

## **ARIMA (Autoregressive Integrated Moving Average)**

### **I. Prinsip Dasar dan Tujuan Analisis**

#### **1.1 Prinsip Dasar**

ARIMA sering juga disebut metode runtun waktu Box-Jenkins. ARIMA sangat baik ketepatannya untuk peramalan jangka pendek, sedangkan untuk peramalan jangka panjang ketepatan peramalannya kurang baik. Biasanya akan cenderung *flat* (mendatar/konstan) untuk periode yang cukup panjang.

Model *Autoregresif Integrated Moving Average* (ARIMA) adalah model yang secara penuh mengabaikan independen variabel dalam membuat peramalan. ARIMA menggunakan nilai masa lalu dan sekarang dari variabel dependen untuk menghasilkan peramalan jangka pendek yang akurat. ARIMA cocok jika observasi dari deret waktu (*time series*) secara statistik berhubungan satu sama lain (*dependent*).

#### **1.2 Tujuan Analisis**

Tujuan model ini adalah untuk menentukan hubungan statistik yang baik antar variabel yang diramal dengan nilai historis variabel tersebut sehingga peramalan dapat dilakukan dengan model tersebut.

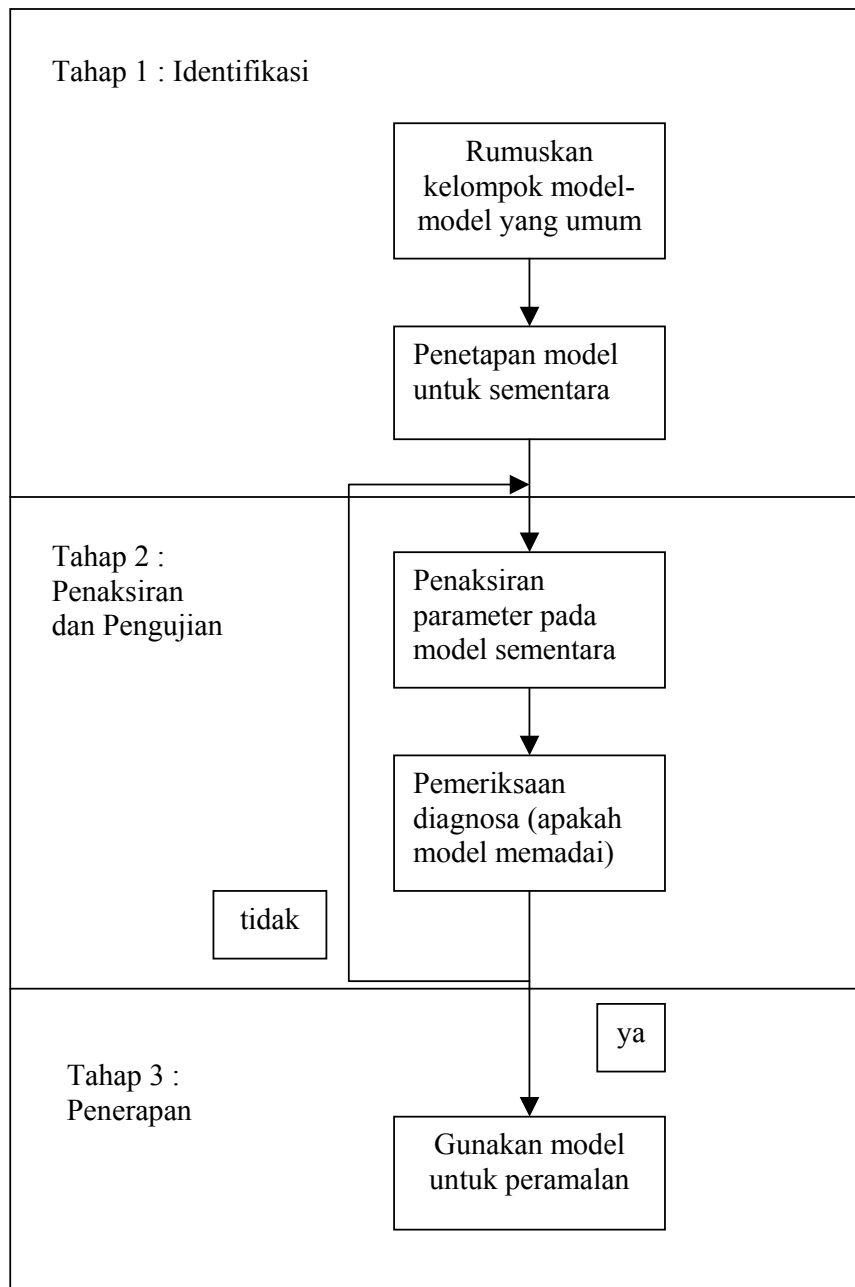
### **II. Format Data Dasar dan Program Komputer yang Digunakan**

