

IMPLEMENTASI SUBSET AUTOREGRESSIVE MENGGUNAKAN PAKET FITAR

Tomí Ardi¹, Rukun Santoso², Alan Prahutama³

^{1,2,3} Departemen Statistika FSM Universitas Diponegoro
rukunsantoso25@gmail.com

ABSTRACT

Time series data analysis is one of the important points in statistics that is a time-dependent analysis. The commonly used model for time series data is ARIMA (Autoregressive Integrated Moving Average) or often also called the Box-Jenkins time series method. A model of ARIMA used in time clock data forecasting is the AR subset (autoregressive). The AR subset model is suitable for a long time series with a more than 5th order lag. The statistical software used is the R. time series AR subset approach on R using the FitAR package. The main function of the FitAR package is `SelectModel` and `FitAR`. `SelectModel` function to get the model automatically while `FitAR` is used to determine the temporary suspect model. Data used in the form of dataset contained in package FitAR that is `SeriesA`. The `SeriesA` data is data about the chemical concentration process observed every 2 hours for 17 days. `SeriesA` is processed using FitAR package so that the best model is AR [1,2,7].

Keywords : Time Series, Time Series Non-stasioner, Subset AR, FitAR Package

1. PENDAHULUAN

Analisis data runtun waktu (*time series*) merupakan salah satu bahasan penting dalam ilmu statistika. Data runtun waktu merupakan jenis data yang dikumpulkan menurut urutan waktu dalam suatu rentang waktu tertentu. Berdasarkan penjelasan sebelumnya bisa dikatakan bahwa data yang memiliki urutan dan rentang waktu tertentu. Model yang umum digunakan untuk data runtun waktu adalah ARIMA (*Autoregressive Integrated Moving Average*) atau sering juga disebut metode runtun waktu Box-Jenkins.

Salah satu model dari ARIMA yang digunakan dalam peramalan data runtun waktu adalah subset AR (*autoregressive*). Berbeda dengan model AR biasanya, model ini merupakan alternatif sederhana dari mdoel ARMA yang kompleks dimana secara khusus model subset AR cocok untuk deret waktu yang panjang dengan lag orde lebih dari 5. Perbedaan model AR dengan subset AR adalah lag yang muncul pada proses analisisnya. Pada umumnya untuk menentukan model AR dengan melihat lag yang muncul pada plot PACF. Apabila lag muncul secara beurutan dan berhenti di lag 3 maka bisa dikatakan model tersebut model AR(3), namun bila lag yang muncul secara acak misalkan lag 1,3,5 maka bisa menggunakan model subset AR untuk melakukan analisisnya.

Dalam perkembangannya banyak *software-software* yang digunakan untuk menganalisis metode statistik dan satu di antaranya menggunakan *software R*. R merupakan salah satu *software* untuk komputasi statistik dan grafik. Sebagian besar fungsinya berasal dari paket yang ada pada *software* atau tersedia pada situs web R. Paket adalah suatu kumpulan perintah atau fungsi yang digunakan untuk melakukan analisis tertentu.

Ada beberapa paket yang digunakan untuk analisis runtun waktu, salah satu paket yang tersedia adalah FitAR. Di dalam paket ini tersedia analisis runtun waktu khususnya untuk model AR dan juga subset AR dengan dua fungsi utama yaitu `SelectModel` dan `FitAR` untuk pemilihan model secara otomatis dan model yang cocok setelah dilakukan diagnosa.

2. TINJAUAN PUSTAKA

2.1 FitAR

Perbedaan R dengan program statistika lain seperti SAS *System* atau Minitab, yaitu R merupakan bahasa pemrograman interpreter yang artinya akan mengeksekusi perintah setiap kali diketikkan di layar. Kelebihan dari R adalah menyediakan penanganan dan penyimpanan data, mendukung banyak operator perhitungan, menyediakan banyak alat untuk analisis data, dan kemampuan grafik yang baik (www.r-project.org).

Paket FitAR menyediakan pendekatan untuk subset time series AR (*autoregressive*). Dua fungsi utama dalam paket ini adalah `SelectModel` dan `FitAR` untuk pemilihan model secara otomatis dan estimasi model, paket ini menyediakan analisis untuk model AR biasa dan Subset AR.

