

Una introducción muy general a los modelos ARIMA

WILFREDO TOLEDO, Ph.D.

Taller Ofrecido el 20 de diciembre de 2013 en La Junta de
Planificación de Puerto Rico

Modelos de Predicción

Univariables

- ▶ Tasas de crecimiento
- ▶ Modelos de tendencia
- ▶ Modelos ARIMAs

Las predicciones de los modelos ARIMAs son óptimas dentro de los modelos univariables.

Métodos multivariados

- Uniecuacionales
- Modelos multi-ecuacionales estructurales
Tienen variables exógenas
¿Cómo predecirlas?
- Modelos VAR
Todas las variables endógenas

Dos tipos de datos para los análisis estadísticos:

Corte seccional: **Muestras aleatorias**. Son observaciones que provienen de algún experimento aleatorio. En el caso de muestras aleatorias simples, toda muestra de tamaño "n" tiene la misma probabilidad de ser seleccionada.

Todas la inferencia estadística clásica está basada en muestras aleatorias donde las observaciones son i.i.d. (X_i)

Series cronológicas (de tiempo): Son observaciones secuenciales ordenadas a través del tiempo (X_t o $X(t)$).

No son i.i.d. Exhiben correlación temporal.

¿Cómo es posible utilizar X_t para hacer análisis estadísticos?

Claramente no son observaciones i.i.d.

$\text{Corr}(X_t, X_{t-k} \neq 0)$

Fundamento Teórico

- ▶ Existe un proceso estocástico que genera la serie $(X(t))$.
- ▶ La serie observada $[X(t)]$ es una realización del proceso. Pudo haber sucedido otra serie con distintos valores.

Corte seccional

Universo (población) y muestra

Series de tiempo

Proceso y realización

Condiciones para utilizar análisis estadístico inferencial con datos de series de tiempo

1. Estacionariedad

- a. Definición: Un proceso estocástico es **estacionario** en sentido *estricto* o fuerte cuando la distribución de probabilidad conjunta de la serie es invariante con respecto al tiempo.

$$F(x_t, x_{t+1}, \dots, x_{t+k}) = F(x_{t+k+1}, x_{t+k+2}, \dots, x_{t+k+s})$$

Condiciones para utilizar análisis estadístico inferencial con datos de series de tiempo

B. Un proceso estocástico es **estacionario** en el sentido débil si los primeros dos momentos (promedio y varianza) son constantes a través del tiempo.

▶ $E(x(t))=U$

▶ $\text{Var}(X(t))=\text{Sigma}^2$

2. Ergodicidad

Las observaciones muy lejanas en el tiempo no están correlacionadas.

Es necesaria para poder contar con suficientes observaciones independientes para estimar los parámetros del modelo.

Transformaciones útiles

1. Si la variable no tiene un promedio constante, Box y Jenkins recomiendan usar las primeras diferencias de la serie.

(Raíces unitarias, tendencias estocásticas, tendencias deterministas: temas para otro taller)

2. Si la variabilidad aumenta a través del tiempo usar logaritmos.

3. Si suceden ambas condiciones:

$$D\text{Log}(X(t)) = \text{Log } X(t) - \text{Log}(X(t-1))$$

Algunos Procesos estocásticos

Si X es una variable estacionaria:

a. Proceso puramente aleatorio (ruido blanco)

$$X_t = \epsilon_t \quad E[\epsilon_t] = 0 \text{ y } \text{Var}(\epsilon_t) = \sigma^2$$

b. AR(1)

en general AR(p)

$$X_t = \phi X_{t-1} + \epsilon_t$$

c. MA(1)

en general MA(q)

$$X_t = \psi \epsilon_{t-1} + \epsilon_t$$

Los 2 últimos procesos pueden tener intercepto

Un proceso puede representarse de ambas formas:

AR(1) se puede representar como un MA(∞)

MA(1) se puede representar como un AR(∞)

Procesos más complejos se pueden representar como una combinación de ambos tipos de procesos:

ARMA(1,1) en general ARMA(p,q)

$$X(t) = \phi X_{t-1} + \psi \epsilon_{t-1} + \epsilon_t$$

Algunos Procesos estocásticos

ARIMA(p,d,q) d= las veces que la serie tiene que ser diferenciada para que sea estacionaria