ARIMA hanya menggunakan suatu variabel (*univariate*) deret waktu. Misalnya: variabel IHSG. Program komputer yang dapat digunakan adalah EViews, Minitab, SPSS, dll.

### **III. Model Matematis dan Algoritma Pokok Analisis**

Model ARIMA terdiri dari tiga langkah dasar, yaitu tahap identifikasi, tahap penaksiran dan pengujian, dan pemeriksaan diagnostik. Selanjutnya model ARIMA dapat digunakan untuk melakukan peramalan jika model yang diperoleh memadai.

## SKEMA PENDEKATAN BOX JENKINS



### Stasioneritas dan Nonstasioneritas

Hal yang perlu diperhatikan adalah bahwa kebanyakan deret berkala bersifat nonstasioner dan bahwa aspek-aspek AR dan MA dari model ARIMA hanya berkenaan dengan deret berkala yang stasioner.

Stasioneritas berarti tidak terdapat pertumbuhan atau penurunan pada data. Data secara kasarnya harus horizontal sepanjang sumbu waktu. Dengan kata lain, fluktuasi data

berada di sekitar suatu nilai rata-rata yang konstan, tidak tergantung pada waktu dan varians dari fluktuasi tersebut pada pokoknya tetap konstan setiap waktu.

Suatu deret waktu yang tidak stasioner harus diubah menjadi data stasioner dengan melakukan *differencing*. Yang dimaksud dengan *differencing* adalah menghitung perubahan atau selisih nilai observasi. Nilai selisih yang diperoleh dicek lagi apakah stasioner atau tidak. Jika belum stasioner maka dilakukan *differencing* lagi. Jika varians tidak stasioner, maka dilakukan transformasi logaritma.

### Klasifikasi model ARIMA

Model Box-Jenkins (ARIMA) dibagi kedalam 3 kelompok, yaitu: model *autoregressive* (AR), *moving average* (MA), dan model campuran ARIMA (*autoregressive moving average*) yang mempunyai karakteristik dari dua model pertama.

#### 1) Autoregressive Model (AR)

Bentuk umum model *autoregressive* dengan ordo  $p$  (AR( $p$ )) atau model ARIMA ( $p,0,0$ ) dinyatakan sebagai berikut:

$$X_t = \mu' + \phi_1 X_{t-1} + \phi_2 X_{t-2} + \dots + \phi_p X_{t-p} + e_t [0]$$

dimana:  $\mu'$  = suatu konstanta

$\phi_p$  = parameter autoregresif ke- $p$

$e_t$  = nilai kesalahan pada saat  $t$

#### 2) Moving Average Model (MA)

Bentuk umum model *moving average* ordo  $q$  (MA( $q$ )) atau ARIMA ( $0,0,q$ ) dinyatakan sebagai berikut:

$$X_t = \mu' + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \dots - \theta_q e_{t-k}$$

dimana:  $\mu'$  = suatu konstanta

$\theta_1$  sampai  $\theta_q$  adalah parameter-parameter *moving average*

$e_{t-k}$  = nilai kesalahan pada saat  $t - k$

#### 3) Model campuran

##### a. Proses ARMA

Model umum untuk campuran proses AR(1) murni dan MA(1) murni, misal ARIMA (1,0,1) dinyatakan sebagai berikut:

$$X_t = \mu' + \phi_1 X_{t-1} + e_t - \theta_1 e_{t-1}$$

atau

$$(1 - \phi_1 B)X_t = \mu' + (1 - \theta_1 B)e_t$$

AR(1)                      MA(1)

b. Proses ARIMA

Apabila nonstasioneritas ditambahkan pada campuran proses ARMA, maka model umum ARIMA  $(p,d,q)$  terpenuhi. Persamaan untuk kasus sederhana ARIMA  $(1,1,1)$  adalah sebagai berikut:

