

# **LAPORAN**

## **PENGEMBANGAN MODEL DERET WAKTU DENGAN TUJUAN MEMPREDIKSI (*FORECASTING*) HASIL PENJUALAN BULANAN UNTUK TOKO SUVENIR DI KOTA RESOR PANTAI DI QUEENSLAND, AUSTRALIA**

Ditulis untuk memenuhi tugas mata kuliah Analisis Deret Waktu

Oleh:

Kanaya Tabhita Djie (01112180013)



**PROGRAM STUDI MATEMATIKA  
FAKULTAS SAINS DAN TEKNOLOGI  
UNIVERSITAS PELITA HARAPAN  
TANGERANG  
2022**

# **BAB I**

## **PENDAHULUAN**

### **Deskripsi Masalah**

Di masa terkini di mana jumlah informasi sudah tidak terhitung banyaknya dan bahkan terus bertambah secara cepat seiring berjalannya waktu, tidak memungkinkan jika harus membaca data yang jumlahnya bisa ribuan bahkan jutaan secara mentah-mentah, dan kemudian langsung menarik sebuah kesimpulan dari data tersebut. Oleh karena itu, kita dapat memanfaatkan deret waktu untuk mengobservasi data tersebut secara statistik maupun visual, sehingga informasi-informasi penting dari data tersebut dapat diekstrak dengan baik dan dimanfaatkan dengan baik.

Pada kasus ini, sebuah toko souvenir di resor pantai, Queensland, Australia memiliki sekian banyak data dan akan dianalisis dengan deret waktu. Berdasarkan karakteristik data tersebut, akan dibuat model yang sesuai dan model tersebut akan dimanfaatkan untuk *forecasting* atau memprediksi hasil penjualan dua tahun ke depannya.

### **Deskripsi Data**

Data yang digunakan dalam pengembangan model deret waktu adalah data penjualan bulanan untuk toko souvenir di kota resor pantai, Queensland, Australia dari Januari 1987 hingga Desember 1993. Sumber data diperoleh dari <https://robjhyndman.com/tsdldata/data/fancy.dat>.

## BAB II

### PENGEMBANGAN MODEL

#### *Step 1: Import the Data*

Pertama, data akan diimpor dari <https://robjhyndman.com/tsdldata/data/fancy.dat> dan kemudian disimpan di variabel `sales`. Kelas dari data `sales` masih berupa numerik sehingga akan diubah menjadi kelas deret waktu atau `ts` dan karena data hasil penjualan dimulai dari tahun 1987, atur *argument* `start` menjadi dari tahun 1987 dengan *frequency* sebesar 12 karena berupa data perbulan. Data deret waktu ini kemudian akan disimpan di variabel `sales.ts`.

#### *Step 2: Data Exploration*

Berikut adalah tampilan data hasil penjualan toko souvenir yang akan digunakan pada penelitian ini. Data hasil penjualan berupa data penjualan per bulan dari tahun 1987-1993.

**Gambar 2.1: Data Hasil Penjualan Toko Souvenir**

```
> sales.ts
      Jan      Feb      Mar      Apr      May      Jun
1987 1664.81 2397.53 2840.71 3547.29 3752.96 3714.74
1988 2499.81 5198.24 7225.14 4806.03 5900.88 4951.34
1989 4717.02 5702.63 9957.58 5304.78 6492.43 6630.80
1990 5921.10 5814.58 12421.25 6369.77 7609.12 7224.75
1991 4826.64 6470.23 9638.77 8821.17 8722.37 10209.48
1992 7615.03 9849.69 14558.40 11587.33 9332.56 13082.09
1993 10243.24 11266.88 21826.84 17357.33 15997.79 18601.53
      Jul      Aug      Sep      Oct      Nov      Dec
1987 4349.61 3566.34 5021.82 6423.48 7600.60 19756.21
1988 6179.12 4752.15 5496.43 5835.10 12600.08 28541.72
1989 7349.62 8176.62 8573.17 9690.50 15151.84 34061.01
1990 8121.22 7979.25 8093.06 8476.70 17914.66 30114.41
1991 11276.55 12552.22 11637.39 13606.89 21822.11 45060.69
1992 16732.78 19888.61 23933.38 25391.35 36024.80 80721.71
1993 26155.15 28586.52 30505.41 30821.33 46634.38 104660.67
```

Total ada 84 data hasil penjualan souvenir dengan nilai maksimum \$104,661 pada bulan Desember 1993 dan nilai minimum \$1,665 pada bulan Januari 1987. Rata-rata hasil penjualan tiap bulannya adalah \$14,316.