Pada paket FitAR, fungsi yang digunakan yaitu:

1. Tahap seleksi model
 - `TimeSeriesPlot` : digunakan untuk menampilkan plot Time Series
 - `acf` : digunakan untuk menampilkan plot ACF
 - `pacf` : digunakan untuk menampilkan plot PACF
 - `SelectModel` : digunakan untuk menduga model awal
 - `BoxCox` : digunakan untuk menguji kestasioneran data dalam varian
2. Tahap Estimasi Model
 - `FitAR` : digunakan untuk mengestimasi model
3. Tahap Pemeriksaan Model
 - `JarqueBeraTest` : digunakan untuk menguji normalitas
 - `LjungBoxTest` : digunakan untuk menguji korelasi residual antar lag

Dalam paket ini mencakup fungsi untuk plot *time series*, pemeriksaan diagnostik, Box-Cox analisis, dan prosedur *time series* yang biasa digunakan.

2.2 Analisis Runtun Waktu

George E.P.Box dan Gwilym M. Jenkins dalam Makridakis et.al (1999) memperkenalkan analisis runtun waktu, yaitu pengamatan sekarang (Z_t) tergantung pada satu atau beberapa pengamatan sebelumnya (Z_{t-k}). Metode peramalan yang sering digunakan antara lain adalah metode ARIMA Box-Jenkins yang digunakan untuk mengolah runtun waktu yang univariat.

Menurut Wei (2006), suatu runtun waktu harus memenuhi syarat stasioneritas, yaitu nilai mean $E(Z_t) = \mu$ dan varians $Var(Z_t) = E(Z_t - \mu)^2 = \sigma^2$ konstan.

Uji stasioneritas data dalam *mean* digunakan Uji Dickey Fuller. Pengujian stasioneritas dalam *mean* juga dapat dilakukan dengan beberapa cara (Rosadi, 2012), yaitu menggunakan plot *time series*, plot fungsi autokorelasi (ACF) dan plot fungsi autokorelasi parsial (APCF) apakah data mengandung *trend* atau tidak. Jika data tidak stasioner dalam mean maka dilakukan differensi. Sedangkan untuk mendeteksi ketidakstasioneran dalam varian dapat digunakan plot ACF/PACF dari residual kuadrat atau menggunakan plot Box-Cox. Untuk melihat dan mengatasi ketidakstasioneran dalam varian dapat digunakan transformasi Box-Cox (Wei, 2006).

$$T(Z_t) = \frac{Z_t^\lambda - 1}{\lambda}$$

Salah satu model runtun waktu non-musiman adalah ARIMA (p, d, q). Bentuk umum model ini sebagai berikut :

$$\phi_p(B)(1-B)^d Z_t = \theta_q(B)a_t$$

dimana $\phi(B) = (1 - \phi_1 B - \dots - \phi_p B^p)$ merupakan operator AR(p) yang stasioner dan $\theta(B) = (1 - \theta_1 B - \dots - \theta_q B^q)$ merupakan operator MA(q) yang *invertible* dengan a_t independen dan berdistribusi normal dengan mean 0 dan varians σ_a^2 (Soejoeti, 1987).

Sedangkan menurut Zhang & McLeod (2006) model subset AR merupakan bagian dari model AR yang tergeneralisasi, sehingga tidak dapat dinyatakan dalam bentuk umum. Contoh model subset AR ([1,4]) dapat ditulis sebagai :

$$Z_t = 1 - \phi_1 B - \phi_4 B^4$$

Pemilihan model terbaik didasarkan pada nilai AIC terkecil pada masing-masing model yang diverifikasi (Wei, 2006).

3. METODOLOGI PENELITIAN

3.1 Sumber Data

Data yang akan digunakan pada karya ilmiah ini berupa *dataset* yang ada pada paket FitAR yaitu SeriesA. SeriesA adalah *dataset* mengenai proses konsentrasi kimia yang diamati tiap 2 jam selama 17 hari. Dalam pengerjaannya fungsi-fungsi pada paket yang digunakan yaitu:

Seleksi Model : TimeSeriesPlot, acf, pacf, SelectModel, BoxCox
 Estimasi Model : FitAR
 Pemeriksaan Model : JarqueBeraTest, LjungBoxTest

3.2 Teknik Pengolahan Data

1. Tahap Identifikasi

Pada tahap ini dilakukan pengujian stasioner data dalam mean dengan menggunakan plot *time series* dan uji formal menggunakan uji Dickey-Fuller serta untuk uji stasioner dalam varian dengan menggunakan plot Box-Cox. Selanjutnya membuat plot ACF dan PACF untuk menduga model awal.

