BAB II

KAJIAN TEORI

A. Statistik Dasar

1. *Average* (Rata-rata)

Menurut Spiegel,dkk (1996:45) rata-rata yaitu sebuah nilai yang khas atau yang mewakili suatu himpunan data.

Menurut Supranoto (2001:14) Rata – rata (μ) dari distribusi probabilitas adalah nilai harapan dari variabel acaknya, dimana nilai harapan variabel acak diskrit adalah rata – rata tertimbang terhadap seluruh kemungkinan hasil dimana penimbangnya adalah nilai probabilitas yang dihubungkan dengan setiap hasil.

$$\mu = \sum_{i=1}^{n} x_i \, p(x_i)$$

Dimana :

μ = rata-rata
$\sum_{i=1}^{n}$ = banyak data dari data ke 1 sampai data ke N
x_i = nilai ke *i* dari variabel acak X
$p(x_i)$ = probabilitas terjadinya x_i

2. Standar Deviasi (σ)

Standar deviasi adalah penyimpangan data dari rata-ratanya atau akar kuadrat dari varians.

Standar deviasi dirumuskan sebagai berikut :

$$\sigma = \sqrt{E\,(X - \mu)^2} \tag{2.1}$$

Atau

$$\sigma = \sqrt{\sum_{i=1}^{N} (x_i - \mu)^2\, p\,(x_i)}$$

Dimana :

σ	= standar deviasi
$E\ atau \sum_{i=1}^{n}$	= banyak data dari data ke 1 sampai data ke N
x_i	= nilai ke *i* dari variabel acak X
μ	= rata-rata
$p(x_i, y_i)$	= probabilitas terjadinya x_i dan y_i

3. **Varian**

Menurut Supranoto (2001:15) Varian dapat diartikan sebagai standar deviasi yang di kuadratkan, atau rata-rata tertimbang dari kuadrat selisih antara setiap kemungkinan hasil dan rata-ratanya, dengan rumus :

$$\sigma^2 = E(X - \mu)^2 \tag{2.2}$$

Atau

$$\sigma^2 = \sum_{i=1}^{N} (x_i - \mu)^2\, p\,(x_i)$$

Dimana :

σ^2	= varian
$E\ atau \sum_{i=1}^{n}$	= banyak data dari data ke 1 sampai data ke N
x_i	= nilai ke *i* dari variabel acak X
μ	= rata-rata
$p(x_i, y_i)$	= probabilitas terjadinya x_i dan y_i

4. **Kovarian**

Menurut Supranoto (2001:19) Kovarian adalah suatu pengukuran statistik untuk mengukur berapa jauh hubungan antara dua variabel yang bergerak bersama. Jika kedua kombinasi variabel bergerak searah maka menunjukan kovarian positif begitu juga sebaliknya, jika kovarian negative maka akan berlawanan.

Rumus kovarian :

$$Cov = \sum_{i=1}^{n} \left[X_{i} - E(X)\right]\left[Y_{i} - E(Y)\right] p(x_{i}, y_{i}) \quad (2.3)$$

Dimana :

X_i = nilai variabel X ke *i*
Y_i = nilai variabel Y ke *i*
$p(x_i, y_i)$ = probabilitas terjadinya x_i dan y_i
i = 1,2,…., N

5. Koefisien Korelasi

Menurut Makridakis (1995:189) Koefisien korelasi (r) adalah suatu ukuran asosiasi (linier) relative antara dua variabel dan dapat bervariasi dari 0 (yang menunjukkan tidak ada korelasi) hingga ± 1 (yang menunjukan korelasi sempurna). Jika korelasi lebih besar dari 0, dua variabel dikatakan berkorelasi positif dan jika kurang dari 0 dikatakan berkorelasi negative.

Koefisien korelasi (r) dapat dirumuskan sebagai berikut :

$$\rho_{k} = \frac{cov_{XY}}{\sqrt{var\, X\, var\, Y}}$$

Dimana :

P_k = koefisien korelasi
Cov $_{xy}$ = nilai kovarian dari X dan Y
Var X,Y = nilai varian dari X dan Y

B. Kestasioneran Data

Dalam analisis deret berkala, data harus berada dalam keadaan stasioner, yaitu tidak terdapat perubahan yang sistematis dalam mean (tidak terdapat suatu kecenderungan dalam mean) dan varians. Untuk memeriksa kestasioneran ini dapat digunakan diagram deret waktu (*time series plot*) yaitu diagram pencar antara nilai peubah Z_t dengan waktu *t*. Jika diagram deret waktu berfluktuasi di sekitar garis yang sejajar sumbu waktu (*t*) maka dikatakan deret stasioner dalam rata – rata.