ARIMA(1,1,1)

$$\Delta X_t = \phi \Delta X_{t-1} + \psi \epsilon_{t-1} + \epsilon_t$$

El Procedimiento Box-Jenkins, desarrollado por: BOX, G. E. P., AND G. M. JENKINS y que apareció en el libro: Time Series Analysis, Forecasting and Control, (Holden- Day, San Francisco, CA, 1970 (y la 2nda edición en 1976)), Parte de la premisa de que toda (o casi toda) serie de tiempo puede ser ajustada por un modelo ARIMA.

Procedimiento Box-Jenkins

El procedimiento de modelación Box-Jenkins consiste en:

1. Identificación: Se escoge un modelo ARIMA o más como posibles candidatos

2. Estimación

3. Diagnóstico del modelo, si el mismo es adecuado:

4. Realizar la Predicción

Identificación de los modelos

(Box y Jenkins (B-J) sugieren un mínimo de 50 observaciones para los análisis)

1. Herramientas para la identificación de los modelos

A. Funciones de auto-correlación (ACF) y auto-correlación parcial (PACF)

ACF: Correlación de $X(t)$ con sus valores pasados:

$$\rho_k = \text{Cov}(X_t, X_{t-k}) / \text{Var}(X_t)$$

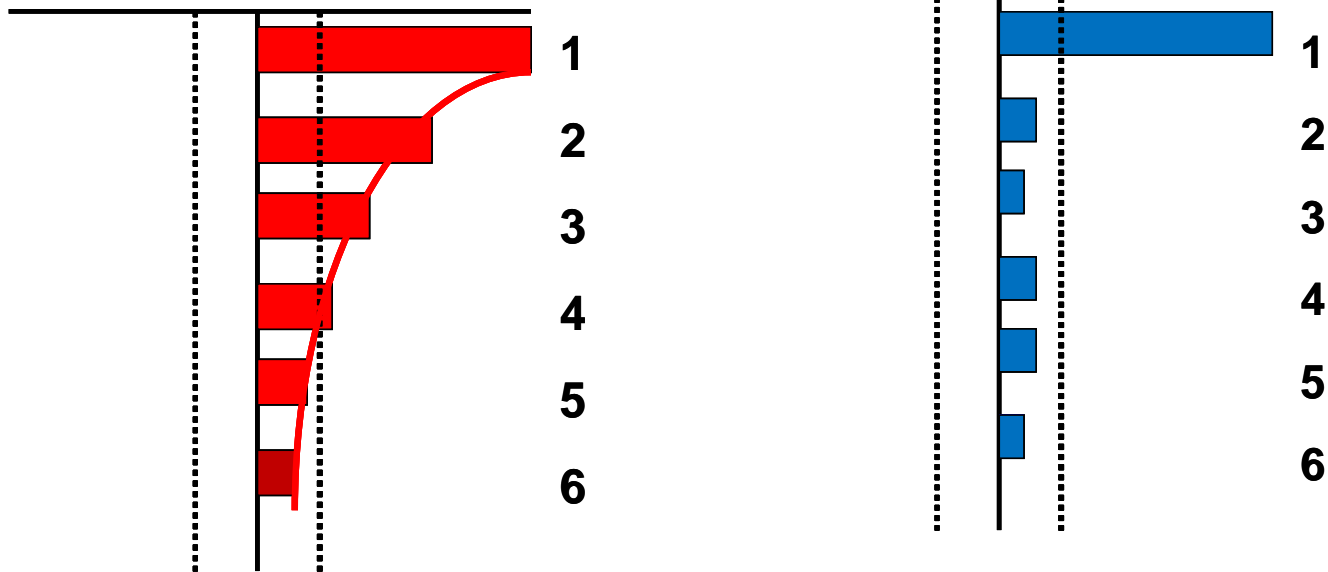
B. Funciones de auto-correlación parcial: Correlación de $X(t)$ con $X(t-K)$, descontando el efecto de la correlación entre $X(t)$ con $(X(t-(k+s)), k > 0, s > 0)$.

