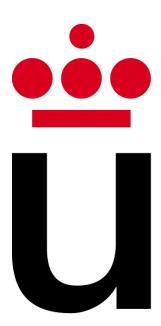
TRABAJO FINAL Análisis de Datos Avanzados

Análisis de series temporales

José Ignacio Escribano



Móstoles, 20 de marzo de 2016

Índice

Ĭn	dice de figuras	C
Ín	dice de tablas	d
1.	Introducción	1
2.	Resolución de las series temporales 2.1. Índice de empleo de un determinado país	
3.	Conclusiones	25
4.	Código EViews 4.1. Índice de empleo	26 26 26

Índice de figuras

1.	Serie temporal empleo	2
2.	Correlograma de la serie temporal empleo	3
3.	Serie transformada tomando una diferencia regular	3
4.	Correlograma de la serie temporal transformada dempleo	4
5.	Estimación del modelo ARIMA(1,1,0)	5
6.	Estimación del modelo ARIMA(1,1,2)	6
7.	Estimación del modelo ARIMA(0,1,2)	7
8.	Correlograma de los residuos del modelo ARIMA(1,1,0)	8
9.	Residuos del modelo ARIMA(1,1,0)	9
10.	Correlograma de los residuos del modelo ARIMA(0,1,2)	10
11.	Residuos del modelo ARIMA(0,1,2)	11
12.	Serie temporal puros	12
13.	Correlograma de la serie temporal puros	13
14.	Serie temporal puros	14
15.	Correlograma de la serie temporal diferenciada dpuros	15
16.	Serie temporal ddpuros12	16
17.	Correlograma de la serie temporal ddpuros12	17
18.	Estimación del modelo ARIMA $(0,1,1) \times (0,1,1)_{12} \dots \dots \dots$	18
19.	Estimación del modelo ARIMA $(2,1,0) \times (0,1,1)_{12} \dots \dots \dots$	19
20.	Residuos del modelo ARIMA $(0,1,1) \times (0,1,1)_{12} \dots \dots \dots \dots$	20
21.	Correlograma de los residuos del modelo ARIMA $(0,1,1) \times (0,1,1)_{12}$	21
22.	Residuos del modelo ARIMA $(2,1,0) \times (0,1,1)_{12} \dots \dots \dots \dots$	22
23.	Correlograma de los residuos del modelo ARIMA $(2,1,0) \times (0,1,1)_{12}$	23
24.	Predicciones (e intervalo de confianza) de ventas de puros para el año 1997	24
25.	Serie temporal de puros junto a las predicciones para el año 1997	24

Índice de tablas

1.	Principales características de la función de autocorrelación y de autoco-	
	rrelación parcial de los principales modelos estacionarios	1
2.	Análisis de la estimación de los modelos ARIMA(1,1,0) y ARIMA(0,1,2)	7
3.	Predicciones del índice de empleo para el año 1994	10
4.	Análisis de la estimación de los modelos ARIMA $(0,1,1) \times (0,1,1)_{12}$ y	
	$ARIMA(2,1,0) \times (0,1,1)_{12} \dots \dots$	19
5	Predicciones de ventas de nuros para el año 1997	22

1. Introducción

En este caso práctico utilizaremos la metodología Box-Jenkins para analizar dos series temporales. La primera es el índice de empleo de un determinado país, y la segunda es el volumen de ventas mensual de puros de una empresa tabacalera. En ambos casos, se trata de obtener un modelo que se ajuste lo máximo posible a la serie temporal.

La metodología Box-Jenkins recoge los pasos necesarios para obtener el modelo más adecuado de serie temporal:

1. Especificación inicial: consiste en determinar el orden de integración de la serie temporal y naturaleza de diferencias que se requerirán para convertir en estacionaria la serie temporal. En este paso se usa el análisis gráfico de la serie, además de los correlogramas simple y parcial de la serie. Una vez hecho lo anterior, habrá que decidir los órdenes de los polinomios autorregresivo y de medias móviles. De nuevo, se hará uso del correlograma simple y parcial de la serie. La Tabla 1 recoge las principales características de la función de autocorrelación y de autocorrelación parcial de los principales modelos estacionarios.