$$(1 - B)(1 - \phi_1 B)X_t = \mu' + (1 - \theta_1 B)e_t$$

pembedaan      AR(1)              MA(1)  
pertama

### Musiman dan Model ARIMA

Musiman didefinisikan sebagai suatu pola yang berulang-ulang dalam selang waktu yang tetap. Untuk data yang stasioner, faktor musiman dapat ditentukan dengan mengidentifikasi koefisien autokorelasi pada dua atau tiga *time-lag* yang berbeda nyata dari nol. Autokorelasi yang secara signifikan berbeda dari nol menyatakan adanya suatu pola dalam data. Untuk mengenali adanya faktor musiman, seseorang harus melihat pada autokorelasi yang tinggi.

Untuk menangani musiman, notasi umum yang singkat adalah:

$$\text{ARIMA } (p,d,q) (P,D,Q)^S$$

Dimana  $(p,d,q)$  = bagian yang tidak musiman dari model

$(P,D,Q)$  = bagian musiman dari model

S = jumlah periode per musim

### Identifikasi

Proses identifikasi dari model musiman tergantung pada alat-alat statistik berupa autokorelasi dan parsial autokorelasi, serta pengetahuan terhadap sistem (atau proses) yang dipelajari.

### Penaksiran Parameter

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

- a. Dengan cara mencoba-coba (*trial and error*), menguji beberapa nilai yang berbeda dan memilih satu nilai tersebut (atau sekumpulan nilai, apabila terdapat lebih dari

satu parameter yang akan ditaksir) yang meminimumkan jumlah kuadrat nilai sisa (*sum of squared residual*).

- b. Perbaiki secara iteratif, memilih taksiran awal dan kemudian membiarkan program komputer memperhalus penaksiran tersebut secara iteratif.

### **Pengujian Parameter Model**

1. Pengujian masing-masing parameter model secara parsial (*t-test*)
2. Pengujian model secara keseluruhan (*Overall F test*)

Model dikatakan baik jika nilai error bersifat random, artinya sudah tidak mempunyai pola tertentu lagi. Dengan kata lain model yang diperoleh dapat menangkap dengan baik pola data yang ada. Untuk melihat kerandoman nilai error dilakukan pengujian terhadap nilai koefisien autokorelasi dari error, dengan menggunakan salah satu dari dua statistik berikut:

- 1) Uji Q Box dan Pierce:

$$Q = n' \sum_{k=1}^m r_k^2$$

- 2) Uji Ljung-Box:

$$Q = n'(n'+2) \sum_{k=1}^m \frac{r_k^2}{(n'-k)}$$

Menyebar secara Khi Kuadrat ( $\chi^2$ ) dengan derajat bebas (db)=(k-p-q-P-Q) dimana:

$n' = n-(d+SD)$

$d$  = ordo pembedaan bukan faktor musiman

$D$  = ordo pembedaan faktor musiman

$S$  = jumlah periode per musim

$m$  = lag waktu maksimum

$r_k$  = autokorelasi untuk time lag 1, 2, 3, 4,...,  $k$

Kriteria pengujian:

Jika  $Q \leq \chi^2_{(\alpha, db)}$ , berarti: nilai error bersifat random (model dapat diterima).

Jika  $Q > \chi^2_{(\alpha, db)}$ , berarti: nilai error tidak bersifat random (model tidak dapat diterima).

## Pemilihan Model Terbaik

Untuk menentukan model yang terbaik dapat digunakan *standard error estimate* berikut:

$$S = \left[ \frac{SSE}{n - n_p} \right]^{\frac{1}{2}} = \left[ \frac{\sum_{t=1}^n (Y_t - \hat{Y}_t)^2}{n - n_p} \right]^{\frac{1}{2}}$$

dimana:

$Y_t$  = nilai sebenarnya pada waktu ke-t

$\hat{Y}_t$  = nilai dugaan pada waktu ke-t

Model terbaik adalah model yang memiliki nilai *standard error estimate* (S) yang paling kecil.

Selain nilai *standard error estimate*, nilai rata-rata persentase kesalahan peramalan (MAPE) dapat juga digunakan sebagai bahan pertimbangan dalam menentukan model yang terbaik yaitu:

$$MAPE = \frac{\sum_{t=1}^T \frac{|Y_t - \hat{Y}_t|}{Y_t}}{T} \times 100\%$$

dimana:

$T$  = banyaknya periode peramalan/dugaan.