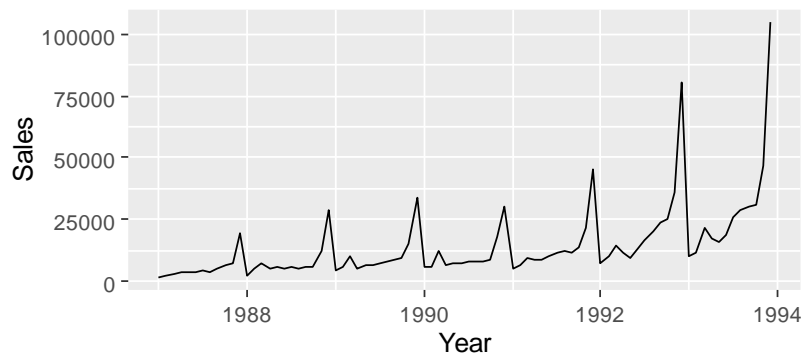
**Gambar 2.2: Summary Data Hasil Penjualan Toko Souvenir**

```
> summary(sales)
   Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
   1665   5884   8772  14316  16889 104661
```

### Step 3: Trend Analysis

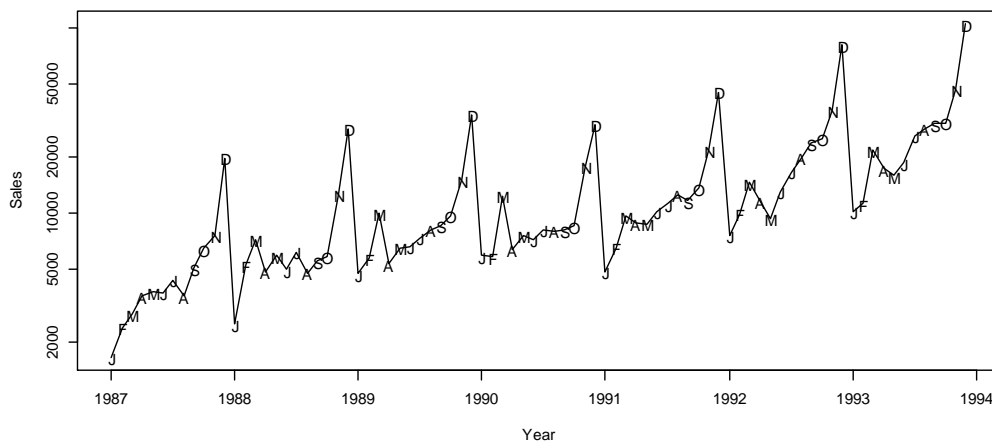
Langkah selanjutnya ketika data deret waktu telah diperoleh adalah membuat *plot* deret waktu. Data observasi hasil penjualan toko souvenir akan di-*plot* terhadap waktu pengamatan secara berurutan dari Januari 1987 hingga Desember 1993. Data perbulannya akan dihubungkan menggunakan garis lurus.

