdots - \emptyset_p B^p$

$\Phi_p(B^s) \quad = 1 - \Phi_1 B - \Phi_2 B^{2S} - \cdots - \Phi_p B^{PS}$

$\theta_q(B) \quad = 1 - \theta_1 B - \theta_2 B^2 - \cdots - \theta_p B^q$

$\Theta_Q(B^S) \quad = 1 - \Theta_1 B - \Theta_2 B^{2S} - \cdots - \Theta_p B^{QS}$

Dalam penelitian ini, ada dua jenis variabel tambahan, yaitu variabel *dummy* untuk efek variasi kalender saja, dan variabel *dummy* untuk efek variasi kalender dan tren deterministik. Model pertama dikenal sebagai ARIMAX dengan tren stokastik serta menerapkan perbedaan non musiman dan / musiman, dan yang kedua dengan tren deterministik. Model ARIMA musiman umum, yang ditunjukkan dalam persamaan (1) dapat ditulis sebagai berikut :

$$Y_t = \frac{\theta_q(B)\ \Theta_Q(B^S)}{\emptyset_p(B)\Phi_p(B^s)(1-B)^d(1-B^s)^D}\varepsilon_t \qquad (5)$$

Selain itu, model ARIMAX dengan tren stokastik serta musiman ialah :

$$Y_t = \sum_j \alpha_j D_{j,t} + \sum_j \gamma_j D_{j,t-1} + \frac{\theta_q(B)\ \Theta_Q(B^S)}{\emptyset_p(B)\Phi_p(B^s)(1-B)^d(1-B^s)^D}\varepsilon_t \quad (6)$$

Terdapat pula model ARIMAX dengan tren deterministik serta musiman ialah :

$$Y_t = \delta_1 t + \beta_1 M_{1,t} + \beta_2 M_{2,t} + \cdots + \beta_s M_{s,t} + \sum_j \alpha_j D_{j,t} + \sum_j \gamma_j D_{j,t-1} + \frac{\theta_q(B)}{\emptyset_p(B)}\varepsilon_t \qquad (7)$$

Dimana $M_{s,t}$ adalah variabel dummy bagi satuan waktu yang sesuai untuk menangkap tren deterministik.

### 3.3.3 Pemeriksaan Diagnostik

Menurut (Wei, 2006) pemeriksaan diagnostik perlu dilakukan, untuk mengetahui apakah data residual telah memenuhi asumsi *white noise* dan berdistribusi normal.

1. *White Noise*

   Proses $\alpha_t$ disebut proses white noise apabila tidak ada korelasi dalam variabel acak dengan nilai mean konstan, hipotesis yang digunakan dalam pengujian ini adalah sebagai berikut :

   Hipotesis

   $H_0 : \rho_1 = \rho_2 = \cdots = \rho_k = 0$

   $H_1$ : terdapat $\rho_i \neq \rho_j$ ; i,j = 1,2, ..., k; i = j

2. Distribusi Normal

   Untuk melihat distribusi data memenuhi asumsi normalitas atau tidak dapat dilakukan dengan melihat plot residual atau dapat juga dilakukan dengan pengujian *Kolmogorov-Smirnov*. Hipotesis yang digunakan untuk menguji data dengan uji *Kolmogorov-Smirnov*.

   Hipotesis

   $H_0$ : $F(x) = F_0(x)$ (Residual berdistribusi normal)

   $H_1$ : $F(x) \neq F_0(x)$ (Residual berdistribusi normal)

   $Statistik\ Uji$

   $D = SUP\ X\ |S(x) - F_0(x)|$

   Dimana :

   S(x) : Fungsi peluang kumulatif yang dihitung dari data sampel

   $F_0(x)$ : Fungsi peluang kumulatif dari distribusi normal

   SUP X : Nilai supremum untuk semua x dari $|S(x) - F_0(x)|$

   Dengan kriteria : tolak $H_0$ jika $D > D_{(1-\alpha,n)}$ atau p-value < $\alpha$