## Peramalan Dengan Model ARIMA

Notasi yang digunakan dalam ARIMA adalah notasi yang mudah dan umum. Misalkan model ARIMA (0,1,1)(0,1,1)<sup>12</sup> dijabarkan sebagai berikut:

$$(1 - B)(1 - B^{12})X_t = (1 - \theta_1 B)(1 - \Theta_1 B^{12})e_t$$

Tetapi untuk menggunakannya dalam peramalan mengharuskan dilakukan suatu penjabaran dari persamaan tersebut dan menjadikannya sebuah persamaan regresi yang lebih umum. untuk model diatas bentuknya adalah:

$$X_t = X_{t-1} + X_{t-12} - X_{t-13} + e_t - \theta_1 e_{t-1} - \Theta_1 e_{t-12} + \theta_1 \Theta_1 e_{t-13}$$

Untuk meramalkan satu periode ke depan, yaitu  $X_{t+1}$  maka seperti pada persamaan berikut:

$$X_{t+1} = X_t + X_{t-1} - X_{t-2} + e_{t+1} - \theta_1 e_t - \Theta_1 e_{t-1} + \theta_1 \Theta_1 e_{t-2}$$

Nilai  $e_{t+1}$  tidak akan diketahui, karena nilai yang diharapkan untuk kesalahan random pada masa yang akan datang harus ditetapkan sama dengan nol. Akan tetapi dari model yang disesuaikan (*fitted model*) kita boleh mengganti nilai  $e_t$ ,  $e_{t-1}$  dan  $e_{t-2}$  dengan nilai-nilai mereka yang ditetapkan secara empiris (seperti yang diperoleh setelah iterasi terakhir algoritma Marquardt). Tentu saja bila kita meramalkan jauh ke depan, tidak akan kita peroleh nilai empiris untuk “ $e$ ” sesudah beberapa waktu, dan oleh sebab itu nilai harapan mereka akan seluruhnya nol.

Untuk nilai  $X$ , pada awal proses peramalan, kita akan mengetahui nilai  $X_t$ ,  $X_{t-1}$ ,  $X_{t-2}$ . Akan tetapi sesudah beberapa saat, nilai  $X$  akan berupa nilai ramalan (*forecasted value*), bukan nilai-nilai masa lalu yang telah diketahui.

#### IV. Struktur Informasi Pokok Hasil Analisis (Cara Interpretasi)

1. Identifikasi.
  - a. Berdasarkan plot data aktual dapat diketahui apakah data sudah stasioner. Jika belum stasioner maka data harus distasionerkan terlebih dahulu.
  - b. Tentukan kombinasi model ARIMA yang mungkin. Dari plot autokorelasi tentukan ordo MA (q), dari plot autokorelasi parsial tentukan orde AR (p).
2. Estimasi dan pengujian model ARIMA yang mungkin serta pemilihan model terbaik.
3. Tentukan persamaan dan nilai ramalan model ARIMA terbaik.

#### V. Contoh Aplikasi Analisis

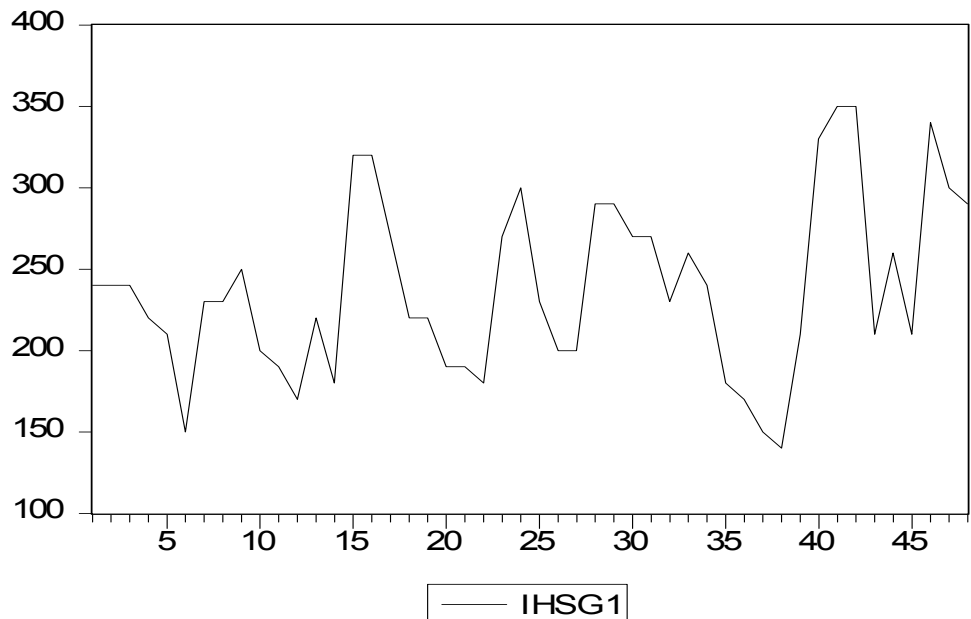
Misalkan kita ingin meramalkan nilai IHSG harian untuk jangka pendek.