**Gambar 2.3: Plot Deret Waktu Hasil Penjualan Toko Souvenir Tahun 1987-1993**



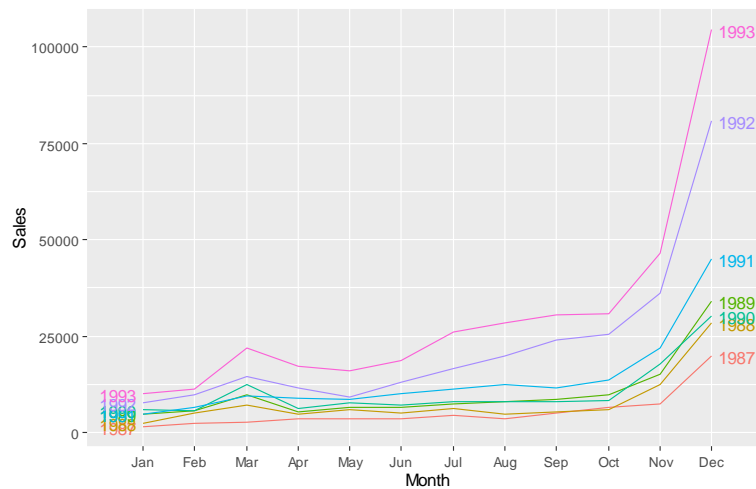
Dari Gambar 2.3 terlihat hasil penjualan cenderung sangat tinggi di akhir tahun sehingga ada kemungkinan data ini memiliki pola musiman atau *seasonal*. Untuk memastikan dugaan ini, akan ditambahkan *plotting symbols* sesuai dengan bulannya. Maka semua data bulan Januari akan ditandai dengan huruf J, data bulan Februari dengan F, data bulan Maret dengan M, dan seterusnya.

**Gambar 2.4: Plot Deret Waktu Hasil Penjualan Toko Souvenir Tahun 1987-1993 dengan Plotting Symbols**



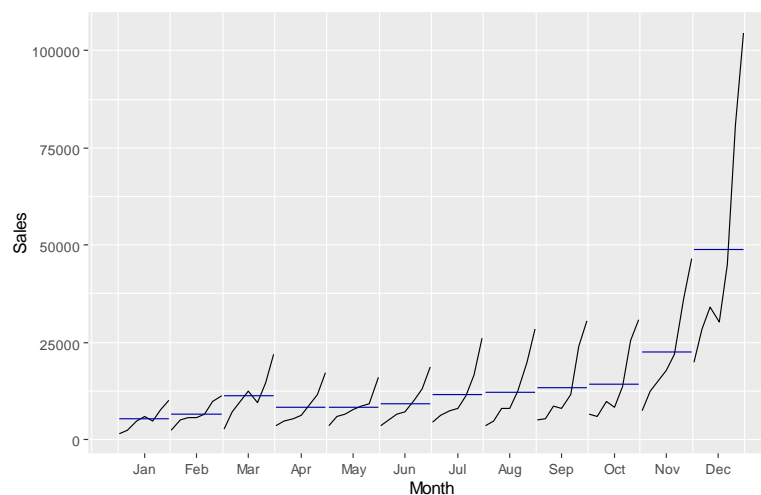
Dari Gambar 2.4 terlihat hasil penjualan toko souvenir setiap tahunnya akan mencapai puncak tertinggi di bulan Desember dan akan jatuh kembali di bulan Januari tahun depannya. Maka dapat dipastikan data bersifat musiman. Pada Gambar 2.5 digambarkan deret waktu hasil penjualan yang dibagi berdasarkan tahunnya, sehingga akan terlihat lebih jelas bahwa setiap tahunnya, hasil penjualan akan naik pada bulan Desember dan turun di bulan Januari.

**Gambar 2.5: Plot Deret Waktu Hasil Penjualan Toko Souvenir per Tahun**



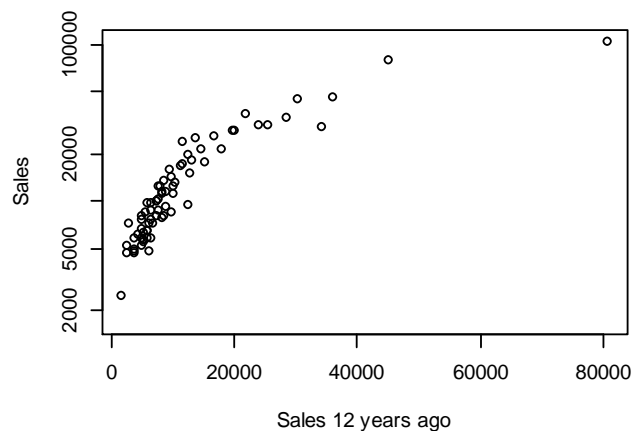
Pada Gambar 2.6 ditunjukkan hasil penjualan keseluruhan tahun 1987-1993 dan dibagi berdasarkan bulannya. Garis horizontal menunjukkan rata-rata penjualan tiap bulan dan terlihat rata-rata penjualan di bulan Desember jauh lebih tinggi dibandingkan bulan-bulan lainnya.