Tabla 1: Principales características de la función de autocorrelación y de autocorrelación parcial de los principales modelos estacionarios

Modelo	Función de autocorrelación	Función de autocorrelación parcial
AR(p)	Decrecimiento rápido hacia cero, sin llegar a anularse	p primera autocorrelaciones distintas de cero, y el resto cero
MA(q)	<i>q</i> primeras autocorrelaciones significativas, y el resto cero	Decrecimiento rápido hacia cero, sin llegar a anularse
ARMA(p,q)	Decrecimiento rápido hacia cero, sin llegar a anularse	Decrecimiento rápido hacia cero, sin llegar a anularse

- 2. Estimación: en este paso, se procede a estimar los modelos propuestos, normalmente mediante máxima verosimilitud o mínimos cuadrados no lineales.
- 3. Chequeo o validación: en este paso, se validan los posibles modelos y se escoge el que parezca más adecuado para describir la serie temporal.
- 4. Utilización del modelo: el modelo escogido se puede utilizar para predecir futuros valores de la serie, entre otras opciones.

2. Resolución de las series temporales

A continuación, aplicamos la metodología Box-Jenkins para obtener un modelo que se adecue a cada una de las series temporales planteadas.

2.1. Índice de empleo de un determinado país

La primera serie temporal es el índice de empleo de un determinado país. La serie está corregida de estacionalidad y tiene frecuencia trimestral. El período muestral abarca desde el primer trimestre del año 1962 hasta el cuarto trimestre del año 1994.

Comenzamos representando la serie temporal (Figura 1).

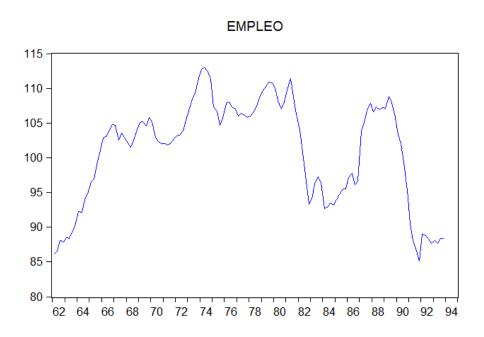


Figura 1: Serie temporal empleo

Se observa que podría haber tendencia en la serie original. Para verificarlo, usamos el correlograma de la serie que se puede ver en la Figura 2.

Como se observa un decrecimiento lento en la parte positiva del eje X, estamos ante una serie que presenta tendencia, por lo que estamos ante una serie no estacionaria. Tenemos que eliminar la tendencia, haciendo uso de las diferencias regulares de la serie original.

Tomamos la primera diferenciación para convertir la serie en estacionaria, que llamamos dempleo. Representamos la nueva serie para comprobar que hemos eliminado la tendencia de la serie original.

La nueva serie parece indicar que estamos ante una serie estacionaria, ya que hemos eliminado la tendencia tomando una diferencia regular, y la serie carecía de estacionalidad de acuerdo al enunciado. Esto se puede comprobar mirando el correlograma de esta nueva serie (Figura 4).

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
1	1	1	0.949	0.949	118.05	0.000
	<u> </u>	2	0.877	-0.244	219.60	0.000
	' ['	3	0.795	-0.100	303.65	0.000
	'E '	4	0.707	-0.070	370.75	0.000
	'E '	5	0.617	-0.066	422.20	0.000
1	(6	0.526	-0.047	459.91	0.000
ı	(7	0.437	-0.031	486.23	0.000
· 🗀	t t	8	0.351	-0.049	503.32	0.000
· 🗀		9	0.258	-0.151	512.62	0.000
' 	'E '	10	0.163	-0.071	516.36	0.000
ı j ı ı		11	0.073	-0.010	517.12	0.000
1 1	1 1 1	12	-0.005	0.016	517.13	0.000
ı (<u> </u>	13	-0.062	0.125	517.69	0.000
ı п -		14	-0.103	0.033	519.24	0.000
' -	ינם י	15	-0.125	0.069	521.55	0.000
' □ '	1 11	16	-0.135	0.016	524.24	0.000