Contoh: data IHSG harian (fiktif) selama 48 periode:

Hari	IHSG	Hari	IHSG	Hari	IHSG	Hari	IHSG	Hari	IHSG	Hari	IHSG
1	240	9	250	17	270	25	230	33	260	41	350
2	240	10	200	18	220	26	200	34	240	42	350
3	240	11	190	19	220	27	200	35	180	43	210
4	220	12	170	20	190	28	290	36	170	44	260
5	210	13	220	21	190	29	290	37	150	45	210
6	150	14	180	22	180	30	270	38	140	46	340
7	230	15	320	23	270	31	270	39	210	47	300
8	230	16	320	24	300	32	230	40	330	48	290

Data tersebut cenderung sudah stasioner, artinya nilai tengah dan varian tetap tidak tergantung pada perubahan waktu. Plot data adalah sebagai berikut:

Gambar 1. Plot data IHSG



Tabel 1. Hasil *Unit Root Test* dari Data IHSG

ADF Test Statistic	-3.735113	1% Critical Value*	-3.5778
		5% Critical Value	-2.9256
		10% Critical Value	-2.6005

\*MacKinnon critical values for rejection of hypothesis of a unit root.

Tabel 2. Plot autokorelasi data IHSG

Date: 09/04/04 Time: 22:37  
Sample: 1 48  
Included observations: 48

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob	
.  ****	.  ****	1	0.497	0.497	12.615	0.000
.  *	.  *	2	0.126	-0.160	13.449	0.001
** .	*** .	3	-0.291	-0.387	17.959	0.000
** .	.  **	4	-0.206	0.206	20.268	0.000
.  *	.  .	5	-0.122	-0.042	21.100	0.001
.  .	.  .	6	0.041	-0.051	21.197	0.002
.  .	.  .	7	0.026	-0.010	21.236	0.003
.  .	.  .	8	0.034	0.013	21.304	0.006
.  *	** .	9	-0.147	-0.232	22.625	0.007
.  *	.  .	10	-0.179	-0.038	24.640	0.006
.  *	.  *	11	-0.096	0.163	25.244	0.008
.  .	.  *	12	0.028	-0.126	25.295	0.013
.  *	.  .	13	0.102	-0.002	26.005	0.017
.  .	.  .	14	0.055	0.022	26.221	0.024



.  *	.  *	15	0.087	0.128	26.774	0.031
.  *	.  *	16	0.184	0.193	29.303	0.022
.  *	.  *	17	0.161	-0.085	31.310	0.018
.  *	.  *	18	0.089	0.036	31.950	0.022
.  *	.  *	19	-0.113	-0.145	33.010	0.024
**  *	.  *	20	-0.216	-0.098	36.991	0.012

Dari plot autokorelasi dan plot autokorelasi parsial, terlihat bahwa lag 1 signifikan. Sehingga ordo p dan q yang mungkin adalah 1. Kombinasi model ARIMA yang mungkin: ARIMA (0,0,1), ARIMA (1,0,1), ARIMA (1,0,0).