**Gambar 2.6: Plot Deret Waktu Total Hasil Penjualan Toko Souvenir per Bulan**



Selanjutnya akan ditunjukkan korelasi data lag waktu sebesar 12 menggunakan *scatterplot*. Pada Gambar 2.7 terlihat ada tren positif sehingga ada korelasi positif antara hasil penjualan bulan Desember dengan hasil penjualan bulan Desember sebelumnya.

**Gambar 2.7: Scatterplot Hasil Penjualan Bulan Desember dengan Hasil Penjualan Bulan Desember Sebelumnya**



#### ***Step 4: Model Specification***

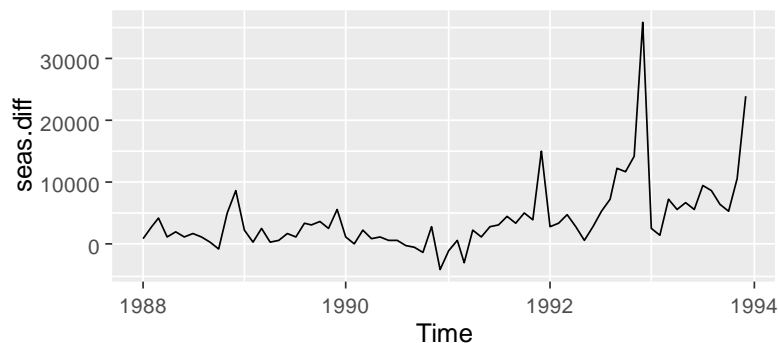
Berdasarkan analisis trend sebelumnya, diketahui data deret waktu hasil penjualan toko souvenir memiliki tren musiman yang terulang setiap 12 bulan. Oleh karena itu, akan digunakan model *seasonal* ARIMA untuk kasus ini, yaitu:

$$ARIMA(p, d, q)(P, D, Q)_s$$

Di mana  $p$ ,  $d$ , dan  $q$  adalah bagian *non-seasonal* dan  $P$ ,  $D$ , dan  $Q$  adalah bagian *seasonal* dari model ARIMA, sedangkan  $s$  menunjukkan berapa lama periode pola tersebut terjadi sebelum terulang kembali.

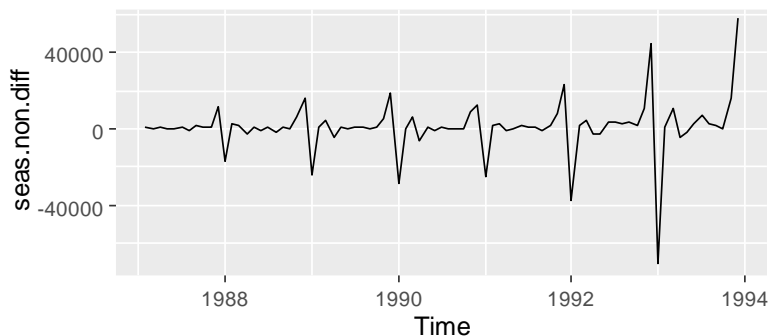
Pola musiman ini menyebabkan data menjadi tidak stasioner, sehingga perlu dilakukan *seasonal differentiation* terlebih dahulu. *Plot* deret waktu untuk data yang sudah didiferensiasi secara musiman ditunjukkan di Gambar 2.8.

**Gambar 2.8: *Plot Deret Waktu Data (Seasonal Differentiation)***



Terlihat data masih belum stasioner sehingga akan dilakukan diferensiasi sekali lagi dan hasilnya ditunjukkan pada Gambar 2.9. Terlihat data sudah stasioner.