Figura 2: Correlograma de la serie temporal empleo

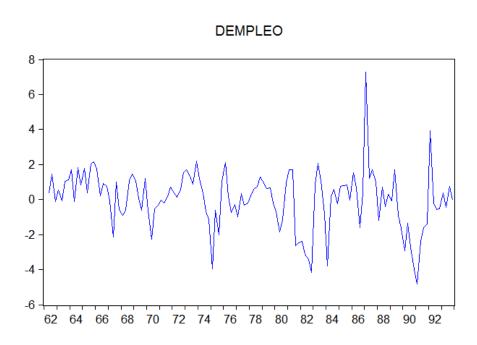


Figura 3: Serie transformada tomando una diferencia regular

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
-	·	1	0.463	0.463	27.822	0.000
· 🗀		2	0.242	0.035	35.483	0.000
ı 🗀 ı		3	0.118	-0.009	37.307	0.000
ı j ı		4	0.073	0.019	38.011	0.000
- I) I		5		-0.023	38.084	0.000
- I) I		6	0.015	0.009	38.114	0.000
- I) I	1 1	7	0.013	0.007	38.138	0.000
۱ ال ا	' <u> </u>	8	0.075	0.083	38.904	0.000
ı اا ا ا		9	0.081	0.023	39.826	0.000
	'[['	10		-0.057	39.876	0.000
' " '	'🖣 '	11	-0.093		41.099	0.000
-	" '	ı	-0.197		46.637	0.000
-	' '	ı	-0.178		51.194	0.000
-	'['		-0.191		56.505	0.000
'□ '	']'		-0.129	0.019	58.928	0.000
'□ '	'[]'		-0.124		61.183	0.000
'□ '	'🖣 '	17			64.041	0.000
<u>'</u>		ı	-0.125		66.380	0.000
' ¶ '		19	-0.094		67.710	0.000
<u> </u>	<u> </u>	20	0.040	0.177	67.955	0.000
<u>'</u>	<u>'</u>	21		-0.132	68.654	0.000
9 !	<u>"</u> ".	22	-0.165		72.912	0.000
<u> </u>	<u> </u>	23	-0.063	0.080	73.542	0.000
' '	<u> </u>	24	0.048	0.081	73.908	0.000
!] !	<u> </u>	25		-0.068	73.915	0.000
<u> </u>	'¶'	26	-0.018		73.968	0.000
<u> </u>	<u> </u>	27	0.014	0.046	73.999	0.000
. □	! ₽ !	28	0.149	0.134	77.676	0.000
. □	<u> </u>	29	0.195	0.051	84.043	0.000
¦ ₽ .		30	0.162	0.006	88.482	0.000
¦ ₽!	<u> </u>	31	0.134	0.047	91.567	0.000
' <u> </u>	<u>'</u> L'	32		-0.046	92.166	0.000
¦ ₽;		33	0.146	0.106	95.890	0.000
: P:	'.∤'.	34	0.134	0.029	99.048	0.000
' D '		35		-0.016	100.57	0.000
	' '	36	0.089	0.020	102.00	0.000

Figura 4: Correlograma de la serie temporal transformada dempleo

Así pues, tenemos que d=1 y D=0, y

$$dempleo = (1 - B)empleo$$

Es decir, dempleo es un modelo integrado de orden 1.