Selanjutnya cari nilai koefisiennya (penaksiran parameter) dengan menggunakan EViews didapatkan hasil sbb:

Hasil pengolahan output model ARIMA:

#### ARIMA (0,0,1)

Dependent Variable: IHSG1  
Method: Least Squares  
Date: 09/04/04 Time: 22:44  
Sample: 1 48  
Included observations: 48  
Convergence achieved after 52 iterations  
Backcast: 0

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	238.2940	9.780283	24.36474	0.0000
MA(1)	0.386536	0.138313	2.794646	0.0076
R-squared	0.202873	Mean dependent var		237.9167
Adjusted R-squared	0.185544	S.D. dependent var		54.34164
S.E. of regression	49.04183	Akaike info criterion		10.66400
Sum squared resid	110634.6	Schwarz criterion		10.74196
Log likelihood	-253.9360	F-statistic		11.70721
Durbin-Watson stat	1.701802	Prob(F-statistic)		0.001316
Inverted MA Roots	-.39			

Date: 09/04/04 Time: 22:42

Sample: 1 48

Included observations: 48

Q-statistic  
probabilities  
adjusted for 1  
ARMA term(s)

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob	
.  *	.  *	1	0.140	0.140	0.9959	
.  **	.  **	2	0.220	0.204	3.5161	0.061
***  *	***  *	3	-0.356	-0.435	10.279	0.006
.  *	.  *	4	-0.060	0.021	10.478	0.015
.  *	.  *	5	-0.142	0.069	11.601	0.021
.  *	.  *	6	0.097	-0.040	12.142	0.033
.  *	.  *	7	-0.025	-0.053	12.178	0.058
.  *	.  *	8	0.109	0.087	12.891	0.075
.  *	.  *	9	-0.141	-0.159	14.120	0.079
.  *	.  *	10	-0.123	-0.188	15.075	0.089
.  *	.  *	11	-0.081	0.158	15.505	0.115

.   .	.   .	12	0.014	-0.026	15.517	0.160
.   *	.   *	13	0.095	-0.059	16.132	0.185
.   .	.   .	14	0.007	-0.015	16.136	0.242
.   .	.   .	15	0.038	0.056	16.240	0.299
.   *	.   **	16	0.163	0.251	18.221	0.251
.   *	.   .	17	0.084	-0.033	18.761	0.281
.   *	.   .	18	0.108	0.027	19.702	0.290
.   *	.   .	19	-0.102	-0.048	20.564	0.302
.   *	.   **	20	-0.158	-0.200	22.711	0.250

### ARIMA(1,0,1)

Dependent Variable: IHSG1

Method: Least Squares

Date: 09/04/04 Time: 22:46

Sample(adjusted): 2 48

Included observations: 47 after adjusting endpoints

Convergence achieved after 11 iterations

Backcast: 1

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	238.7642	13.66360	17.47448	0.0000
AR(1)	0.424111	0.269218	1.575342	0.1223
MA(1)	0.116910	0.293080	0.398903	0.6919
R-squared	0.258929	Mean dependent var		237.8723
Adjusted R-squared	0.225244	S.D. dependent var		54.92826
S.E. of regression	48.34797	Akaike info criterion		10.65643
Sum squared resid	102851.2	Schwarz criterion		10.77452
Log likelihood	-247.4260	F-statistic		7.686773
Durbin-Watson stat	1.952965	Prob(F-statistic)		0.001371
Inverted AR Roots	.42			
Inverted MA Roots	-.12			

Date: 09/04/04 Time: 22:45

Sample: 2 48

Included observations: 47

Q-statistic  
probabilities  
adjusted for 2  
ARMA term(s)

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob
.   .	.   .	1	0.021	0.021	0.0211
.   *	.   *	2	0.130	0.130	0.8844
***   .	***   .	3	-0.394	-0.406	9.0056 0.003
.   .	.   .	4	-0.042	-0.028	9.0986 0.011
.   *	.   .	5	-0.097	0.017	9.6175 0.022
.   *	.   .	6	0.113	-0.038	10.338 0.035
.   .	.   .	7	-0.017	-0.046	10.354 0.066
.   *	.   *	8	0.155	0.138	11.778 0.067
.   *	.   *	9	-0.126	-0.129	12.743 0.079
.   *	.   **	10	-0.116	-0.203	13.585 0.093
.   *	.   *	11	-0.069	0.116	13.887 0.126
.   .	.   .	12	0.022	-0.023	13.918 0.177
.   *	.   .	13	0.106	-0.035	14.681 0.198
.   .	.   *	14	-0.047	-0.080	14.836 0.251
.   .	.   .	15	-0.012	0.000	14.847 0.317
.   *	.   **	16	0.149	0.247	16.492 0.284
.   .	.   .	17	0.063	-0.001	16.793 0.331
.   *	.   *	18	0.123	0.094	17.986 0.325