**Gambar 2.9: *Plot Deret Waktu (First Differentiation)***



Untuk membuktikan data sudah stasioner, dapat dilakukan uji *Augmented Dickey-Fuller* (ADF). Dari Gambar 2.10, didapatkan  $p - value = 0.02655$  di mana nilai tersebut lebih kecil secara signifikan dari  $\alpha = 0.05$ . Oleh karena itu, hipotesis nol dapat ditolak dan terbukti data tersebut stasioner.

**Gambar 2.10: *Uji Augmented Dickey-Fuller (ADF)***

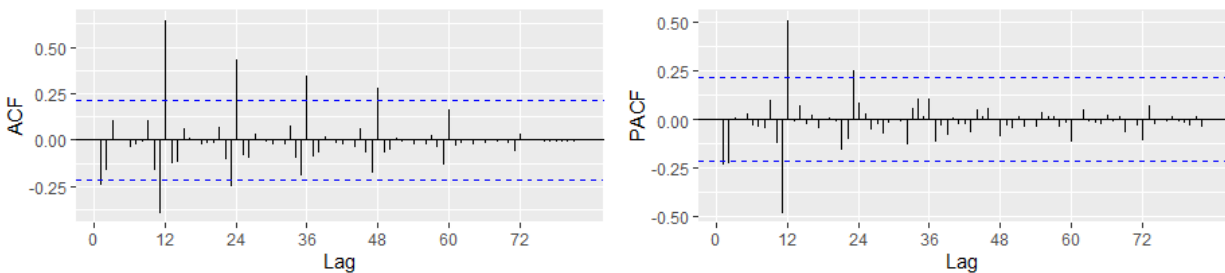
```
> adf.test(seas.non.diff)

Augmented Dickey-Fuller Test

data: seas.non.diff
Dickey-Fuller = -3.7374, Lag order = 4, p-value = 0.02655
alternative hypothesis: stationary
```

Setelah mendapatkan data yang stasioner, akan ditentukan model ARIMA yang tepat berdasarkan ACF dan PACF pada Gambar 2.11. Pertama akan ditentukan bagian *seasonal* dari model terlebih dahulu, yaitu dengan mengobservasi pola ACF dan PACF pada lag 12, 24, 36, dan seterusnya. Perhatikan pola nilai ACF adalah *tails-off* dan PACF *cut-off* setelah lag ke-12, sehingga diduga bagian *seasonal* mengikuti model AR(1). Sekarang akan ditentukan bagian *non-seasonal* dari model ARIMA. Perhatikan pola nilai ACF adalah *tails-off* dan PACF *cut-off* setelah lag ke-2, sehingga diduga bagian *non-seasonal* mengikuti model AR(2). Berdasarkan hasil observasi di atas, dapat diduga bahwa model tersebut adalah  $ARIMA(2, 1, 0)(1, 1, 0)_{12}$ .

**Gambar 2.11: ACF dan PACF untuk Spesifikasi Model**



#### **Step 5: Parameter Estimation**

Berdasarkan langkah sebelumnya, model yang digunakan untuk memprediksi hasil penjualan toko suvenir adalah  $ARIMA(2, 1, 0)(1, 1, 0)_{12}$ . Sekarang diperlukan estimasi parameter untuk *non-seasonal*  $\hat{\Phi}_1, \hat{\Phi}_2$  dan *seasonal*  $\hat{\Phi}_1$ . Estimasi parameter akan dilakukan dengan metode *maximum likelihood*. Hasil estimasi ditunjukkan pada Gambar 2.12.

**Gambar 2.12: Estimasi Parameter Model Pertama**

```
Call:
arima(x = seas.non.diff, order = c(2, 1, 0), seasonal = c(1, 1, 0), method = "ML")

Coefficients:
      ar1      ar2      sar1
    -0.9431  -0.5567   0.6275
s.e.    0.1000   0.0966   0.1213

sigma^2 estimated as 31246801:  log likelihood = -706.92,  aic = 1419.85
```

Dari hasil estimasi tersebut, didapatkan nilai untuk bagian *non-seasonal* yaitu  $\hat{\Phi}_1 = -0.9431$  dan  $\hat{\Phi}_2 = -0.5567$ . Perhatikan *standard error* (s.e.), estimasi untuk koefisien lag 1 AR



