

**DETEKSI *OUTLIER* PADA DATA *TIME SERIES***

**OLEH**

**A ROFIQI MAULANA**

**125090500111025**



**PROGRAM STUDI STATISTIKA**

**JURUSAN MATEMATIKA**

**FAKULTAS MATEMATIKA DAN ILMU PENGETAHUAN ALAM**

**UNIVERSITAS BRAWIJAYA**

**2015**

## **BAB I**

### **PENDAHULUAN**

Model Autoregressive Integrated Moving Average (ARIMA) merupakan model yang digunakan untuk meramalkan data time series. Model ARIMA menghendaki data time series memenuhi asumsi stasioner pada varian maupun mean. Peristiwa yang terjadi diluar kendali, dimungkinkan dapat mempengaruhi stasioner data time series. Peristiwa tersebut dinamakan intervensi, suatu intervensi dapat berupa perubahan keadaan ekonomi nasional, bencana alam, kebijakan, promosi, dan peristiwa tidak terduga lainnya.

Pada sekumpulan data deret waktu, kadang-kadang terdapat data yang nilainya jauh berbeda dari data lainnya dan tidak mencerminkan karakteristik data secara umum. Nilai ini dinamakan dengan outlier. Pada tahun 1960 Gumbel mende\_nisikan bahwa outlier adalah nilai yang tampak terlalu besar atau terlalu kecil dibandingkan dengan keseluruhan pengamatan. Grubbs (1969) mende\_nisikan bahwa pengamatan outlier adalah salah satu pengamatan yang tampak menyimpang tajam dari pengamatan-pengamatan lainnya. Pada pengolahan data deret waktu, Gounder et.al (2007) menjelaskan bahwa kehadiran outlier akan mempengaruhi analisis data seperti identifikasi model, estimasi parameter dan peramalan. Untuk mendeteksi kehadiran outlier tersebut Fox (1972) memperkenalkan metode untuk deteksi outlier padadata deret waktu yang dikenal dengan outlier tipe 1 atau Additive Outliers (AO) dan tipe 2 atau Innovation Outliers (IO).

Permasalahan inflasi di Indonesia memegang peranan yang sangat penting dalam perekonomian nasional. Berbagai permasalahan makroekonomi seperti kemiskinan, pengangguran, dan pertumbuhan ekonomi sangat dipengaruhi oleh tingkat inflasi yang terjadi. Apa lagi jika mempertimbangkan bahwa hampir separuh dari sumber pertumbuhan ekonomi Indonesia berasal dari konsumsi, baik rumah tangga, pemerintah, swasta maupun Non Government Organization.

Tidak seperti negara-negara maju yang memiliki tingkat inflasi sangat rendah, tingkat inflasi di Indonesia berfluktuasi cukup tinggi sejalan dengan pertumbuhan ekonomi. Hal ini tentu memberikan dampak langsung terhadap menurunnya daya beli masyarakat. Namun karena adanya ekspektasi terhadap inflasi juga memberikan rangsangan terhadap tumbuhnya perekonomian, pemerintah melalui bank sentral selaku pemegang regulasi berkepentingan untuk tetap mempertahankan tingkat inflasi nasional pada level tertentu.

Mengingat betapa pentingnya pengetahuan mengenai model tingkat inflasi bagi perencanaan kebijakan pembangunan kedepan, maka dalam penelitian ini akan dilakukan pemodelan dengan beberapa metode, diantaranya ARIMA dengan deteksi outlier.