Menurut Makridakis dkk (2006), konsep stasioner ini secara statistik dapat dijelaskan sebagai berikut :

a. Apabila suatu data deret waktu dibuat diagramnya dan kemudian tidak ada perubahan rata-rata yang jelas dari waktu ke waktu, dikatakan bahwa deret data tersebut stasioner pada rata-rata.

b. Apabila diagram data deret waktu tidak memperlihatkan adanya perubahan variansi (ragam) yang jelas dari waktu ke waktu, dapat dikatakan deret data tersebut stasioner pada variansi.

c. Apabila rata-rata mengalami perubahan dari waktu ke waktu, dengan kata lain menyimpang dengan beberapa pola siklus kecenderungan (*trend cycle*), maka deret waktu tersebut mempunyai rata-rata yang tidak stasioner.

d. Apabila rata-rata suatu data deret waktu menyimpang (berubah setiap waktu) dan variansi (simpangan bakunya) tidak konstan setiap waktu, deret waktu teresbut memiliki rata-rata dan variansi yang tidak stasioner.

C. *Differencing*

Differncing adalah menghitung perubahan atau selisih nilai observasi. Proses ini dilakukan apabila data tidak stasioner yaitu dengan mengganti data asli dengan perbedaan data asli tersebut, atau dapat dirumuskan sebagai berikut :

$$B^{d} Z_{t} = Z_{t-d} \tag{2.11}$$

Dengan kata lain, notasi B yang dipasang pada Z_t mempunyai pengaruh menggeser data pada d pada periode ke belakang. Sebagai deret stasioner, model umumnya adalah sebagai berikut :

$$W_t = (1-B)^d Z_t \tag{2.12}$$

Dimana :

d	: 1, 2, … (biasanya 1 dan 2)
$(1-B)^d$	: orde *differencing* non musiman
B	: backshift operator (operator mundur)
Z_t	: observasi pada waktu ke t

Differencing dilakukan hingga data tersebut stasioner, jika varians tidak stasioner maka dilakukan transformasi.

D. *White Noise*

Proses α_t disebut proses *white noise* apabila tidak ada korelasi dalam variabel acak dengan nilai mean konstan E(α_t) = μ_a, nilai mean ini biasanya diasumsikan sebagai nol, variansi konstan (α_t) = σ_a^2 dan γ_k = cov(α_t, α_{t+k}) = 0, untuk k ≠ 0. Dari definisi tersebut dapat diketahui bahwa proses *white noise* adalah stasioner dengan fungsi autokovarian (sukarna : 2006)

Fungsi autokovariani :

$$\gamma_k = \begin{cases} \sigma_a^2, & untuk\ k = 0 \\ 0, & untuk\ k \neq 0 \end{cases}$$

Fungsi autokorelasi :

$$\rho_k = \begin{cases} 1, & untuk\ k = 0 \\ 0, & untuk\ k \neq 0 \end{cases}$$

Fungsi autokorelasi parsial :

$$\phi_{kk} = \begin{cases} 1, & untuk\ k = 0 \\ 0, & untuk\ k \neq 0 \end{cases}$$

E. Regresi

Regresi merupakan model statistik yang mempelajari tentang hubungan ketergantungan dari sebuah variabel, yaitu variabel dependen terhadap satu atau lebih variabel yang lain yaitu variabel-variabel penjelas (variabel independen) dengan tujuan untuk menaksir dan/atau meramal nilai rata-rata variabel dependen dengan dasar nilai tertentu dari variabel penjelas.

Model regresi umumnya dapat di definisikan sebagai berikut :

$$Y = \alpha + \beta X \qquad (2.13)$$

Dimana :

$$\beta = \frac{Cov_{xy}}{Var_{x}} = \frac{n\sum XY - (\sum X)(\sum Y)}{n\sum X^2 - (\sum X)^2}$$

dan

$$a = \frac{\sum Y}{n} - b\frac{\sum X}{n} = \bar{Y} - \beta\bar{X}$$

α dan β adalah parameter – parameter tetap (tetapi tidak diketahui), X diasumsikan sebagai suatu ukuran tanpa kesalahan.

Namun dalam prakteknya regresi yang digunakan adalah :

$$Y_i = \alpha + \beta X_i + e_i \text{ untuk } i = 1, 2, \ldots, n, \qquad (2.14)$$

Dimana α dan β merupakan variabel random, X tak mungkin diukur tanpa kesalahan, e_i adalah kesalahan taksiran untuk observasi ke *i*.

F. Autoregressive (AR) dan Moving Average (MA)

1. Model Autoregressive (AR)

Model autoregressive adalah model hasil regresi tetapi tidak berfungsi untuk menghubungkan nilai variabel bebas dengan variabel tak bebas, melainkan menghubungkan nilai-nilai sebelumnya dengan variabel pada *time lag* (selang waktu) yang bermacam-macam. Jadi, model *autoregressive* menyatakan suatu ramalan sebagai fungsi nilai-nilai sebelumnya dari deret berkala tertentu (Makridakis,dkk, 1995:513)

Pada autoregressive berorde p menyatakan bahwa nilai pengamatan pada periode ke-*t* (Z_t) merupakan hasil regresi dari nilai-nilai pengamatan sebelumnya selama p periode. Bentuk umum model *autoregressive* orde p, AR (p) adalah:

$$Z_t = \phi_1 Z_{t-1} + \phi_2 Z_{t-2} + \ldots + \phi_p Z_{t-p} + \alpha_t \quad (2.5)$$

Dimana :

Z_t : nilai variabel dependen waktu t.