(ar1) signifikan terhadap nol, sedangkan koefisien lag 2 AR (ar2) tidak signifikan terhadap nol. Untuk bagian *seasonal*, didapatkan nilai  $\hat{\Phi}_1 = 0.6275$ . Nilai koefisien lag 1 AR (sar1) sudah signifikan terhadap nol. Nilai AIC untuk model ini adalah 1,419.85. Karena nilai ar2 tidak signifikan, kita bisa menghilangkan koefisien tersebut dengan mengatur nilai  $\hat{\Phi}_2$  menjadi nol. Setelah itu, lakukan estimasi parameter kembali untuk mendapatkan nilai ar1 dan sar1 yang baru. Hasil estimasi ditunjukkan di Gambar 2.13.

**Gambar 2.13: Estimasi Parameter Model Kedua**

```
Call:
arima(x = seas.non.diff, order = c(2, 1, 0), seasonal = c(1, 1, 0), transform.pars = FALSE,
      fixed = c(NA, 0, NA), method = "ML")

Coefficients:
          ar1  ar2      sar1
      -0.5957    0  0.5739
s.e.    0.0942    0  0.1313

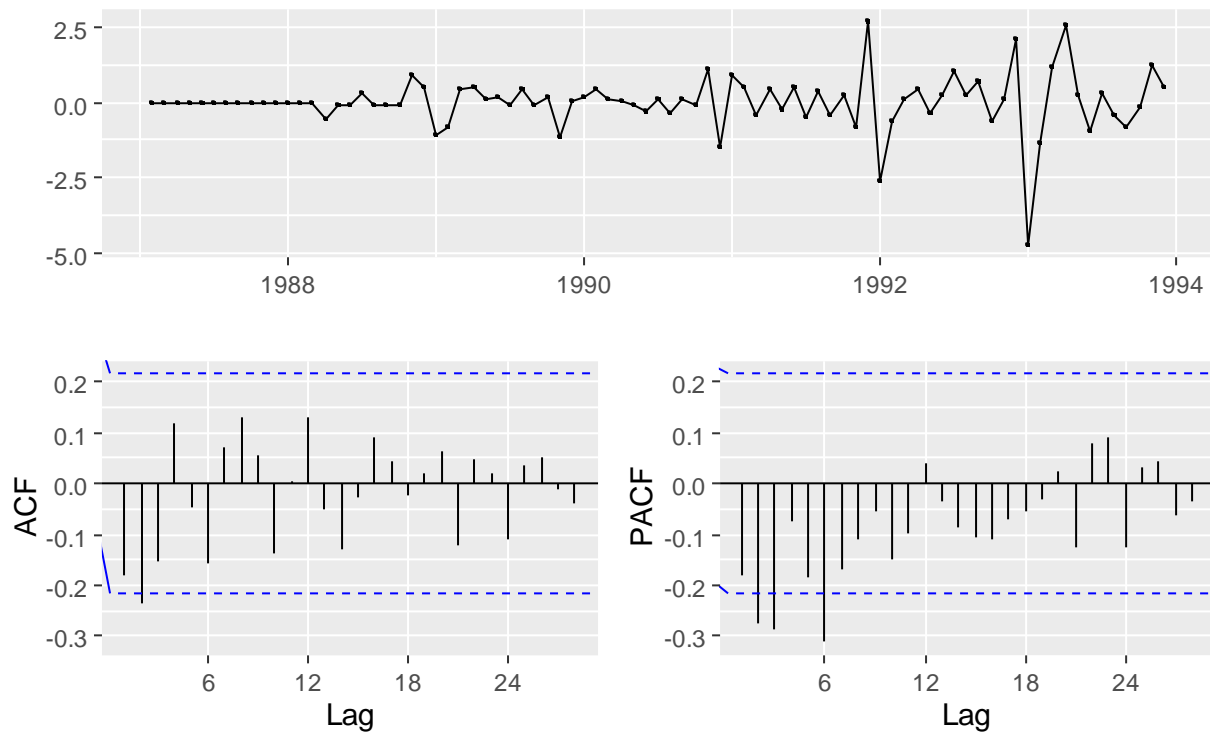
sigma^2 estimated as 46888031:  log likelihood = -720.16,  aic = 1444.32
```

Dari hasil estimasi model kedua, didapatkan nilai ar1 sebesar  $-0.5957$  dan sar1 sebesar  $0.5793$  yang signifikan terhadap nol, namun nilai AIC lebih besar yaitu  $1,444.32$ . Model terbaik adalah model dengan nilai AIC terkecil, sehingga akan dipilih model pertama.