$$X_{t+1} = \phi_{11} X_{t-1} + \epsilon_t \quad \text{correlación parcial de orden 1}$$

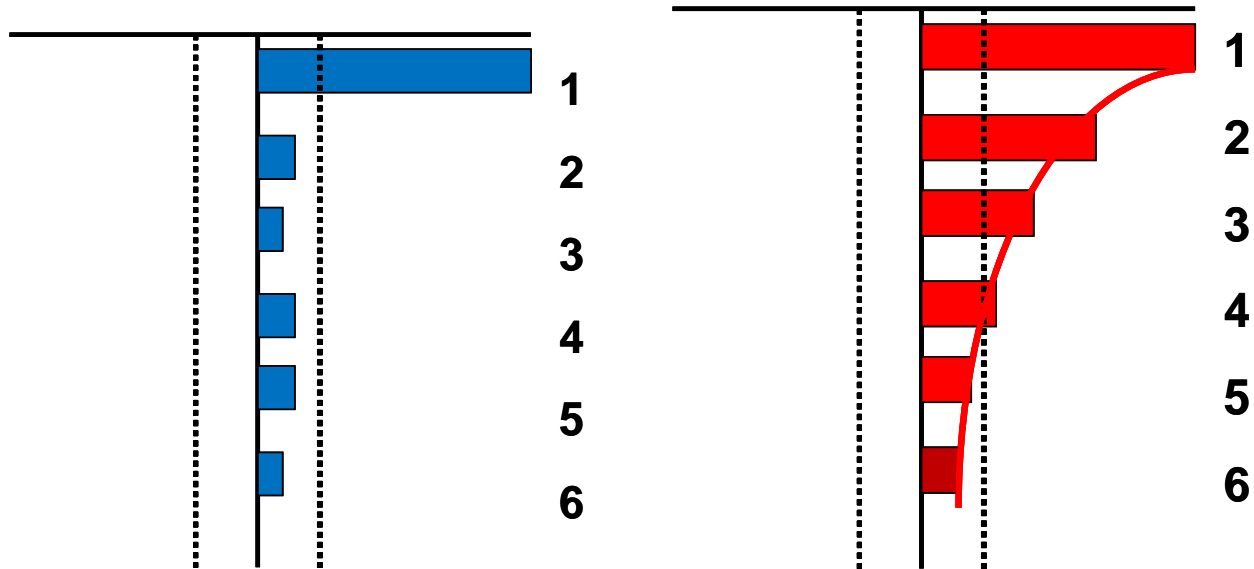
$$X_{t+2} = \phi_{21} X_{t-1} + \phi_{22} X_{t-2} + \epsilon_t \quad \text{correlación parcial de orden 2}$$

¿Cuántas correlaciones usar? B-J sugieren $T/4$.

ACF y PACF de un AR(1)



ACF Y PACF de MA(1)



Identificación de los modelos: Resumen

PROCESO	ACF	PACF
AR(p)	Decrece exponencialmente.	Hace un pico en el rezago p y se va a cero.
MA(q)	Hace un pico en el rezago q y se va a cero.	Decrece exponencialmente.
ARMA(1,1)	Decrece exponencialmente.	Decrece exponencialmente.
ARMA(p,q)	Decrece exponencialmente.	Decrece exponencialmente.

Estimación

Los modelos AR puros pueden ser estimados por MCO.

Los MAs se estiman por mínimos cuadrados no lineales (MCNL).

En general los ARIMA se estiman por MCNL

Diagnóstico:

1. Examinar estacionariedad: AR(1)

Valor absoluto de $\phi_1 < 1$

2. Examinar invertibilidad: MA(1)

Valor absoluto de $\psi_1 < 1$

Otros procesos son más complicados, pero usualmente los programas estadísticos indican si el proceso es estacionario o no.