Z_{t-p} : variabel independent yang dalam hal ini merupakan *lag* (beda waktu) dari variabel dependen pada satu periode sebelumnya hingga p periode sebelumnya.

α_t : nilai kesalahan.

ϕ_1 , ϕ_2 , ϕ_3 : parameter dari model autoregressive

(Makridakis,dkk, 1995:385)

2. Model Moving Average (MA)

Dalam peramalan Box_jenkis model *moving average* (rata-rata bergerak) dapat diartikan bahwa nilai deret berkala pada waktu t dipengaruhi unsur kesalahan pada saat ini dan unsur kesalahan pada masa lalu (Sukarna, 2006:35).

Bentuk umum model rata-rata bergerak orde q, atau MA (q) adalah

$$Z_t = \alpha_t - \theta_t \alpha_{t-1} - \theta_2 \alpha_{t-2} - \ldots - \theta_q \quad Z_{t-q} \tag{2.6}$$

Dimana :

Z_t : variabel dependen pada waktu t.

$\theta_1, \theta_2, \theta_3$: koefisien orde q

α_t : suatu proses white noise atau galat pada waktu ke t yang diasumsikan mempunyai rata-rata 0 dan variansi konstanta $\sigma^2 \alpha$

(Sukarna, 2006:55)

3. Model Campuran Autoregresive-Moving Average (ARMA)

Model ini merupakan model campuran/gabungan kedua model autoregressive (AR) dan rata-rata bergerak (MA). Bentuk umum model autoregressive rata-rata bergerak atau disebut ARMA (p,q) adalah :

$$Z_t = \phi_1 Z_{t-1}. + \phi_2 Z_{t-2} + \ldots + \phi_p \ Z_{t-p} + \alpha_t - \theta_t \alpha_{t-1} - \theta_2 \alpha_{t-2} - \ldots - \theta_q \ Z_{t-q} \tag{2.7}$$

G. Fungsi Autokorelasi dan Fungsi Autokorelasi Parsial

Nilai tengah (mean) dan varian dari suatu data deret berkala mungkin tidak bermanfaat apabila deret tersebut tidak stasioner, karena statistic kunci dari deret waktu adalah koefisien korelasi (korelasi deret berkala dengan deret berkala itu sendiri dengan selisih waktu (lag) 0,1,2 periode atau lebih).

akan tetapi nilai minimum dan maksimum dapat digunakan untuk tujuan ploting. Koefisien korelasi adalah suatu fungsi yang menunjukan besarnya korelasi (hubungan linier) antara pengamatan pada waktu ke t dengan pengamatan pada waktu – waktu sebelumnya.

Nilai autokorelasi untuk *time lag 1,2,3,…, k* adalah :

$$\rho_k = \frac{\sum_{t=1}^{n-k}(Z_t - \bar{Z})(Z_{t+k} - \bar{Z})}{\sum_{t=1}^{n}(Z_t - \bar{Z})^2}$$

Di dalam analisis regresi, apabila variabel tidak bebas Y diregresikan kepada variabel terhadap peubah-peubah bebas X_1 dan X_2 akan timbul pertanyaan " Sejauh mana peubah X_1 mampu menerangkan keadan dari Y apabila mula-mula X_2 dipisahkan (*partialled out*) ". Ini berarti meregresikan lagi nilai sisa tersebut atas Y atas X_2 dan menghitung kesalahan sisa (*residual eror*), kemudian meregresikan lagi nilai sisa tersebut atas X_1.

Autokorelasi parsial digunakan untuk tingkat keeratan (association) antara Z_t dan Z_{t-k}, apabila pengaruh dari lag waktu (time lag) *1, 2, 3, … , k–1* dianggap terpisah. Fungsi autokorelasi parsial (PCAF) adalah suatu fungsi yang menunjukan besarnya korelasi parsial antara pengamatan waktu ke *t* (dinotasikan denga Z_t) dengan pengamataan pada waktu-waktu sebelumnya (dinotasikan dengan Z_{t-k}, Z_{t-1}, Z_{t-2} … , Z_{t-k}.

Rumus autokorelasi parsial (ϕ_{kk}) adalah :

$$\phi_{kk} = \text{corr}(Z_t, Z_{t+k} | Z_{t-1}, Z_{t-2}, \dots, Z_{t+k-1}) \tag{2.9}$$

Dimana $Z_{t+k} = \phi_{k1} Z_{t+k+1} + \phi_{k2} Z_{t+k+2} + \ldots + \phi_{kk} Z_t + \alpha_{t+k}$ sedangkan ϕ_{ki} adalah parameter regresi ke-*i* dan α_{t+k} adalah bentuk sisa (*error term*) yang tidak berkorelasi dengan Z_{t+k-j} untuk $j \geq 1$.