### **Step 6: Model Diagnostics**

Setelah mengestimasi parameter dari model, perlu dilakukan *model diagnostics* untuk menguji kemampuan model untuk melakukan *fitting* terhadap data. Pertama akan dilakukan analisis residual dari model yang sudah ditentukan. *Plot* analisis residual yang sudah distandarisasi pada Gambar 2.14 menunjukkan nilai residual mencapai hampir  $-5.0$  di awal tahun 1993. Nilai residu yang tiba-tiba besar ini tidak normal dan sebaiknya harus dicek kembali apakah ada faktor luar yang mempengaruhi hal tersebut.

**Gambar 2.14: Analisis Residu**



Berdasarkan *plotting* nilai ACF pada Gambar 2.14, pada lag kedua nilainya melebihi batas standar error sehingga ini menandakan adanya autokorelasi antar residual di model ini. Residual seharusnya tidak saling berkorelasi dan untuk memastikan dugaan ini, dapat dilakukan uji Ljung-Box. Hasil uji ini ditampilkan pada Gambar 2.15. Nilai p-value yang didapatkan adalah sebesar 0.07026 yang lebih besar dari  $\alpha = 0.05$ , sehingga gagal tolak hipotesis nol dan dapat disimpulkan bahwa tidak ada korelasi antar residual di model ini.

**Gambar 2.15: Uji Ljung-Box**

```
> checkresiduals(mod1)

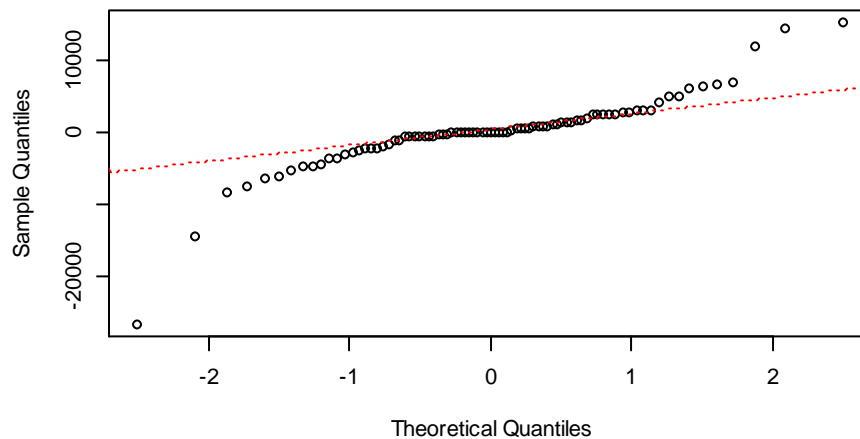
Ljung-Box test

data:  Residuals from ARIMA(2,1,0)(1,1,0)[12]
Q* = 22.427, df = 14, p-value = 0.07026

Model df: 3.    Total lags used: 17
```

Residual juga harus terdistribusi secara normal, maka akan dibuat *Quantile-Quantile plot* pada Gambar 2.16. Terlihat hampir seluruhnya titik-titik observasinya mengikuti garis normal, namun masih ditemukan beberapa titik ektrim yang sangat jauh dari garis normal. Titik-titik tersebut adalah pencilan.

**Gambar 2.16: *Quantile-Quantile Plot***



Untuk membuktikan bahwa residual sudah terdistribusi secara normal, akan dilakukan uji Shapiro-Wilk. Hasil uji ditunjukkan pada Gambar 2.17 dan terlihat nilai p-valuenya cukup signifikan sehingga tolak hipotesis nol. Oleh karena itu, dapat disimpulkan residual model ini tidak terdistribusi secara normal.