## BAB II TINJAUAN PUSTAKA

### 2.1 Analisis *Outlier*

Data deret waktu merupakan serangkaian data yang berurutan berdasarkan waktu. Data deret waktu dapat digunakan untuk menduga proyeksi masa depan dari suatu variabel berdasarkan pada data masa lalu dan sekarang. Analisis data deret waktu telah banyak diterapkan pada berbagai bidang seperti ekonomi, sika, demogra, biomedis, dan lain-lain. Dalam pengolahan data dengan teori statistika seringkali asumsi yang digunakan berhubungan dengan contoh pengamatan yang saling bebas. Namun pada kenyataannya pengamatan data deret waktu sering tidak saling bebas, sehingga diperlukan pemilihan teknik yang sesuai untuk pengolahan data deret waktu tersebut. Pada sekumpulan data deret waktu, kadang-kadang terdapat data yang nilainya jauh berbeda dari data lainnya dan tidak mencerminkan karakteristik data secara umum. Nilai ini dinamakan dengan outlier. Pada tahun 1960 Gumbel mende\_nisikan bahwa outlier adalah nilai yang tampak terlalu besar atau terlalu kecil dibandingkan dengan keseluruhan pengamatan. Grubbs (1969) mende\_nisikan bahwa pengamatan outlier adalah salah satu pengamatan yang tampak menyimpang tajam dari pengamatan-pengamatan lainnya. Pada pengolahan data deret waktu, Gounder et.al (2007) menjelaskan bahwa kehadiran outlier akan mempengaruhi analisis data seperti identifikasi model, estimasi parameter dan peramalan. Untuk mendeteksi kehadiran outlier tersebut Fox (1972) memperkenalkan metode untuk deteksi outlier padadata deret waktu yang dikenal dengan outlier tipe 1 atau Additive Outliers (AO) dan tipe 2 atau Innovation Outliers (IO).

Kajian tentang deteksi outlier menjadi sangat penting karena kehadiran outlier bisa menyebabkan estimasi parameter model menjadi tidak tepat. Pengisolasian outlier dapat memberikan model yang lebih bagus. Jika data yang mengandung outlier tidak diperlakukan secara benar maka dapat menghasilkan model yang tidak mencerminkan data yang sesungguhnya. Autoregressive Moving Average (ARMA) adalah salah satu model pada data deret waktu. Pendeteksian outlier pada data deret waktu untuk model ARMA merupakan masalah yang menarik untuk diteliti karena kemampuannya dalam memberikan model data deret waktu untuk permasalahan aktual yang lebih tepat. Oleh karena itu, dalam penelitian ini akan dibahas proses pendeteksian outlier dengan menggunakan model ARMA dan penggunaannya pada data inasi kota Padang.

#### 2.1.1. Additive Outlier (AO)

Additive Outlier adalah kejadian yang mempunyai efek pada data deret waktu hanya pada satu periode saja. Dengan menggunakan persamaan (1.3) bentuk umum sebuah Additive Outliers (AO) dalam proses ARMA diuraikan sebagai berikut

$$I_t^{(T)} = \begin{cases} 1, & t = T \\ 0, & t \neq T \end{cases}$$

$$Z_t = X_t + \omega I_t^{(T)}$$

$$Z_t = \frac{\theta(B)}{\phi(B)} a_t + \omega I_t^{(T)}$$

$\{X_t\}$  adalah deret pengamatan yang tidak mengandung outlier dan  $\{Z_t\}$  adalah deret pengamatan dengan outlier.  $\omega$  adalah bobot dari AO dan IT t adalah se-buah variabel waktu yang digunakan untuk menunjukkan terjadinya sebuah AO, dimana

$$I_t^{(T)} = \begin{cases} 1, & t = T \\ 0, & t \neq T \end{cases}$$

#### 2.1.2 Innovation Outliers (IO)

Innovation Outliers adalah kejadian yang efeknya mengikuti proses ARMA. Dengan menggunakan (1.3) bentuk umum sebuah Innovation Outliers didefinisikan sebagai berikut

$$Z_t = X_t + \omega I_t^{(T)}$$

$$Z_t = \frac{\theta(B)}{\phi(B)} a_t + \frac{\theta(B)}{\phi(B)} \omega I_t^{(T)} = \frac{\theta(B)}{\phi(B)} (a_t + \omega I_t^{(T)})$$

### 2.1.3 Level Shift (LS)

Merupakan modifikasi intervensi dan pencilan yang mempengaruhi hingga akhir pengamatan.

$$Z_t = \frac{\theta(B)}{\phi(B)} a_t + \frac{1}{(1-B)} \omega_L I_t^{(T)}(t)$$

### 2.1.4 Transient Change (TC),

Merupakan modifikasi antara pencilan dan intervensi, pengaruhnya turun secara eksponensial.

$$Z_t = \frac{\theta(B)}{\phi(B)} a_t + \frac{1}{(1-\delta B)} \omega_T I_t^{(T)}$$