Kestasioneran data dapat diuji dengan cara plot data dan menghitung *autocorrelation function* (ACF). Jika autokorelasinya cenderung turun lambat atau mendekati garis lurus maka dapat dikatakan bahwa data belum stasioner dalam rata-rata, sebaliknya jika autokorelasinya cenderung turun secara drastis sampai nol setelah time lag dua atau tiga, maka data dapat dikatakan stasioner (Sukarna, 2006:25).

Pola ACF dan PACF bisa berpola *cut off* dan *dies down*. Pola *cut off* adalah pola ketika garis ACF dan PACF pnifikan pada lag pertama atau kedua tetapi kemudian tidak ada garis ACF dan PACF yang pnifikan pada lag berikutnya. Untuk pola *cut off*, perbedaan antara ACF dan PACF yang pnifikan dengan yang tidak pnifikan adalah besar sehingga garis ACF dan PACF terlihat terpotong (cut off). ACF dan PACF dikatakan memiliki perilaku *dies down* jika kedua fungsi tersebut tidak terpotong, melainkan menurun secara bertahap. Bentuk penurunannya bisa tanpa ataupun dengan berbentuk gelombang sinus. Model AR digunakan jika plot ACFnya *dies down* sementara PACF-nya *cut off*. Model MA digunakan jika plot ACFnya *cut off* dan plot ACF-nya *dies down*. Sedangkan jika kedua plot ACF dan PACF sama-sama *dies down*, maka model tersebut adalah arima Apabila disajikan secara grafik, autokorelasi suatu data yang tidak stasioner menunjukan suatu trend searah diagonal dari kanan ke kiri bersama dengan meningkatnya jumlah *time lag* (selisih waktu). Adanya suatu *trend* (linier atau tidak linier) dalam data berarti bahwa setiap nilai yang berturut-turut akan berkorelasi positif satu sama lainya. Autokorelasi untuk suatu *time lag* r_1, relative akan besar dan positif. Autokorelasi untuk dua time lag juga akan relative besar dan positif, tetapi tidak sebesar r_1, karena komponen kesalahan random telah dimasukan dua kali. Demikian pula, secara umum r_k untuk data yang tidak stasioner akan relative besar dan positif, sampai nilai *k* menjadi cukup besar sehingga komponen kesalahan random mulai mendominasi autokorelasi.

2.1 Tabel Bentuk ACF dan PACF yang stasioner

No	Bentuk	ACF	PACF
1	Autoregresive AR (p)	*Dies down* (turun cepat secara eksponensial/ nusoidal)	*Cuts off after lag p* (terputus setelah lag p)
2	Moving Average MA (q)	*Cuts off after lag q* (terputus setelah lag q)	*Dies down* (turun cepat secara eksponensial / sinusoidal)
3	ARMA (p,q)	*Dies down* (turun cepat secara eksponensial / sinusoidal)	*Dies down* (turun cepat secara eksponensial / sinusoidl))
4	AR(*p*) atau MA(*q*)	*Cuts off after lag q* (terputus setelah lag *q*)	*Cuts off after lag p* (terputus setelah lag *p*)
5	*White noise* (Random)	Tidak ada yang pnifikan (tidak ada yang keluar batas)	Tidak ada yang pnifikan (tidak ada yang keluar batas)

2.2 Tabel Bentuk ACF dan PACF musiman yang stasioner

No	Proses	ACF	PACF
1	AR(P)^S	*Dies down* (turun cepat secara eksponensial / sinusoidal) pada lag kS, dengan $k=1,2,3,\ldots$	*Cuts off* (terputus) setelah lag PS
2	MA(Q)^S	*Cuts off* (terputus) setelah lag QS	*Dies down* (turun cepat secara eksponensial / sinusoidal) pada lag kS, dengan $k=1,2,3,$
3	ARMA(P,Q)^S	*Dies down* (turun cepat secara eksponensial / sinusoidal) pada lag kS, dengan $k=1,2,3,\ldots$	*Dies down* (turun cepat secara eksponensial / sinusoidal) pada lag kS, dengan $k=1,2,3,\ldots$
4	AR(P)^S atau MA(Q)^S	*Cuts off* (terputus) setelah lag QS	*Cuts off* (terputus) setelah lag PS
5	*White noise* (Random)	Tidak ada yang pnifikan (tidak ada yang keluar batas)	Tidak ada yang pnifikan (tidak ada yang keluar batas)

Suhartono (2008:177)

H. Model Autoregressive Integrated Moving Average (ARIMA)

Model – model yang mungkin dihasilkan dari pengidentifikasian data deret waktu dapat berupa model *autoregressive* (AR), *integrated* (I) dan *moving average* (MA) atau kombinasi dari dua komponen model (ARI, IMA, ARMA) atau kombinasi dari tiga komponen model (ARIMA). Dimana model autoregressive merupakan model yang menggambarkan hubungan antara peubah terikat Y pada waktu sebelumnya, sedangkan model moving average merupakan model yang menggambarkan ketergantungan peubah terikat *Y* terhadap nilai-nilai galat pada waktu sebelumnya yang berurutan.