Observando el correlograma de la Figura 4 podemos sugerir que la serie temporal puede venir dada por un modelo AR(1), ya que la función de autocorrelación decrece rápidamente hacia cero, sin llegar a anularse, y en la función de autocorrelación parcial,

hay un valor no nulo positivo, y el resto es cero. También podría tratarse de un modelo MA(2), ya que en la función de autocorrelación hay dos valores no nulos positivos y el resto es cero, y en la función de autocorrelación parcial hay un decrecimiento rápido sin llegar a anularse. Así planteamos los siguientes modelos:

1. ARIMA(1,1,0), que escrito en forma de ecuación es

$$(1 - \theta_1 B)(1 - B)$$
empleo = a_t

2. ARIMA(0,1,2), que escrito en forma de ecuación es

$$(1 - \theta_1 B)(1 - B)$$
empleo = $(1 - \theta_1 B - \theta_2 B^2)a_t$

3. ARIMA(1,1,2), que escrito en forma de ecuación es

$$(1 - B)$$
empleo = $(1 - \theta_1 B - \theta_2 B^2)a_t$

Estimamos cada uno de los modelos propuestos. En primer lugar estimamos el modelo ARIMA(1,1,0). La salida de EViews de este modelo se muestra en la Figura 5.

Dependent Variable: D(EMPLEO,1)

Method: ARMA Maximum Likelihood (OPG - BHHH)

Date: 02/14/16 Time: 12:10 Sample: 1962Q2 1993Q4 Included observations: 127

Convergence achieved after 12 iterations

Coefficient covariance computed using outer product of gradients

Variable	Coefficient	Std. Error	Std. Error t-Statistic		
AR(1) SIGMASQ	0.459214 2.139284	0.064247 0.157305			
R-squared Adjusted R-squared S.E. of regression Sum squared resid Log likelihood Durbin-Watson stat	0.214012 0.207724 1.474284 271.6891 -228.6135 2.024943	Mean dependent var S.D. dependent var Akaike info criterion Schwarz criterion Hannan-Quinn criter.		0.017041 1.656314 3.631709 3.676500 3.649907	
Inverted AR Roots	.46				

Figura 5: Estimación del modelo ARIMA(1,1,0)

Se observa que todos los parámetros del modelo son significativos, por lo que lo consideraremos adecuado. Además, se obtiene un valor de \mathbb{R}^2 ajustado de 0.214.

Estimamos nuestro segundo modelo, es decir, el modelo ARIMA(0,1,2) usando EViews. La salida que produce este programa se puede ver en la Figura 6.

Dependent Variable: D(EMPLEO,1)

Method: ARMA Maximum Likelihood (OPG - BHHH)

Date: 02/14/16 Time: 12:28 Sample: 1962Q2 1993Q4 Included observations: 127

Convergence achieved after 12 iterations

Coefficient covariance computed using outer product of gradients

Variable	Coefficient	Std. Error	t-Statistic	Prob.
MA(1) MA(2) SIGMASQ	0.434809 0.194397 2.157131	0.075247 0.093151 0.187482	0.0000 0.0389 0.0000	
R-squared Adjusted R-squared S.E. of regression Sum squared resid Log likelihood Durbin-Watson stat	0.207454 0.194671 1.486378 273.9556 -229.1323 1.962206	Mean depend S.D. depende Akaike info cri Schwarz criter Hannan-Quin	0.017041 1.656314 3.655626 3.722812 3.682923	
Inverted MA Roots	2238i	22+.38i		

Figura 6: Estimación del modelo ARIMA(1,1,2)

Observando la salida de EViews, tenemos que todos los parámetros del modelo son significativos, por lo que consideraremos este modelo adecuado para representar la serie de tiempo del índice de empleo. Además, tiene tiene un coeficiente \mathbb{R}^2 ajustado de 0.195.

Por último, estimamos nuestro último modelo propuesto: el ARIMA(1,1,2). De nuevo, usamos EViews que nos devuelve la salida de la Figura 7.

Observando la salida de este modelo, vemos que todos los parámetros no son significativos, por lo que desechamos este modelo ya que no parece adecuado para describir la serie temporal que estamos tratando.