3. Significancia estadística de los parámetros estimados

Diagnóstico:

4. Residuos deben ser un proceso puramente aleatorio
5. Coeficiente de determinación alto: ajuste del modelo
6. Estabilidad de los parámetros estimados
7. Estimar modelo en sub-periodos y evaluar predicción

Docimasia para aleatoriedad de los residuos: Estadístico-Q

$H_0: \rho_1 = \rho_2 = \dots = \rho_k = 0$

$H_a: H_0$ no rige.

$$Q_k = (T(T + 2) \sum_{k=1}^K (n - k)^{-1} \hat{\rho}_k^2) \sim \chi^2(K - M)$$

$M = \#$ de parámetros estimados

Series con componente estacional

Es necesario remover ese elemento antes del modelaje de la serie.

Diferenciación estacional

$DX_t = X_t - X_{t-s}$; $s=4$ con datos trimestrales, $s=12$ con datos mensuales.

Resumen: características de un buen modelo ARIMA

Es estacionario, si AR

Es invertible, si MA

Parámetros estimados de alta calidad (estadístico-t)

Tiene el número mínimo de parámetros: es parsimonioso

Residuos son un proceso de ruido blanco

Ajusta bien los datos (R^2)

Tiene un error de predicción relativamente bajo.

Predicciones

AR(1)

$$X_t = C + \phi X_{t-1} + \epsilon_t \quad C = \text{Constante}$$

$$X_{t+1} = C + \phi X_t + \epsilon_{t+1}$$

$$E[X_{t+1} / I_t] = C + \phi X_t + \epsilon_{t+1} = C + \phi X_t$$

I_t = conjunto de información disponible en el período t.

Regla recursiva.

Predicciones

MA(1)

$$X_{(t)} = D + \psi\epsilon_{t-1} + \epsilon_t$$

Un período hacia adelante:

$$X_{t+1} = D + \psi\epsilon_t + \epsilon_{t+1}$$

D= Constante

$$E[X_{t+1}/I_t] = D + \psi\epsilon_t$$

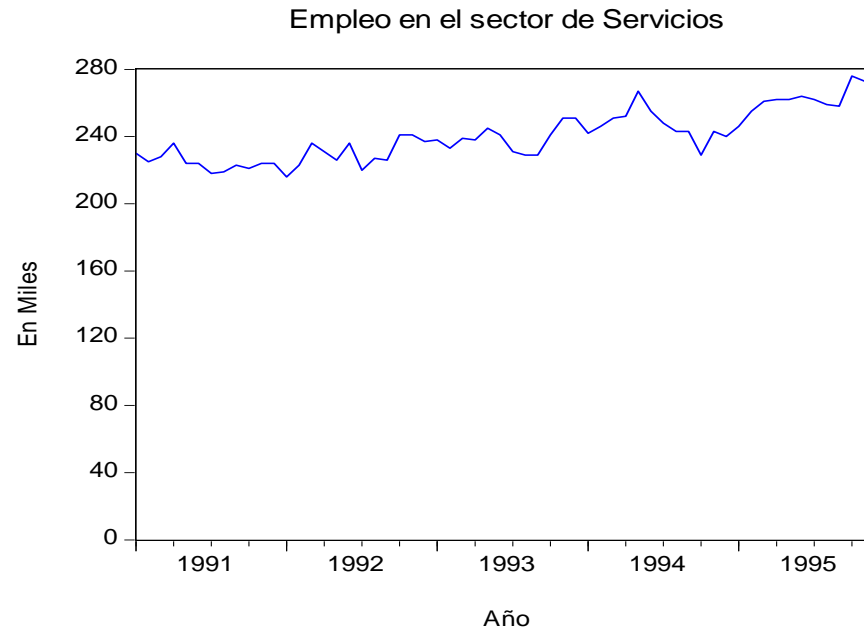
Dos períodos hacia adelante:

$$X_{t+2} = D + \psi\epsilon_{t+1} + \epsilon_{t+2}$$

$$E[X_{t+2}/I_t] = D$$

Ejemplos

Ejemplo 1: Empleo de Servicios (LSERV)



Correlograma(ACF y PACF): Niveles de la serie(LSERV)