Secara umum bentuk model ARIMA Box-Jenkins atau ARIMA (p, d, q) adalah sebagai berikut :

$$\phi^{p}(\mathrm{B})(1-\mathrm{B})^{d} Z_{t} = \theta^{q}(\mathrm{B})a_{t} \quad (2.10)$$

dimana :
p adalah orde AR
d adalah orde differencing
q adalah orde MA

Model ARIMA (p,d,q) merupakan model yang telah mengalami proses *differencing*. Bentuk umum dari model ARIMA pada orde ke p,q dengan proses *differencing* sebanyak d adalah sebagai berikut :

$$\phi_{p}(B)\,(1-B)^{d} Z_{t} = \theta_{q}(B)\alpha_{t} \quad (2.10a)$$

yang didefernsialkan dengan :

$$\phi_{p}(B) = (1 - \phi_{1}B - \phi_{2}\ B^{2} - \ldots - \phi_{p}B^{p})\ \text{dan} \quad (2.10b)$$

$$\theta_{q}(B) = (1 - \theta_{t}B - \theta_{2}B^{2}\ - \ldots - \theta_{q}B^{q})$$

I. Model SARIMAX (Seasonal Autoregressive Integrated Moving Average with Exogenous Input)

SARIMAX adalah model *Seasonal* ARIMA dengan variabel tambahan, variabel tambahan yang dimaksud adalah variabel *dummy (t)* untuk efek variasi kalender dimana efek variasi tersebut dibagi menjadi dua bagian yaitu variasi perdagangan dan variasi liburan, namun yang digunakan di Indonesia adalah efek lebaran.

1) Model ARIMA Musiman

Secara umum bentuk model ARIMA musiman atau *(p, d, q)(P, D, Q)* ^{S} adalah :

$$\phi_p(B)\Phi_p(B^S)(1-B)^d(1-B^S)^D Z_t = \theta_q(B)\Theta_Q(B^S)\alpha_t \qquad (5.1)$$

dimana :

p, d, q	: orde AR, *differencing*, dan MA non-musiman
P, D, Q	: orde AR, *differencing*, dan MA musiman
S	: jumlah periode musiman
$\phi_p(B)$	: $1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^P$
$\Phi_p(B^S)$	: $1 - \Phi_1 B^S - \Phi_2 B^{2S} - \dots - \phi_p B^{pS}$
$(1-\mathrm{B})^d$	: orde *differencing* non-musiman
$(1-\mathrm{B}^S)^D$	: orde *differencing* non-musiman
$\theta_q(B)$	: $1 - \theta_1 B^S - \theta_2 B^2 - \dots - \theta_q B^q$
$\Theta_Q(B^S)$	: $1 - \Theta_1 B^S - \Theta_2 B^{2S} - \dots - \Theta_Q B^{QS}$
Z_t	: $Z_t - \mu$

2) Variabel *Dummy*

Menurut Makridakis (1995:260) Model yang memasukan variabel kategorik biasa dinamakan model dengan variabel *dummy*. Jika variabel penjelas kategorik mempunyai dua kategori, maka hanya satu variabel *dummy* yang akan dibutuhkan untuk mewakili dua kategori tersebut. Sebuah variabel *dummy* tertentu (X_d) nilainya ditentukan sebagai :

$$X_d = \begin{cases} 1 \text{ jika observasinya dalam kategori 2} \\ 0 \text{ jika observasinya dalam kategori 1} \end{cases}$$

Variabel dummy yang digunakan dalam penelitian ini adalah variabel *dummy* bulan dalam satu tahun ($D_{1,t}$, $D_{2,t}$, … , $D_{11,t}$) dan variabel *dummy* periode liburan lebaran (PL_t). Berikut ini adalah definisi dari masing – masing variabel *dummy* :

$$D_{1,t} = \begin{cases} 1 \text{ jika periode t adalah bulan januari} \\ 0 \text{ lainya} \end{cases}$$

$$D_{2,t} = \begin{cases} 1 \text{ jika periode t adalah bulan februari} \\ 0 \text{ lainya.} \end{cases}$$

$$\vdots$$

$$D_{11,t} = \begin{cases} 1 \text{ jika periode t adalah bulan november} \\ 0 \text{ lainya.} \end{cases}$$

$D_{1,t}$, $D_{2,t}$, … , $D_{11,t}$ = 0, jika periode t adalah bulan desember.

$$PL_t = \begin{cases} 1 \text{ jika periode t terdapat periode lebaran} \\ 0 \text{ lainya.} \end{cases}$$

Setiap variabel *dummy* ekuivalen dengan sebuah regresor baru akan tetapi bila menghadapi masalah multikolinearitas maka menggunakan (P-1), variabel *dummy* untuk menyatakan P periode yang berbeda. Pengunaan variabel dummy untuk menunjukan sifat musiman telah memberikan perbaikan pada model secara pnifikan.

3) Analisis regresi time series untuk variasi kalender

Regresi dalam time series memiliki bentuk yang sama dengan regresi linier umum dengan mengansumsikan output atau bentuk dependen Y_t untuk $t = 1,2,\ldots, n$, yang dipengaruhi oleh kemungkinan data input atau independen, dimana inputnya merupakan fix atau diketahui, hubungan ini dapat ditunjukan dengan model regresi linier.