2. Tahap Estimasi

Pada tahap ini dilakukan estimasi parameter berdasarkan model yang telah diduga pada tahap sebelumnya. Parameter-parameter tersebut diuji dengan menggunakan uji-t apakah signifikan atau tidak dalam model. Model yang digunakan adalah model yang semua parameternya signifikan.

3. Tahap Diagnosis

Pada tahap ini model yang telah diperoleh dilakukan pengujian asumsi-asumsi:

- a. Uji Independensi (*Ljung-Box test*), dimana akan terpenuhi jika tidak ada korelasi *residual* antar lag.
- b. Uji Normalitas (*Jarque-Bera test*), dimana akan terpenuhi jika *residual* berdistribusi normal.
- c. Uji Homoskedastisitas (*ARCH-LM test*), dimana akan terpenuhi jika residual homogen.

4. Tahap Peramalan

Pada tahap ini meramalkan data dengan model subset AR yang memiliki AIC terkecil.

4. HASIL DAN PEMBAHASAN

4.1 FitAR

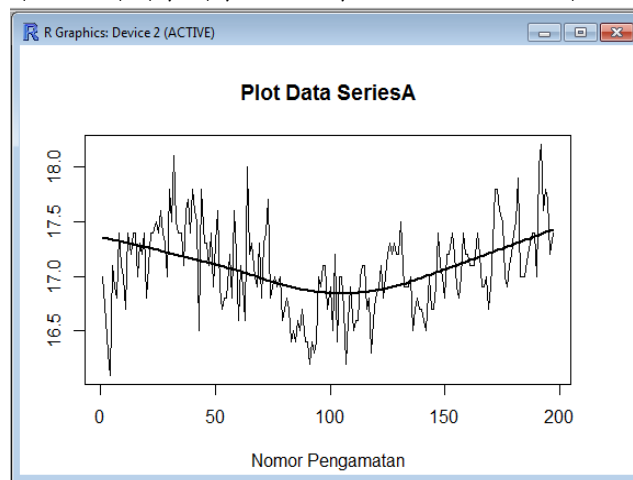
FitAR merupakan salah satu paket yang tersedia dalam *software R* yang mendukung untuk analisis data *time series*. Menurut McLeod & Zhang (2006), tahapan-tahapan yang digunakan untuk melakukan analisis *time series* subset AR bisa di implementasikan kedalam program atau paket FitAR.

Dalam FitAR pemilihan model dilakukan secara otomatis menggunakan fungsi `SelectModel` dan melakukan estimasi model menggunakan fungsi `FitAR`. Langkah awal dalam paket FitAR yaitu menghasilkan *time series* plot dari fungsi `TimeSeriesPlot` yang digunakan untuk uji stasioneritas data. Uji stasioneritas merupakan uji asumsi wajib dalam analisis *time series*. Apabila data stasioner maka dapat langsung diolah menggunakan paket FitAR, jika data tidak stasioner maka harus dilakukan perbaikan data menggunakan differensi atau transformasi. Kegunaan paket FitAR yaitu untuk mengolah data *time series* model AR ataupun subset AR.

4.2 Identifikasi Model

Perintah plot *time series* pada *Software R* untuk menampilkan plot data SeriesA dan merupakan langkah awal dalam uji stasioneritas.