Date: 12/01/13 Time: 10:34
Sample: 1991M01 1995M12
Included observations: 60

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
. *****	. *****	1	0.851	0.851	45.618	0.000
. *****	. .	2	0.730	0.022	79.759	0.000
. ****	. .	3	0.616	-0.034	104.55	0.000
. ****	. *	4	0.569	0.174	126.06	0.000
. ****	. .	5	0.537	0.063	145.58	0.000
. ****	. .	6	0.516	0.040	163.89	0.000
. ***	.* .	7	0.450	-0.117	178.12	0.000
. ***	.* .	8	0.359	-0.126	187.35	0.000
. **	. *	9	0.304	0.080	194.12	0.000
. **	. .	10	0.261	-0.024	199.19	0.000
. **	. *	11	0.285	0.184	205.37	0.000
. **	. .	12	0.287	-0.018	211.73	0.000
. **	. .	13	0.269	-0.044	217.46	0.000
. **	. *	14	0.250	0.096	222.52	0.000
. **	. .	15	0.229	-0.030	226.84	0.000
. **	. .	16	0.215	-0.002	230.76	0.000
. **	. .	17	0.216	0.009	234.78	0.000
. *	. .	18	0.209	-0.057	238.64	0.000
. *	.* .	19	0.145	-0.159	240.54	0.000
. .	.* .	20	0.052	-0.195	240.79	0.000
. .	. .	21	-0.017	0.019	240.82	0.000
. .	. .	22	-0.057	-0.004	241.14	0.000
. .	. *	23	-0.036	0.132	241.27	0.000
. .	.* .	24	-0.048	-0.067	241.50	0.000
.* .	.* .	25	-0.080	-0.068	242.19	0.000
.* .	. .	26	-0.136	-0.009	244.22	0.000
.* .	. .	27	-0.172	-0.010	247.56	0.000
.* .	. .	28	-0.191	-0.038	251.81	0.000

Correlograma(ACF y PACF): Primeras Diferencias de la serie (D(LSERV,1,1))

Date: 12/01/13 Time: 10:40
Sample: 1991M03 1995M12
Included observations: 58

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
**** .	**** .	1	-0.554	-0.554	18.772	0.000
. * .	** .	2	0.102	-0.297	19.420	0.000
. .	** .	3	-0.049	-0.222	19.574	0.000
. .	** .	4	-0.015	-0.217	19.589	0.001
.* .	** .	5	-0.076	-0.343	19.967	0.001
. * .	** .	6	0.107	-0.277	20.739	0.002
. * .	. .	7	0.105	0.025	21.494	0.003
* .	. .	8	-0.156	-0.038	23.187	0.003
. * .	. * .	9	0.123	0.084	24.260	0.004
** .	** .	10	-0.244	-0.225	28.570	0.001
. * .	* .	11	0.190	-0.138	31.239	0.001
. .	. * .	12	0.015	0.079	31.255	0.002
* .	* .	13	-0.093	-0.103	31.926	0.002
. * .	. * .	14	0.176	0.100	34.368	0.002
* .	. .	15	-0.168	0.003	36.665	0.001
. .	. .	16	0.004	-0.043	36.666	0.002
. .	. .	17	-0.024	-0.046	36.716	0.004
. * .	* .	18	0.093	-0.104	37.471	0.005
. .	. .	19	-0.009	0.038	37.479	0.007
. .	. .	20	0.027	0.032	37.544	0.010
. .	. * .	21	0.016	0.146	37.569	0.014
** .	* .	22	-0.221	-0.105	42.281	0.006
. * .	* .	23	0.209	-0.066	46.605	0.003
* .	. * .	24	-0.083	0.083	47.302	0.003

Ejemplo 1: Estimación (LSERV)