Model regresi Linier yang digunakan :

- Jika data y_t mempunyai *trend* :

$$y_t = \beta_0 + \beta_1 t + w_t \qquad (5.2)$$

dimana w_t merupakan residual yang mengalami proses independen dan identik serta berdistribusi normal dengan nilai mean 0 dan varian σ_w^2.

- Bentuk data seasonal $S_{1,t}$, $S_{2,t}$, … , $S_{s,t}$:

  $$y_t = \beta_0 + \beta_1 S_{1,t} + \beta_2 S_{2,t} + ... + \beta_s S_{s,t} + w_t \qquad (5.3)$$

  dimana $S_{1,t}$, $S_{2,t}$, …, $S_{s,t}$ merupakan variabel dummy untuk bentuk seasonal.

- Data yang memiliki variasi kalender :

  $$y_t = \beta_0 + \beta_1 V_{1,t} + \beta_2 V_{2,t} + ... + \beta_p V_{p,t} + w_t \qquad (5.4)$$

  dimana $V_{p,t}$ adalah variabel dummy untuk efek variasi kalender ke-p.

Ide dasar dari model variasi kalender adalah :

$y_t = \mu_t + \varepsilon_t$

dimana :

μ_t : komponen deterministik untuk menghitung variasi kalender. mean deterministik (μ_t) dapat diwakili oleh trend, faktor musiman dan efek variasi kalender.

ε_t : proses ARIMA untuk menghitung sisaan y_t

Sehingga model variasi kalender yang terbentuk adalah sebagai berikut :

$$y_t = \alpha_0 + \alpha_1 t + \beta_1 D_{1,t} + ... + \beta_{(L-1)} D_{(L-1),t} + \gamma CV_t + \frac{\theta_q(B)\theta_Q(B^S)}{\phi_p(B)\Phi_p(B^S)}\alpha_t$$

Dimana :

D_1 ,D_2 , … , $D_{(L-1)}$: dummy waktu dalam satu periode musiman.

CV_t :variabel dummy efek variasi kalender, bernilai 1 jika pada observasi ke –t terjadi efek variasi kalender dan bernilai 0 untuk yang lain.

J. Peramalan

Ramalan merupakan dugaan atau pikiran mengenai terjadinya kejadian atau peristiwa dari waktu yang akan datang (Supranto, 1993).

Peramalan dapat juga diartikan sebagai suatu usaha memperkirakan perubahan, agar tidak disalahpahami bahwa peramalan tidak memberikan jawaban yang pasti tentang apa yang akan terjadi, melainkan akan mencari yang sedekat mungkin dengan apa yang akan terjadi. Peramalan adalah suatu teknik untuk memperkirakan suatu nilai pada masa yang akan datang dengan memperhatikan data masa lalu maupun data saat ini (Sukarna, 2006:1).

Hampir setiap organisasi memerlukan ramalan baik secara eksplisit maupun secara implisit, karena hampir setiap organisasi harus membuat perencanaan agar sesuai dengan kondisi masa depan yang tidak diketahui dengan baik. Selain itu, peramalan dibutuhkan pada semua lini fungsional, begitu pula pada semua jenis organisasi. Peramalan dibutuhkan dalam bidang keuangan, pemasaran, personalia, dan lingkup produksi, dalam pemerintahan dan organisasi pencari laba, dalam klub sosial kecil, dan dalam partai politik nasional.

1. Metode Deret Berkala

Runtun waktu adalah serangkaian pengamatan terhadap suatu peristiwa, kejadian, gejala atau variabel yang diambil dari waktu ke waktu, dicatat secara teliti menurut urut-urutan waktu terjadinya dan kemudian disusun sebagai data statistik.

Pada umumnya pencatatan ini dilakukan dalam jangka waktu tertentu misalnya satu bulan, satu tahun, sepuluh tahunan dan sebagainya. Analisis runtun waktu adalah suatu metode kuantitatif untuk menentukan pola data masa lampau yang telah dikumpulkan secara teratur. Secara umum analisis deret waktu mempunyai beberapa tujuan, yaitu peramalan, pemodelan, dan kontrol.

Peramalan berkaitan dengan pembentukan model dan metode yang dapat digunakan untuk menghasilkan suatu ramalan yang akurat. Sebuah model deret berkala adalah suatu fungsi yang menghubungkan nilai deret berkala dengan nilai awal deret berkala, kesalahannya, atau yang berhubungan dengan deret berkala lainnya. Metode deret berkala meramalkan sifatnya untuk masa depan. Jika ada persamaan yang ditentukan maka sifat sistem dapat diketahui dengan menyelesaikan persamaan tersebut yang kondisi awalnya sudah diketahui. Pada peramalan runtun waktu, persamaan dan kondisi awal mungkin diketahui kedua-duanya atau mungkin hanya salah satu saja. Oleh karena itu diperlukan suatu aturan untuk menentukan perkembangan dan keakuratan sistem. Penentuan aturan tersebut mungkin mengacu dari pencocokkan data masa lalu.