*Outlier* adalah pengamatan yang secara jelas berbeda dengan pengamatan lainnya. Dalam kasus runtun waktu, *outlier* diklasifikasikan menjadi *Additive Outlier* (AO), *Innovative Outlier* (IO), *Level Shift* (LS) dan *Transitory Change* (TC). *Additive Outlier* (AO) hanya berpengaruh pada pengamatan ke-T, sedangkan tiga jenis *outlier* lainnya yaitu *Innovative Outlier* (IO), *Level Shift* (LS) dan *Transitory Change* (TC) berpengaruh pada pengamatan ke-T, T+ , .... . Menurut Wei , secara umum model dengan outlier ditulis sebagai:

$$Z_t = \sum_{j=1}^k \bar{\omega}_j v_j(B) I_j^{T_j} + \frac{\theta(B)}{\phi(B)} a_t$$

dengan

$$v_j(B) = 1 \text{ untuk AO}$$

$$v_j(B) = \frac{\theta(B)}{\phi(B)} \text{ untuk IO}$$

$$v_j(B) = \frac{1}{1-B} \text{ untuk LS}$$

$v_j(B) = \frac{1}{(1-\delta B)}$  ; untuk TC, dan pada TC nilai  $\delta$  yang sering digunakan adalah 0.7. Salah satu cara untuk penanganan outlier adalah dengan cara memasukkan pangamatan outlier ke dalam model.

### BAB III

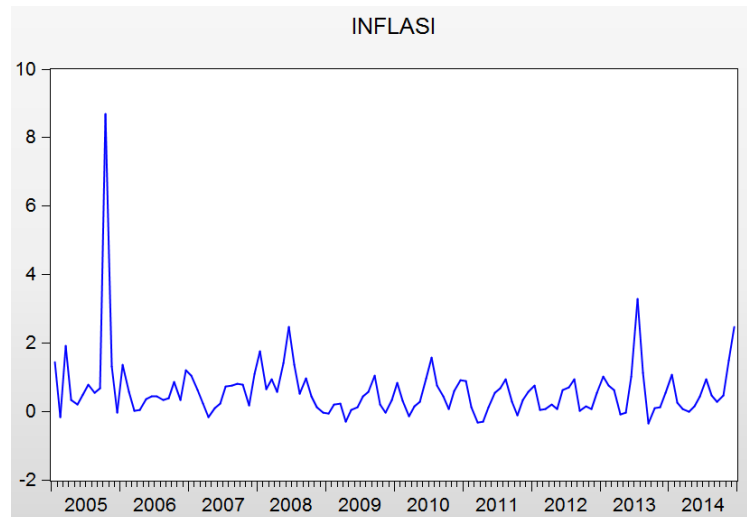
#### ANALISIS DAN PEMBAHASAN

#### 3.1 Pemodelan ARIMA

Data yang digunakan dalam makalah ini adalah data inflasi bulanan Indonesia pada bulan Januari 2005 sampai Desember 2014 yang diperoleh dari [www.bps.go.id](http://www.bps.go.id).

1. Identifikasi data deret waktu

Berikut merupakan plot data inflasi bulanan Indonesia pada bulan Januari 2005 sampai Desember 2014



Gambar 1.1

Berdasarkan gambar 1.1, terdapat pengamtan yang jauh menyimpang yaitu inflasi pada bulan oktober tahun 2005 sekitar 8.7 %, sedangkan pada bulan lainnya cenderung konstan. Hal ini terjadi karena pada bulan oktober 2005 pemerintah kembali menaikkan harga BBM sebesar 87,5% dari Rp2.400/liter menjadi Rp4.500 per liter. Penyebab inflasi tertinggi kedua selanjutnya adalah biaya angkutan atau transportasi untuk dalam kota dan luar kota yang menyumbang 2,08 persen yang keduanya menyumbang inflasi 5,6 persen. Tingginya inflasi selain BBM juga dipengaruhi oleh kondisi menjelang Lebaran.

2. Identifikasi model

Berdasarkan gambar 1.1 dapat dilihat bahwa data inflasi tidak mengalami trend naik maupun turun. Sehingga dapat disimpulkan bahwa data inflasi stasioner terhadap rata-rata. Pengujian stasioneritas data dapat dilakukan menggunakan uji ADF.dengan hipotesis