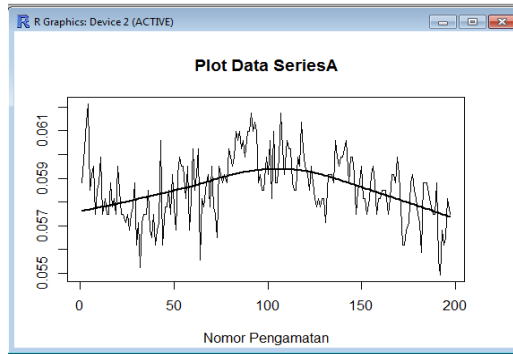
```
TimeSeriesPlot(z, aspect = 0.50, xlab = "Nomor Pengamatan",
ylab=NULL, main="Plot Data SeriesA")
lines(lowess(time(z), z), lwd=2, col="black")
```



Gambar 1 Plot Time Series Data SeriesA

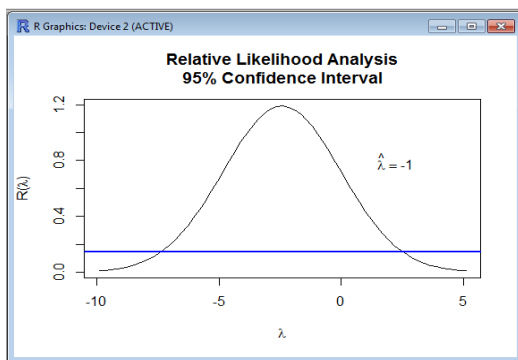
Stasioneritas dalam mean bisa dilihat pada Gambar 1 yang dihasilkan dari perintah `TimeSeriesPlot`. Pada plot tersebut dapat diduga data tidak stasioner karena runtun data tidak berada pada nilai tengah. Kemudian dilakukan pengujian secara formal dilakukan dengan menggunakan perintah `UnitRootTest`, Perintah `UnitRootTest` itu sendiri menggunakan uji Dickey-Fuller. Sehingga Diperoleh nilai $\text{prob} (0.000999001) > \alpha=5\%$ maka data stasioner dalam *mean*.

Uji untuk melihat kestasioneran data dalam varian digunakan uji Box-Cox, sehingga dihasilkan plot pada Gambar 2. Pada Gambar 2, terlihat bahwa nilai $\lambda = -1$ sehingga data SeriesA tidak stasioner dalam varian karena syarat stasioner dalam varian harus $\lambda=1$. Data yang tidak stasioner dalam varian dilakukan proses transformasi $1/Z_t$.

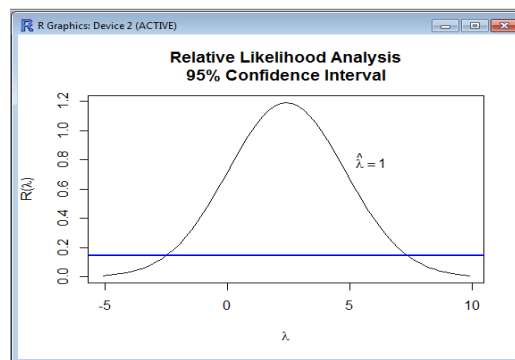


Gambar 2 Plot *Time Series* Setelah di Transformasi

Dapat dilihat pada Gambar 3 bahwa nilai λ sama dengan 1 sehingga dapat disimpulkan bahwa data sudah stasioner. Data SeriesA dapat disimpulkan bahwa data stasioner dalam *mean* dan varian.