**Gambar 2.17: Uji Shapiro-Wilk**

```
> shapiro.test(residuals(mod1))  
  
      shapiro-wilk normality test  
  
data:  residuals(mod1)  
W = 0.81978, p-value = 0.00000001142
```

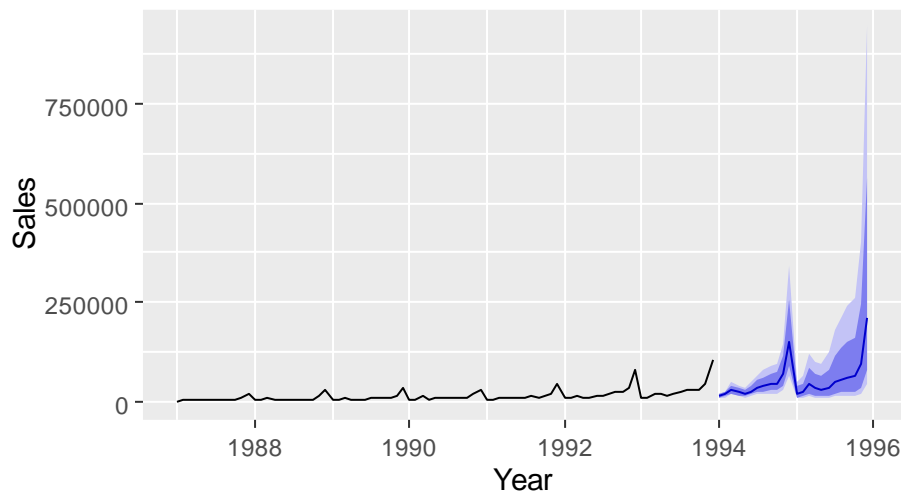
### BAB III

#### PEMBAHASAN

##### Step 7: Forecasting

Setelah melakukan pengembangan model, maka sekarang akan dilakukan *forecasting* menggunakan model tersebut. Hasil *forecasting* untuk dua tahun selanjutnya (1994-1995) ditunjukkan pada gambar 3.1.

**Gambar 3.1: Hasil *Forecasting* untuk Tahun 1994-1995**



Dari Gambar 3.1 terlihat hasil *forecasting* masih mengikuti trend *seasonal* seperti data yang digunakan untuk mengembangkan model. Puncak tertinggi hasil penjualan toko suvenir tetap tercapai di bulan Desember dan akan turun lagi di bulan depannya. Hasil penjualan toko suvenir juga diprediksi akan meningkat dibanding dengan enam tahun sebelumnya. Penjualan diperkirakan akan mencapai sekitar 150,000 pada bulan Desember 1994 dan sekitar 200,000 pada bulan Desember 1995.

## BAB IV

### KESIMPULAN

Model yang digunakan untuk melakukan *forecasting* hasil penjualan dua tahun ke depannya ada model  $ARIMA(2, 1, 0)(1, 1, 0)_{12}$  dengan estimasi parameter bagian *non-seasonal*  $\hat{\Phi}_1 = -0.9431$  dan  $\hat{\Phi}_2 = -0.5567$ , sedangkan untuk bagian *seasonal*, didapatkan nilai  $\hat{\Phi}_1 = 0.6275$ . Dari hasil *model diagnostic*, tidak ditemukan korelasi antar residual, namun masih ditemukan residual yang menyimpang cukup besar dan residual juga tidak terdistribusi secara normal. Oleh karena itu, model dinilai masih kurang baik karena tidak memenuhi seluruh asumsi residual, sehingga tidak menutup kemungkinan model ini dapat dikembangkan lebih lanjut untuk mendapatkan model yang lebih tepat.

Hasil penjualan dua tahun ke depan juga diprediksi masih mengikuti tren *seasonal* dan akan mencapai puncak tertinggi tiap akhir tahun di bulan Desember. Hasil penjualan akhir taun diperkirakan akan terus meningkat. Hasil penjualan toko souvenir yang tinggi setiap akhir tahun bisa diakibatkan karena bulan Desember adalah masa liburan dan cenderung akan banyak pendatang yang mengunjungi kota resor pantai untuk liburan dan membeli souvenir di toko tersebut.