$H_0$ : Data inflasi tidak stasioner

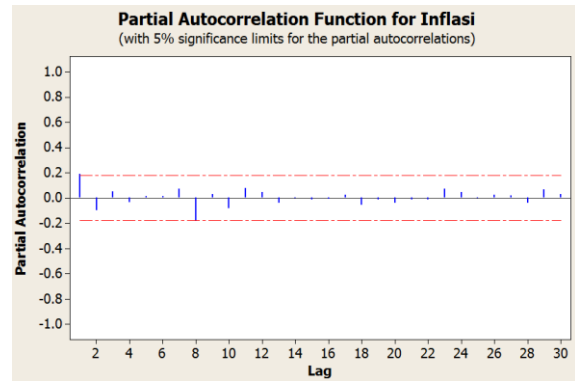
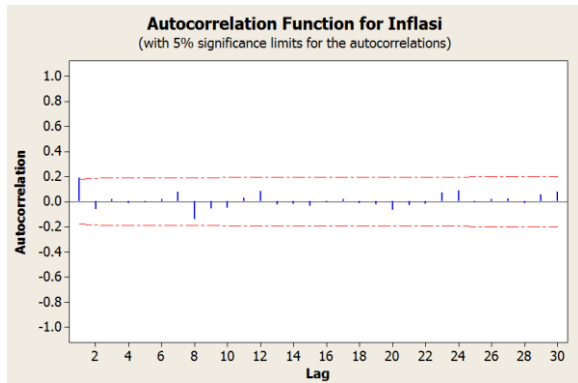
$H_1$ : Data inflasi stasioner

Tabel 1.1 Uji ADF

		t-Statistics	Prob*
Augmented Dickey Fuller test Statistics		-7.354281	0.0000
Nilai titik kritis	Taraf 1%	-3.486551	
	Taraf 5%	-2.886074	
	Taraf 10%	-2.579931	
*MacKinnon (1996) one-side p-value			

Berdasarkan uji ADF diperoleh nilai  $p(0.0000)$  lebih dari  $\alpha(0.05)$  sehingga tolak  $H_0$  dan dapat disimpulkan bahwa data inflasi sudah stasioner terhadap rata-rata.

Berikut merupakan identifikasi model ARIMA berdasarkan ACF dan PACF. Berdasarkan plot ACF, lag yang keluar batas yaitu pada lag 1 (cut off pada lag 1) sehingga orde  $AR = 1$ . Sedangkan pada PACF, lag yang keluar batas pada lag 1 (cut off pada lag 1) sehingga model ARIMA yang terbentuk adalah ARMA(1,1) dengan model tentatif ARMA(1,1), AR(1), dan MA(1).



### 3. Model Terbaik

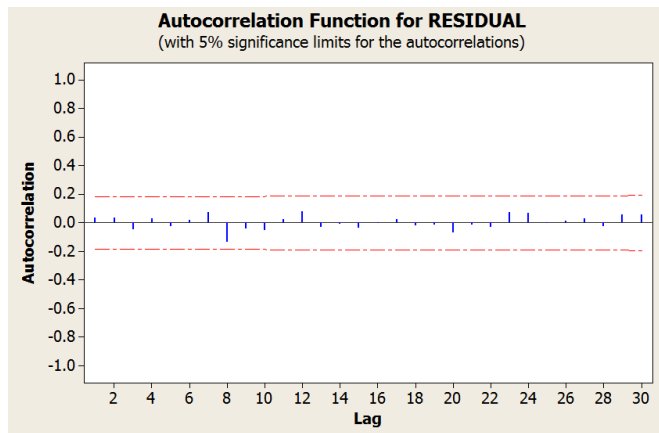
Pemilihan model terbaik dapat dilihat berdasarkan nilai MSE yang terkecil. Pada model ARMA(1,1) semua parameter signifikan dan mempunyai MSE terkecil sehingga dapat disimpulkan bahwa ARMA(1,1) adalah model terbaik dalam memodelkan data inflasi bulanan Indonesia.

Model	Koefisien	Prob	MSE
ARMA(1,1)	$\phi = -0.1743$	0.001	0.859
	$\theta = -0.8192$	0.000	
	$\theta_0 = 1.0009$	0.000	
AR(1)	$\phi = 0.1956$	0.035	0.873
	$\theta_0 = 0.499$	0.000	
MA(1)	$\theta = -0.241$	0.009	0.865
	$\theta_0 = 0.621$	0.000	

### 4. Uji Diagnostik

#### ➤ ACF residual dan Statistik uji Ljung Box

Berdasarkan plot ACF Residual, dapat dilihat bahwa tidak terdapat lag yang keluar batas signifikansi. Sehingga dapat dikatakan bahwa tidak terdapat korelasi antar sisaan white noise).