Pada saat t_1, t_2, t_3 ,… t_n pengamatan suatu deret berkala membentuk suatu deret dan mempunyai variabel random Z_{t1} , Z_{t2} , Z_{t3} ,… Z_{tn} dengan fungsi distribusi bersama F(Z_{t1}, Z_{t2}, , Z_{t3} ... Z_{tn}) Dalam analisis *time series*, data pengamatan yang disimbolkan dengan Z_t disyaratkan mengikuti proses stokastik

Proses stokastik adalah suatu kelompok data berdasarkan waktu yang tersusun oleh variabel *Z (ω , t)* random dimana ω adalah ruang sampel dan *t* adalah indeks waktu

Faktor utama yang mempengaruhi pemilihan teknik peramalan untuk data deret waktu adalah identifikasi dan pemahaman pola historis data. Pola data tersebut terbagi menjadi empat, yaitu:

1) Pola Horisontal

   Pola ini terjadi pada saat data observasi berfluktuasi disekitar nilai rata-rata konstan. Pola ini disebut juga pola stasioner.

2) Pola *Trend*

   Pola ini muncul ketika observasi data menaik atau menurun pada periode yang panjang. Contoh dari rangkaian *trend* adalah pertumbuhan populasi, inflasi harga, perubahan teknologi, preferensi konsumen dan kenaikan produktifitas.

3) Pola Siklis (*cyclus*)

Pola ini muncul pada saat observasi data memperlihatkan kenaikan dan penurunan pada periode yang tidak tetap. Komponen siklik mirip fluktuasi gelombang disekitar *trend* yang sering dipengaruhi oleh kondisi ekonomi.

4) Pola Musiman (*seasonality*)

Pola terjadi pada saat data observasi dipengaruhi oleh faktor musiman. Komponen musiman mengacu pada suatu pola perubahan yang berulang dengan sendirinya dari tahun ketahun. Untuk deret bulanan, komponen musiman mengukur keragaman deret dari setiap Januari, setiap Februari dan seterusnya. Untuk deret triwulanan, ada empat elemen musim, masing-masing satu untuk setiap triwulan.

2. Deret waktu musiman

Musiman berarti kecenderungan mengulangi pola tingkah gerak dalam periode musim, biasanya satu tahun untuk data bulanan. Karena itu, deret waktu musiman mempunyai karakteristik yang ditunjukan oleh adanya korelasi beruntun yang kuat pada jarak semusim.

Deret waktu musiman adalah serangkaian pengamatan terhadap suatu peristiwa atau variabel yang diambil dari waktu ke waktu, dicatat secara teliti menurut urutan – urutan waktu terjadi dan disusun sebagai statistik. Dari suatu rangkaian waktu akan dapat diketahui apakah peristiwa yang diamati itu berkembang mengikuti pola – pola perkembangan yang teratur, maka dapat dibuat ramalan yang cukup kuat mengenai perilaku gejala yang dicatat dan atas ramalan inilah di dapat rencana–rencana yang dapat dipertanggung jawabkan. Peramalan deret waktu sangat dibutuhkan dalam pengambilan keputusan dalam masa – masa yang akan datang.

K. Langkah analisis *Seasonal* ARIMAX

a) Identifikasi

1. Plot Data

Plot data digunakan untuk mengetahui pola data dan *trend* (penyimpangan nilai tengah) pengamatan serta menghilangkan pengaruh musim pada data, dengan melihat grafik hasil plot data, kita dapat melihat gambaran bahwa data tersebut stasioner (tidak terdapat pertumbuhan atau penurunan data) atau tidak stasioner.

2. Membuat plot ACF dan PCAF.

3. Pembentukan variabel dummy, Variabel dummy yang digunakan adalah variabel *dummy* bulan dalam satu tahun ($D_{1,t}$, $D_{2,t}$, … , $D_{11,t}$) dan variabel *dummy* periode liburan lebaran (PL_t). Berikut ini adalah definisi dari masing – masing variabel *dummy* :

$$D_{1,t} = \begin{cases} 1 \text{ jika periode t adalah bulan januari} \\ 0 \text{ lainya} \end{cases}$$

$$D_{2,t} = \begin{cases} 1 & \text{jika periode t adalah bulan februari} \\ 0 & \text{lainya.} \end{cases}$$

$\vdots$

$$D_{11,t} = \begin{cases} 1 & \text{jika periode t adalah bulan november} \\ 0 & \text{lainya.} \end{cases}$$

$D_{1,t}$, $D_{2,t}$, … , $D_{11,t}$ = 0, jika periode t adalah bulan desember.

$$PL_t = \begin{cases} 1 & \text{jika periode t terdapat periode lebaran} \\ 0 & \text{lainya.} \end{cases}$$

4. Melakukan pemodelan regresi dummy, dengan variabel respon jumlah penumpang kereta api tiap bulan, dan variabel prediktornya adalah waktu (t), dummy 12 bulan ($D_{1,t}$, $D_{2,t}$, … , $D_{11,t}$) dan dummy periode lebaran (PL_t). Dengan pemodelan regresi, maka dapat mengetahui variabel bulan yang berpengaruh terhadap jumlah penumpang.