Sólo tenemos dos modelos que validar: el modelo ARIMA(1,1,0) y el ARIMA(0,1,2). La Tabla 2 muestra una comparativa entre la estimación de los dos modelos.

El modelo ARIMA(1,1,0) tiene un menor error (1.474) que el modelo ARIMA(0,1,2) (1.487). Además tanto los estadísticos de Akaike como de Scharwz son menores en el modelo ARIMA(1,1,0) que en el modelo ARIMA(0,1,2). Lo anterior parece indicar que el modelo más adecuado es el ARIMA(1,1,0).

Dependent Variable: D(EMPLEO,1)

Method: ARMA Maximum Likelihood (OPG - BHHH)

Date: 02/14/16 Time: 12:36 Sample: 1962Q2 1993Q4 Included observations: 127

Convergence achieved after 12 iterations

Coefficient covariance computed using outer product of gradients

Variable	Coefficient	Std. Error	t-Statistic	Prob.		
AR(1)	0.489035	0.327248 1.494388		0.1376		
MA(1)	-0.045889	0.344651	0.344651 -0.133146			
MA(2)	0.017759	0.192464	0.092271	0.9266		
SIGMASQ	2.136406	0.185092	0.0000			
R-squared	0.215069	Mean depend	0.017041			
Adjusted R-squared	0.195924	S.D. depende	nt var	1.656314		
S.E. of regression	1.485221	Akaike info cri	iterion	3.661878		
Sum squared resid	271.3235	Schwarz crite	rion	3.751459		
Log likelihood	-228.5293	Hannan-Quin	n criter.	3.698274		
Durbin-Watson stat	1.992379					
Inverted AR Roots	.49					
Inverted MA Roots	.02+.13i	.0213i				

Figura 7: Estimación del modelo ARIMA(0,1,2)

Tabla 2: Análisis de la estimación de los modelos ARIMA(1,1,0) y ARIMA(0,1,2)

	ARIMA(1,1,0)	ARIMA(0,1,2)
R^2	0.214	0.207
\mathbb{R}^2 ajustado	0.208	0.195
Akaike Info Criterion	3.631	3.656
Schwarz Criterion	3.677	3.723
Error de regresión	1.474	1.487

Para confirmar nuestras sospechas, hacemos un análisis de los residuos de ambos modelos.

Comenzamos con el modelo ARIMA(1,1,0). En la Figura 8 se puede ver el correlograma de los residuos del modelo.

Date: 02/14/16 Time: 15:49 Sample: 1962Q1 1993Q4 Included observations: 127

Q-statistic probabilities adjusted for 1 ARMA term

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
1 (1	1 11	1	-0.013	-0.013	0.0219	
1 11 1		2	0.034	0.033	0.1695	0.681
1 1	1 1	3	-0.003	-0.002	0.1704	0.918
ı j ı		4	0.030	0.029	0.2887	0.962
1 1		5	-0.015	-0.015	0.3207	0.988
1 1	1 1	6	0.002	-0.001	0.3211	0.997
1 (1	'('	7	-0.032	-0.031	0.4601	0.998
1 j j 1	וון ו	8	0.059	0.058	0.9453	0.996
۱ ال ا	יום י	9	0.071	0.075	1.6384	0.990
יוןי	ינוו	10	0.036	0.034	1.8194	0.994
1 [1	'['	ı	-0.039		2.0386	0.996
' - '	" '	ı	-0.145		5.0349	0.929
1 [1	'[['		-0.047		5.3560	0.945
'- '	'¶'	ı	-0.116		7.2940	0.886
' 	' '		-0.014		7.3241	0.921
1 🛭 1	' '		-0.034		7.4949	0.942
' 🗓 '	'🗓 '		-0.068		8.1890	0.943
' ! '	<u>'</u> <u>'</u> '		-0.056		8.6677	0.950
<u>'¶'</u>	'¶_'		-0.095		10.033	0.931
<u> </u>		20	0.156	0.184	13.758	0.798
<u> </u>	│	ı	-0.031	0.016	13.905	0.835
<u> </u>	9'		-0.178		18.852	0.595
' [<u>' [</u>	23		-0.035	18.984	0.646
! ₽ !	l '.₽'	24	0.107	0.096	20.817	0.592
!]!		25	-0.008	0.000	20.826	0.649
! ! !	! 9 !	26		-0.075	21.086	0.688
<u>: ¶ :</u>	<u>'</u> ¶:	27		-0.049	21.584	0.711
: E:	l : E:	28	0.107	0.088	23.476	0.659
: P:		29	0.119	0.088	25.837	0.582
, D ,	' '	30	0.057	0.025	26.389	0.605
	: . :	31	0.078	0.080	27.430	0.601
٦	'0''	32		-0.058 0.066	28.351	0.603
' D ' ' D '	I F	33	0.113		30.557	0.540
; , ;	' '	34	0.067	0.067	31.343	0.550
ľ	1 '1'	35	0.012	0.005	31.369	0.597
1 p 1	יקי	36	0.065	0.059	32.138	0.607