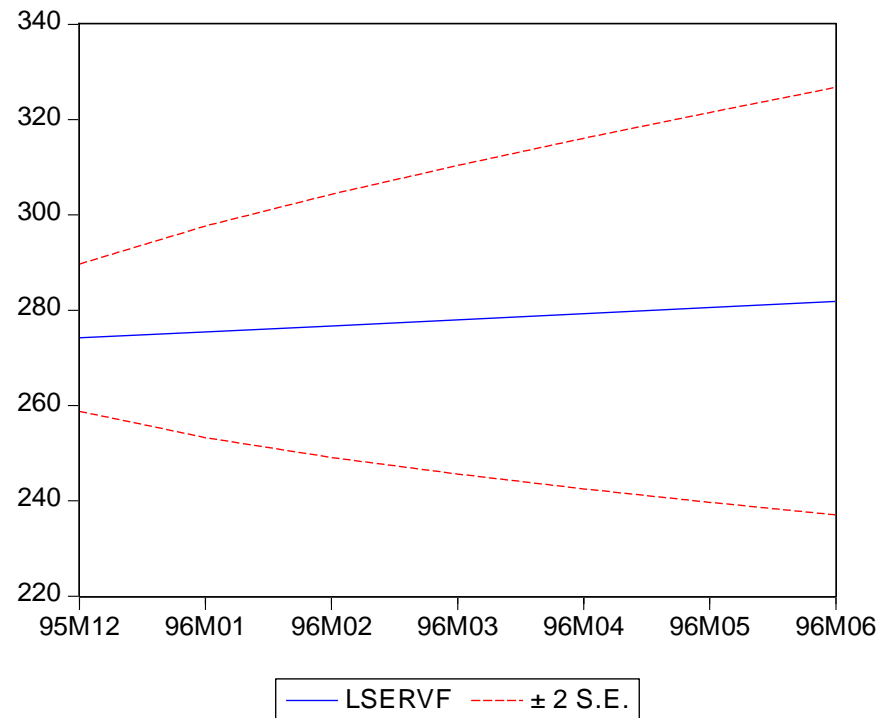
Dependent Variable: D(LSERV,1,1)				
Method: Least Squares				
Date: 12/01/13 Time: 10:42				
Sample (adjusted): 1991M03 1995M12				
Included observations: 58 after adjustments				
Convergence achieved after 16 iterations				
MA Backcast: 1991M02				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	0.016581	0.070250	0.236025	0.8143
MA(1)	-0.968411	0.025897	-37.39415	0.0000
R-squared	0.565116	Mean dependent var		-0.017241
Adjusted R-squared	0.557350	S.D. dependent var		11.60836
S.E. of regression	7.723269	Akaike info criterion		6.960227
Sum squared resid	3340.337	Schwarz criterion		7.031276
Log likelihood	-199.8466	Hannan-Quinn criter.		6.987902
F-statistic	72.76994	Durbin-Watson stat		2.353432
Prob(F-statistic)	0.000000			
Inverted MA Roots	.97			

Correlograma de los residuos (LSERV)

Date: 12/01/13 Time: 10:43
Sample: 1991M03 1995M12
Included observations: 58
Q-statistic probabilities adjusted for 1 ARMA term(s)

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
. * .	. * .	1	-0.197	-0.197	2.3627	
. .	. * .	2	-0.053	-0.096	2.5384	0.111
. * .	. * .	3	-0.132	-0.170	3.6354	0.162
. * .	** .	4	-0.120	-0.207	4.5676	0.206
. * .	** .	5	-0.086	-0.216	5.0564	0.282
. *	. .	6	0.141	-0.000	6.3932	0.270
. *	. *	7	0.126	0.089	7.4736	0.279
. * .	. * .	8	-0.154	-0.173	9.1312	0.243
. .	. * .	9	-0.052	-0.152	9.3244	0.316
** .	** .	10	-0.236	-0.343	13.373	0.146
. *	. .	11	0.159	-0.042	15.239	0.124
. *	. .	12	0.114	0.018	16.218	0.133
. .	. * .	13	0.020	-0.146	16.250	0.180
. *	. *	14	0.165	0.091	18.399	0.143
. * .	. * .	15	-0.177	-0.125	20.948	0.103
. .	. .	16	-0.063	-0.045	21.275	0.128
. .	. .	17	-0.007	-0.011	21.279	0.168
. *	. .	18	0.134	-0.003	22.845	0.154
. .	. .	19	0.029	0.068	22.921	0.194
. .	. .	20	0.033	0.001	23.022	0.236
. * .	. .	21	-0.080	0.022	23.631	0.259
** .	. * .	22	-0.229	-0.141	28.709	0.121
. *	. *	23	0.155	0.092	31.090	0.094
. .	. *	24	0.001	0.094	31.090	0.121