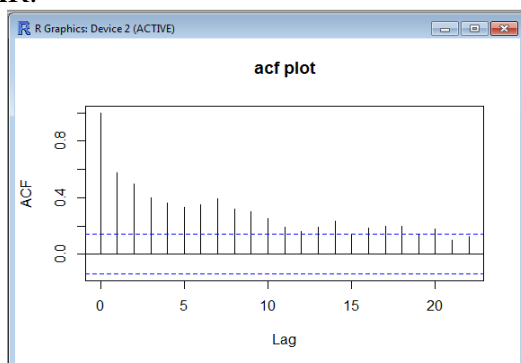


Gambar 3 Plot Box-Cox

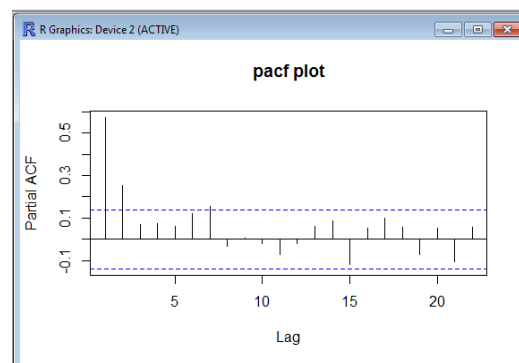


Gambar 4 Hasil Transformasi Box-Cox

Tahapan selanjutnya dalam identifikasi model *time series* adalah pendugaan model awal *time series* dengan melihat plot ACF seperti pada Gambar 4 dan PACF pada Gambar 5. Plot ACF pada Gambar 5 terlihat turun perlahan secara eksponensial dengan nilai ACF yang selalu positif. Sedangkan plot PACF pada Gambar 5 terdapat *cut off* pada lag ke-1, ke-2 dan ke-7. Sehingga dapat dikatakan bahwa data SeriesA mempunyai bentuk umum model AR.



Gambar 4 Plot ACF



Gambar 5 Plot PACF

Paket FitAR mempunyai program dalam penentuan pendugaan model awal *time series* selain menggunakan plot ACF dan PACF, yaitu dengan menggunakan fungsi `SelectModel`. Pada fungsi `SelectModel` dapat juga ditambahkan kriteria yang diinginkan misalnya dengan menampilkan beberapa model awal yaitu dengan perintah:

```
SelectModel(z, lag.max=10, ARModel="ARp", Best=2)
```

Sehingga didapat model awal dari data SeriesA yaitu AR(1) dan AR([1,2,7]).

4.3 Estimasi Parameter Model

Dalam paket FitAR, untuk mengestimasi parameter model digunakan fungsi `FitAR` guna mencari nilai MLE dan output lainnya yang dibutuhkan dalam estimasi parameter. Untuk memunculkan nilai t digunakan fungsi `getT` dimana nilai t akan digunakan dalam uji signifikansi model. Perintah yang digunakan untuk memunculkan nilai-nilai dari estimasi parameter pada model yaitu sebagai berikut:

```
ans <- FitAR(S,p,ARModel="ARp")
coef(ans)
getT(ans)
```

kemudian didapat nilai estimasi pada tiap parameter pada Tabel 1.

Tabel 1 Estimasi Parameter Model

Model	Parameter	Estimasi	Nilai t	Keputusan
AR (1)	ϕ_1	0.57304	-7.3123	H_0 ditolak
Subset AR ([1,2,7])	ϕ_1	0.35887	-9.1306	H_0 ditolak
	ϕ_2	0.22953	-10.9726	H_0 ditolak
	ϕ_7	0.22406	-11.0505	H_0 ditolak

dapat disimpulkan bahwa parameter pada model penduga signifikan.

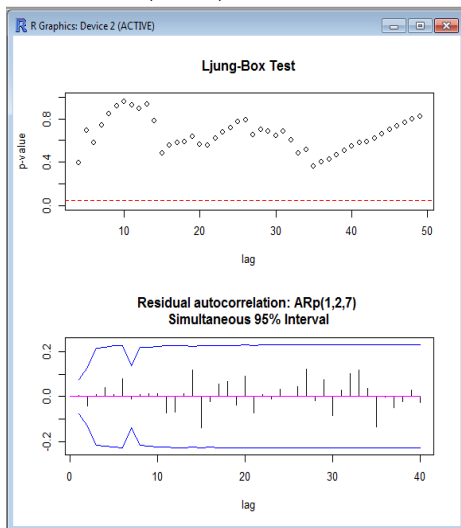
4.4 Pemeriksaan Diagnostik

Pada pemeriksaan diagnostik akan dilakukan uji independensi, uji homoskedastisitas residual, dan uji normalitas residual. Sebuah proses $\{a_t\}$ dikatakan bersifat *white noise* jika deret bersifat independen atau tidak terdapat korelasi serial antar residual dan berdistribusi normal dengan *mean* sama dengan nol dan varians konstan (Tsay, 2005).