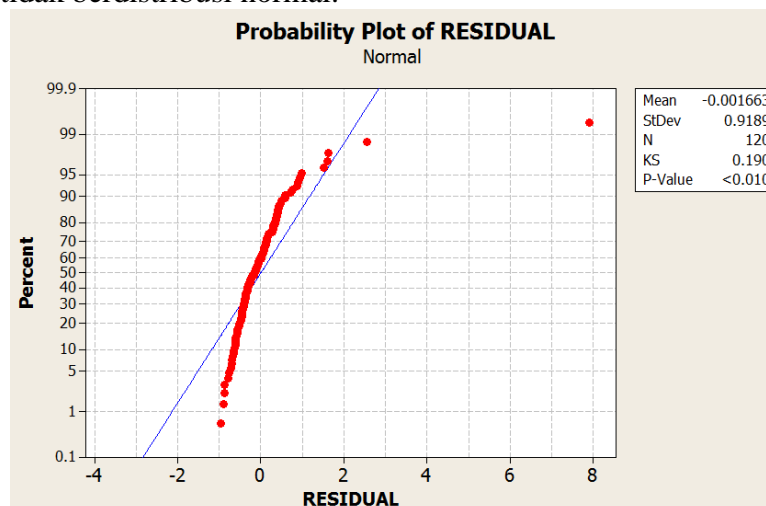
Berdasarkan hasil uji korelasi residual dengan statistik uji Ljung-Box pada lag ke-12, 24, 36 dan 48 memiliki p value lebih dari 0.05 yang artinya menerima  $H_0$ . Dengan demikian dapat disimpulkan model tersebut tidak mengandung korelasi residual antar lag sehingga model dapat dikatakan model sesuai.

Modified Box-Pierce (Ljung-Box) Chi-Square statistic

Lag	12	24	36	48
Chi-Square	5.3	8.0	14.7	21.2
DF	9	21	33	45
P-Value	0.810	0.995	0.997	0.999

➤ Uji Residual Berdistribusi Normal

Uji residual berdistribusi normal dalam penelitian ini menggunakan Kolmogorov-Smirnov dengan signifikansi  $\alpha=5\%$ . Didapatkan nilai Pvalue  $< 0,001$  hal tersebut menunjukkan keputusan menolak  $H_0$ , artinya residual model tidak berdistribusi normal. Hal ini juga dapat dilihat dari P-P plot yang menunjukkan bahwa terdapat satu amatan yang menjauh dari garis vertical yang dapat disimpulkan bahwa residual tidak berdistribusi normal.



### 3.2 Deteksi Outlier

Karena residual tidak berdistribusi normal, maka diperlukan upaya khusus untuk mengatasinya, salah satunya yakni dengan mencari penyebab ketidaknormalan data seperti keberadaan outlier. Oleh karena itu langkah selanjutnya yaitu mendeteksi keberadaan outlier pada data inflasi bulanan Indonesia. Berikut merupakan pendeteksian outlier menggunakan metode iterative dengan bantuan software SPSS.

Outliers				Estimate	SE	t	Sig.
Inflasi-Model_1	3	Additive		1,685	,414	4,067	,000
	10	Innovational		8,181	,422	19,373	,000
	42	Innovational		1,783	,424	4,200	,000
	103	Innovational		2,701	,423	6,385	,000
	119	Transient	Magnitude	1,531	,422	3,626	,000
			Decay factor	,997	,353	2,826	,006

Dari hasil deteksi outlier ARIMA dengan menggunakan tingkat signifikansi 0,05 ternyata diketahui bahwa terdapat lima outlier yang ditemukan yaitu pada pengamatan ke 3, 10, 42, 103 dan 119. Pengamatan ke 3 (Maret 2005) dengan tipe additive. Artinya bahwa efek yang ditimbulkan oleh outlier tersebut hanya terjadi pada satu periode tersebut, atau dapat ditulis sebagai berikut

$$I_t^{(T)} = \begin{cases} 1, & t = 3 \\ 0, & t \neq 3 \end{cases}$$

$$Z_t = \begin{cases} X_t, & t \neq 3 \\ X_t + \omega, & t = 3 \end{cases}$$

$$Z_t = X_t + \omega I_t^{(T)}$$

$$Z_t = \frac{\theta(B)}{\phi(B)} a_t + \omega I_t^{(T)} = \frac{1 - \theta B}{1 - \phi B} a_t + 1.685 I_t^{(3)}$$