.*	.	.	.	19	-0.102	-0.015	18.839	0.338
.*	.	.*	.	20	-0.139	-0.183	20.477	0.307

### ARIMA (1,0,0)

Dependent Variable: IHSG1

Method: Least Squares

Date: 09/04/04 Time: 22:47

Sample(adjusted): 2 48

Included observations: 47 after adjusting endpoints

Convergence achieved after 3 iterations

Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	238.9670	14.22567	16.79829	0.0000
AR(1)	0.507133	0.130226	3.894240	0.0003
R-squared	0.252058	Mean dependent var		237.8723
Adjusted R-squared	0.235437	S.D. dependent var		54.92826
S.E. of regression	48.02888	Akaike info criterion		10.62310
Sum squared resid	103804.8	Schwarz criterion		10.70183
Log likelihood	-247.6429	F-statistic		15.16510
Durbin-Watson stat	1.842423	Prob(F-statistic)		0.000324

Inverted AR Roots .51

Date: 09/04/04 Time: 22:47

Sample: 2 48

Included observations: 47

Q-statistic

probabilities

adjusted for 1

ARMA term(s)

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob	
.   *	.   *	1	0.077	0.077	0.2956	
.   *	.   *	2	0.075	0.070	0.5865	0.444
***   .	***   .	3	-0.405	-0.420	9.1524	0.010
*   .	.   .	4	-0.075	-0.012	9.4560	0.024
*   .	.   .	5	-0.083	-0.003	9.8318	0.043
.   *	.   .	6	0.112	-0.050	10.534	0.061
.   .	.   .	7	0.007	-0.034	10.537	0.104
.   *	.   *	8	0.155	0.140	11.959	0.102
*   .	*   .	9	-0.119	-0.144	12.822	0.118
*   .	*   .	10	-0.129	-0.184	13.853	0.128
*   .	.   *	11	-0.072	0.121	14.187	0.165
.   .	.   .	12	0.029	-0.035	14.243	0.220
.   *	.   .	13	0.106	-0.038	14.998	0.242
.   .	*   .	14	-0.053	-0.092	15.191	0.296
.   .	.   .	15	-0.019	0.007	15.216	0.364
.   *	.   **	16	0.147	0.239	16.829	0.329
*   .	.   .	17	0.080	-0.006	17.317	0.365
.   *	.   *	18	0.122	0.109	18.500	0.358
*   .	.   .	19	-0.103	-0.022	19.370	0.369
*   .	*   .	20	-0.146	-0.166	21.185	0.327

Uji t untuk masing-masing parameter, terlihat bahwa pada ARIMA (0,0,1) dan ARIMA (1,0,0) ternyata t-hitung > dari t-tabel, dan mempunyai nilai prob < 0,005 sehingga dengan alpha 5%,  $H_0$  ditolak, artinya koefisien signifikan (berbeda dari nol). Sementara untuk

ARIMA (1,0,1) mempunyai t-hitung < dari t-tabel, sehingga  $H_0$  diterima, artinya koefisien tidak signifikan.

Pengujian secara keseluruhan (*overall test*):

Hipotesa nol: tidak ada korelasi pada nilai sisa (residual)

Dengan menggunakan Q-Stat, terlihat bahwa pada ARIMA (0,1,1) pada lag 3 yaitu 0.006 lebih besar dari nilai Chi-Square table (nilai prob sebesar 0.006) artinya dengan alpha 5% ada korelasi pada nilai sisa (lag 3) sehingga model tidak cocok.

Jadi dari 3 model yang mungkin, hanya ada 1 model yang memenuhi syarat, yaitu ARIMA (1,0,0). Seandainya ada lebih dari satu model yang memenuhi syarat, maka ambil model yang terbaik sesuai dengan kriteria yang telah dijelaskan sebelumnya.

Didapatkan koefisien AR(1) dengan bentuk persamaan:

$$X_t = 16.79829 + 3.894240X_{t-1} + e_t$$

Model sudah dapat digunakan untuk peramalan

**Referensi:**

Deden. *Summary (Diktat Kuliah ADW)*. STIS. 2004.

Hendranata, Anton. *ARIMA (Autoregressive Moving Average)*, Manajemen Keuangan Sektor Publik FEUI, 2003.

Makridakis, Spyros. , Steven C. Wheelwright, dan Victor E. McGee. *Metode dan Aplikasi Peramalan*, Jakarta: Erlangga, 1999.