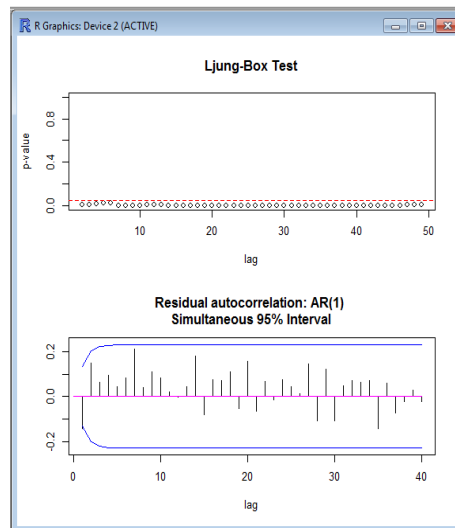
1. Uji Independensi Residual

Uji yang digunakan untuk uji independensi residual secara visual yaitu menggunakan plot Ljung-Box dan uji secara formal menggunakan uji Ljung-Box dengan perintah sebagai berikut :

```
ans <- FitAR(S,p,ARModel="ARp")
res <- resid(ans)
LjungBoxTest(res,k=0,lag.max=30,SquaredQ=FALSE)
Plot(ans)
```



Gambar 6 Model Subset AR([1,2,7])



Gambar 7 Model AR(1)

Secara visual Subset AR([1,2,7]) terdapat korelasi residual antar lag dan AR(1) terdapat korelasi residual antar lag.

Uji independensi residual secara formal menggunakan uji Ljung-Box. Fungsi paket `FitAR` yang digunakan adalah `LjungBoxTest`. Didapatkan hasil pada Tabel 2.

Tabel 2 Nilai statistik Ljung-box

Model	Lag	Qm(Q-Stat)	P-Value
AR([1,2,7])	2	0.37	0.8312
	12	4.21	0.9794
	13	4.24	0.9883
	14	7.26	0.9243
AR(1)	2	8.59	0.0136
	12	26.81	0.0082
	13	27.19	0.0117
	14	34.12	0.0019

Untuk model AR([1,2,7]) nilai p-value > 0.05 sehingga dapat disimpulkan bahwa model AR([1,2,7]) tidak ada korelasi residual antar lag, kemudian untuk model AR(1) ada beberapa nilai p-value < 0.05 sehingga dapat disimpulkan bahwa model AR(1) terdapat korelasi residual antar lag dan untuk model AR(1) tidak dapat digunakan lagi.

2. Uji Normalitas Residual

Digunakan untuk mengetahui residual berdistribusi normal atau tidak. Uji normalitas residual dilakukan menggunakan uji *Jarque Bera*.

`JarqueBeraTest(z)`

Diperoleh hasil p-value (0.8245275) > 0.05 sehingga dapat disimpulkan bahwa residual data SeriesA berdistribusi normal.

3. Uji Homoskedastisitas Residual

Digunakan untuk mengetahui bahwa tidak terdapat heteroskedastisitas dalam residual model. Uji Homoskedastisitas residual menggunakan uji *Lagrange Multiplier (LM)*.

Diperoleh nilai $LM = 82.65923 < \chi^2_{(0.05;1950)} = (230.46)$ maka dapat disimpulkan bahwa residual homogen.

Dari uji signifikansi parameter model dan pemeriksaan diagnostik yang telah dilakukan pada model yang telah diidentifikasi sebelumnya, diperoleh hasil model yang parameternya signifikan, memenuhi asumsi independensi residual, homoskedastisitas residual, dan normalitas residual adalah model subset AR([1,2,7]). Kemudian dibentuk model subset AR([1,2,7]) dengan melihat estimasi parameternya pada Tabel 3 yang akan digunakan dalam proses peramalan.

Tabel 3 Estimasi Parameter model Subset AR

Paramter	Estimate	Std. Error	Z-Ratio
ϕ_1	0.36462	0.07022	5.11098
ϕ_2	0.22351	0.06935	3.31004
ϕ_7	0.22505	0.00039	3.56565

Dengan model yang terbentuk untuk data SeriesA adalah sebagai berikut:

$$\hat{Z}_t = \frac{a_t}{1 - 0.36462B - 0.22351B^2 - 0.22505B^7}$$

4.5 Peramalan

Dari model yang telah didapatkan, selanjutnya dilakukan proses peramalan. Proses peramalan dilakukan sebanyak 10 langkah kedepan. Setelah didapatkan hasil peramalan sebanyak 10 langkah kedepan kemudian apakah model yang didapatkan baik untuk dilakukan proses peramalan secara menyeluruh.