Pada pengamatan ke 10 (Oktober 2005), 42 (Juni 2008) dan 103 (Agustus 2013) memiliki tipe innovational outlier (IO). merupakan kejadian dengan efek yang didasarkan pada proses dari model ARIMA dan mempengaruhi semua nilai pengamatan setelah kejadian tersebut muncul. Sehingga mempengaruhi inflasi setelah periode 103 (Agustus 2013) dapat ditulis sebagai berikut

$$Z_t = X_t + \omega I_t^{(T)}$$

$$Z_t = \frac{\theta(B)}{\phi(B)} a_t + \frac{\theta(B)}{\phi(B)} \omega I_t^{(T)} = \frac{\theta(B)}{\phi(B)} (a_t + \omega I_t^{(T)})$$

$$Z_t = \frac{1 - \theta B}{1 - \phi B} (a_t + 8.18 I_t^{(10)} + 1.78 I_t^{(42)} + 2.70 I_t^{(103)})$$



Pada pengamatan ke 119 ada;ah outlier tipe Transient Change (TC), merupakan modifikasi antara pencilan dan intervensi, pengaruhnya turun secara eksponensial. Dapat ditulis sebagai

$$Z_t = X_t + \omega I_t^{(T)}$$

$$Z_t = \frac{\theta(B)}{\phi(B)} a_t + \frac{1}{(1 - \delta B)} \omega_T I_t^{(T)} = \frac{1 - \theta B}{1 - \phi B} a_t + \frac{1}{(1 - 0.997B)} 1.53 I_t^{(119)}$$

Jadi , model outlier yang didapatkan adalah

$$Z_t = 1.0009 + \frac{1+0.8192B}{1+0.1743B} a_t + 1.685 I_t^{(3)} + \frac{1+0.8192B}{1+0.1743B} \left( 8.18 I_t^{(10)} + 1.78 I_t^{(42)} + 2.70 I_t^{(103)} \right) + \frac{1}{(1-0.997B)} 1.53 I_t^{(119)}$$

## LAMPIRAN

### Lampiran 1. Uji Dicky Fuller

Null Hypothesis: INFLASI has a unit root		
Exogenous: Constant		
Lag Length: 1 (Automatic - based on SIC, maxlag=12)		
	t-Statistic	Prob.*
Augmented Dickey-Fuller test statistic	-7.354281	0.0000
Test critical values: 1% level	-3.486551	
5% level	-2.886074	
10% level	-2.579931	
*MacKinnon (1996) one-sided p-values.		

### Lampiran 2. Pendugaan Parameter

#### ARMA(1,1)

Final Estimates of Parameters

Type	Coef	SE Coef	T	P
AR 1	-0.6199	0.1743	-3.56	0.001
MA 1	-0.8192	0.1256	-6.52	0.000
Constant	1.0009	0.1539	6.50	0.000
Mean	0.61788	0.09503		

Number of observations: 120

Residuals: SS = 100.490 (backforecasts excluded)  
MS = 0.859 DF = 117

Modified Box-Pierce (Ljung-Box) Chi-Square statistic

Lag	12	24	36	48
Chi-Square	5.3	8.0	14.7	21.2
DF	9	21	33	45
P-Value	0.810	0.995	0.997	0.999

#### AR(1)

Final Estimates of Parameters

Type	Coef	SE Coef	T	P
AR 1	0.1956	0.0918	2.13	0.035
Constant	0.49948	0.08532	5.85	0.000
Mean	0.6210	0.1061		

Number of observations: 120

Residuals: SS = 103.024 (backforecasts excluded)  
MS = 0.873 DF = 118

Modified Box-Pierce (Ljung-Box) Chi-Square statistic

Lag	12	24	36	48
Chi-Square	7.9	10.7	17.1	22.8
DF	10	22	34	46
P-Value	0.641	0.978	0.993	0.998

Final Estimates of Parameters

Type		Coef	SE Coef	T	P
MA	1	-0.2410	0.0902	-2.67	0.009
Constant		0.6210	0.1054	5.89	0.000
Mean		0.6210	0.1054		

Number of observations: 120

Residuals: SS = 102.106 (backforecasts excluded)  
MS = 0.865 DF = 118

Modified Box-Pierce (Ljung-Box) Chi-Square statistic

Lag	12	24	36	48
Chi-Square	7.5	10.3	16.3	21.5
DF	10	22	34	46
P-Value	0.680	0.983	0.996	0.999