. 5. Melakukan uji *white noise* dengan uji *L-jung box*, jika residual tidak *white noise* maka dilanjutkan pada tahap selanjutnya, yaitu memodelkan residual dengan model SARIMA.

6. Melakukan pemodelan terhadap residual dengan melihat plot ACF dan PCAF residual, sehingga di dapatkan dugaan sementara.

b) Penaksiran Parameter

Setelah model sementara untuk suatu runtun waktu diidentifikasikan, langkah selanjutnya adalah mencari estimasi terbaik untuk parameter-parameter dalam model ementara tersebut

Terdapat dua cara yang mendasar untuk mendapatkan parameter-parameter tersebut:

1) Dengan cara mencoba-coba (*trial and error*), menguji beberapa nilai yang berbeda dan memilih satuu nilai tersebut (atau sekumpulaan nilai, apabila terdapat lebih dari satu parameter yang akan ditaksir) yang meminimumkan jumlah kuadrat nilai sisa (*sum of squared residuals*).
2) Perbaikan secara iteratif, memilih taksiran awal dan kemudian membiarkan program komputer memperhalus penaksiran tersebut secara iteratif.

c) Pemeriksaan Diagnostik

Setelah menaksir nilai-nilai parameter dari model *Seasonal* ARIMA yang ditetapkan sementara, selanjutnya perlu dilakukan pemeriksaan untuk membuktikan bahwa model tersebut cukup memadai. Diagnosis model dilakukan untuk mendeteksi adanya korelasi dan kenormalan antar *residual*. Dalam runtun waktu (*time series*) ada asumsi bahwa *residual* mengikuti proses *white noise* yang berarti *residual* harus independen (tidak berkorelasi) dan berdistribusi normal dengan rata-rata mendekati 0 ($\mu = 0$) dan standar deviasi (σ) tertentu.

Untuk mendeteksi adanya proses *white noise*, maka perlu dilakukan diagnosis model. Ada beberapa cara dalam hal ini, antara lain:

1. Uji independensi *residual (white noise)*

   Uji dilakukan untuk mendeteksi independensi *residual* antar lag. Dua lag dikatakan independen (tidak berkorelasi) apabila antar lag tidak ada korelasi cukup berarti.

   Dalam penelitian ini, Uji *white noise* menggunakan uji *L-jung box* dengan hipotesis sebagai berikut :

   Dengan membandingkan χ^2 *Ljung−Box* dan χ^2 (α ,*df*) pada output proses *Ljung-Box-Pierce*.

   Hipotesis:

   H0 : 0 , = *at*, *at* + *k* ρ (Tidak ada korelasi antar lag atau minimal ada 1 lag yang 0 , ≠ *at*, *at* + *k* ρ , memenuhi syarat *white noise*)

$H1$: 0 , $\neq$ *at, at + k* ρ (Ada korelasi antar-lag yang artinya tidak memenuhi syarat *white noise*)

Kriteria penolakan *H0* yaitu jika χ^2 *Ljung−Box* > χ^2 (α,*df*) , di mana distribusi χ^2 yang digunakan mempunyai $df = K\text{-}k$.

Selain dengan pengujian hipotesis, *independensi* antar lag akan ditunjukkan pula oleh grafik fungsi autokorelasi (ACF) rasidual. Suatu *residual* model dikatakan telah independen jika tidak ada satu lag pun pada grafik fungsi autokorelasi (ACF) residual yang keluar batas garis .

2. Uji kenormalan *residual*

Uji dilakukan untuk mendeteksi kenormalan *residual* model. Dalam penelitian ini, uji dilakukan hanya dengan membandingkan nilai *P-Value* pada output proses *Ljung-Box-Pierce* dengan level toleransi (α) yang digunakan dalam uji hipotesis.

Hipotesis:

H0: *Residual* model berdistribusi normal dengan rata-rata mendekati 0 (μ= 0)

H1: *Residual* model tidak berdistribusi normal dengan rata-rata mendekati 0 (μ= 0)

Kriteria penolakan H0 yaitu jika *P-Value* < level toleransi (α).

d) Peramalan

.Setelah metode peramalan ditetapkan, maka model *Seasonal* ARIMAX atau SARIMAX dapat diterapkan pada data, dan dapat dilakukan perkiraan pada data tersebut untuk beberapa periode ke depan. Dengan berjalannya waktu perlu dilakukan evaluasi ulang terhadap model yang sudah dipilih karena ada kemungkinan model perlu diperbaiki karena pola data mungkin berubah.

Ada beberapa metode yang digunakan untuk menunjukkan kesalahan yang disebabkan oleh suatu teknik peramalan tertentu. Hampir semua ukuran tersebut menggunakan beberapa fungsi dari perbedaan antara nilai sebenarnya dengan nilai peramalannya. Perbedaan nilai sebenarnya dengan nilai peramalan ini biasanya disebut sebagai residual.