Tabel 4 Hasil Peramalan

No.	Data aktual	Data Peramalan
188	0.0575	0.0581
189	0.0575	0.0580
190	0.0588	0.0580
191	0.0556	0.0551
192	0.0549	0.0551
193	0.0568	0.0572
194	0.0562	0.0567
195	0.0565	0.0561
196	0.0581	0.0579
197	0.0575	0.0579
198		0.0580
199		0.0583
200		0.0584
201		0.0585
202		0.0586
203		0.0586
204		0.0586
205		0.0586
206		0.0586
207		0.0586

Perintah yang digunakan untuk melakukan peramalan dengan menggunakan software R:


```

ans <- SelectModel(S, ARModel="ARp", lag.max=24, Best=1)
vans <- FitAR(S, ans, ARModel="ARp")
out<-predict(vans, n.ahead=10)
zf<-out$Forecasts
zsd<-out$SDForecasts
y<-cts(S, zf)
y

```

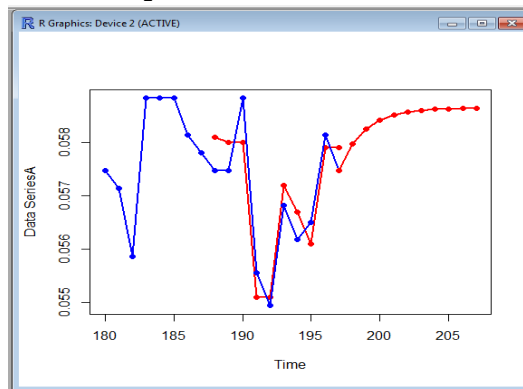
Hasil ramalan data SeriesA Subset AR([1,2,7]) untuk 10 langkah kedepan serta 10 hampir mendekati aktual seperti pada Tabel 4.

Kemudian untunk memunculkan plot data digunakan perintah :

```

plot(window(y, start=190), type="n", ylab="Data SeriesA")
y1<-window(y, start=190, end=197)
lines.ts(y1, col="blue", type="o", lwd=2, pch=16)
y2<-window(y, start=c(197,1))
lines.ts(y2, col="red", type="o", lwd=2, pch=16)
legend(197, 12, legend=c("observed", "forecast"), col=c("red",
"blue"), lwd=c(2, 2), pch=c(16, 16))

```



Gambar 10 Plot Data Hasil Peramalan

5. KESIMPULAN

Subset AR (*Autoregressive*) sangat cocok digunakan apabila dalam suatu analisis data *time series* menemukan kendala dimana pada plot PACF terdapat lag-lag yang *cut-off* tidak secara berurutan sehingga penerapan pada metode AR yang biasa akan sulit untuk dilakukan. Dalam pengerjaannya fungsi-fungsi pada paket yang digunakan yaitu:

Seleksi Model : TimeSeriesPlot, acf, pacf,
SelectModel, BoxCox

Estimasi Model : FitAR, getT

Pemeriksaan Model : JarqueBeraTest, LjungBoxTest

Model terbaik adalah AR ([1,2,7]) dengan nilai AIC -2625.8, model tersebut parameternya sudah signifikan, memenuhi asumsi independensi, homoskedastisitas residual dan normalitas residual.

DAFTAR PUSTAKA

- Makridakis, S., Wheelwright, S.C., and McGee, V.E. 1999. *Metode dan Aplikasi Peramalan*. Jilid satu edisi kedua, Terjemahan Ir. Hari Suminto. Jakarta. Bina Rupa Aksara.
- Rosadi, D. 2012. *Ekonometrika & Analisis Runtun Waktu Terapan dengan Eviews*. Yogyakarta. Penerbit ANDI.
- Soejoeti, Z. 1987. *Materi Pokok Analisis Runtun Waktu*. Jakarta. Karunika.
- Tsay, RS. 2005. *Analysis of Financial Time Series*. John Wiley and Sons, Inc. Chicago.

- Wei, W.W.S. 2006. *Time Series Analysis, Univariate and Multivariate Methods*. Canada. Addison Wesley Publishing Company.
- Zhang, Y and McLeod A.I. 2008. *Improved Subset Autoregression: With R Package*. The American Statistical Association Volume 20, Issue 2.