Predicción: LSERV



Predicción: LSERV

Predicción de la Variable			
Fecha	observación	Predicción	% de error
1996M01	271	275.453	0.0164
1996M02	281	276.704	-0.0153
1996M03	273	277.972	0.0182
1996M04	264	279.257	0.0578
1996M05	276	280.558	0.0165
1996M06	275	281.8756	0.025

Ejemplo 2: Tasa de los Fondos Federales EEUU: (Fed-Fund)

Correlograma de la Serie: Primera diferencia

Date: 12/19/13 Time: 11:13

Sample: 1985Q2 2010Q4

Included observations: 103

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob	
. *****	. *****	1	0.646	0.646	44.241	0.000
. ***	. .	2	0.408	-0.015	62.088	0.000
. **	. .	3	0.284	0.045	70.824	0.000
. *	* .	4	0.112	-0.146	72.206	0.000
. .	. .	5	0.006	-0.031	72.209	0.000
. .	. .	6	-0.046	-0.023	72.444	0.000
* .	. .	7	-0.109	-0.065	73.772	0.000
** .	* .	8	-0.218	-0.172	79.165	0.000
** .	. .	9	-0.250	-0.036	86.330	0.000
** .	* .	10	-0.279	-0.095	95.353	0.000
** .	. .	11	-0.298	-0.061	105.79	0.000
** .	. .	12	-0.258	-0.021	113.68	0.000
** .	** .	13	-0.313	-0.217	125.48	0.000
** .	* .	14	-0.337	-0.105	139.27	0.000
** .	* .	15	-0.330	-0.131	152.69	0.000
** .	. .	16	-0.266	-0.009	161.47	0.000

Estimación: (Fed-Fund)

Dependent Variable: D(FEDFUNDS)				
Method: Least Squares				
Date: 12/19/13 Time: 11:10				
Sample (adjusted): 1985Q3 2010Q4				
Included observations: 102 after adjustments				
Convergence achieved after 3 iterations				
Variable	Coefficient	Std. Error	t-Statistic	Prob.
C	-0.065944	0.104593	-0.630484	0.5298
AR(1)	0.646138	0.075730	8.532156	0.0000
R-squared	0.421289	Mean dependent var		-0.075850
Adjusted R-squared	0.415501	S.D. dependent var		0.488687
S.E. of regression	0.373614	Akaike info criterion		0.888224
Sum squared resid	13.95872	Schwarz criterion		0.939694
Log likelihood	-43.29944	Hannan-Quinn criter.		0.909066
F-statistic	72.79768	Durbin-Watson stat		1.946191
Prob(F-statistic)	0.000000			
Inverted AR Roots	.65			

Correlograma de los residuos (Fed-Fund)

Date: 12/19/13 Time: 11:12
Sample: 1985Q3 2010Q4
Included observations: 102
Q-statistic probabilities
adjusted for 1 ARMA
term(s)

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
. .	. .	1	0.022	0.022	0.0522	
. .	. .	2	-0.031	-0.031	0.1513	0.697
. *	. *	3	0.103	0.105	1.2913	0.524
* .	* .	4	-0.072	-0.080	1.8602	0.602
. .	. .	5	-0.064	-0.054	2.3133	0.678
. .	. .	6	0.021	0.009	2.3613	0.797
. .	. .	7	0.027	0.039	2.4454	0.875
* .	* .	8	-0.115	-0.112	3.9459	0.786
. .	* .	9	-0.065	-0.070	4.4229	0.817
* .	* .	10	-0.070	-0.084	4.9931	0.835
* .	* .	11	-0.142	-0.118	7.3525	0.692
. .	. *	12	0.071	0.074	7.9488	0.718
* .	* .	13	-0.088	-0.116	8.8744	0.714
* .	* .	14	-0.109	-0.100	10.301	0.669
* .	* .	15	-0.115	-0.167	11.903	0.614
* .	* .	16	-0.097	-0.106	13.053	0.598
* .	* .	17	-0.070	-0.096	13.670	0.623
. .	. .	18	0.003	-0.042	13.671	0.690
. .	* .	19	-0.002	-0.099	13.671	0.750
. .	. .	20	0.046	-0.001	13.941	0.787

Serie, Valores Ajustados y Residuos (FED-FUND)