Figura 8: Correlograma de los residuos del modelo ARIMA(1,1,0)

Se puede observar que las autocorrelaciones de los residuos no son significativas y

entran dentro de las bandas de confianza, lo que indica que no son distintas de cero. De la misma forma, el estadístico Q no muestra indicios de autocorrelación de los residuos, por lo que todo parece indicar que estamos ante ruido blanco. Para comprobarlo, representamos el gráfico de residuos, que se puede ver en la Figura 9.

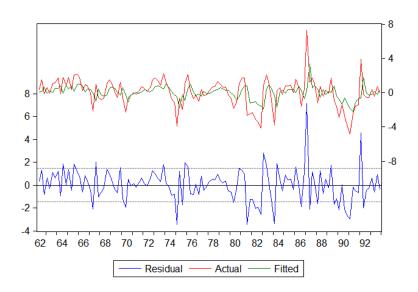


Figura 9: Residuos del modelo ARIMA(1,1,0)

La mayoría de los residuos se encuentran dentro de las bandas de confianza, lo que apoya la teoría de autocorrelación. Todo parece confirmar que los residuos se comportan como ruido blanco, es decir, tienen media 0 y varianza constante.

Notar la presencia de un outlier en el primer trimestre de 1987.

Procedemos de forma similar para comprobar que los residuos del modelo son ruido blanco.

Las Figuras 10 y 11 muestran la presencia de los residuos como ruido blanco.

De todo lo anterior, se deduce que el modelo ARIMA(1,1,0) es superior al modelo ARIMA(0,1,2) ya que tiene mejor coeficiente \mathbb{R}^2 , menor error en la estimación, y menores valores en los estadísticos de Akaike y Schwarz.

Así pues, el índice de empleo viene dado por el modelo ARIMA(1,1,0).

Utilizaremos este modelo para predecir los valores de empleo del año siguiente (1994). Los datos para cada uno de los trimestres se pueden ver en la Tabla 3.

Date: 02/14/16 Time: 15:54 Sample: 1962Q1 1993Q4 Included observations: 127

Q-statistic probabilities adjusted for 2 ARMA terms

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
- 1 1 1		1	0.018	0.018	0.0442	
1 11 1		2	0.040	0.040	0.2550	
ı <u>İ</u> n ı	<u> </u>	3	0.088	0.087	1.2866	0.257
ı j ı		4	0.031	0.027	1.4173	0.492
1 1	1 (1	5	-0.006	-0.013	1.4216	0.700
1 1	1 1	6	0.014	0.004	1.4481	0.836
1 (1		7	-0.024	-0.029	1.5299	0.910
ı j ı	<u> </u>	8	0.056	0.058	1.9640	0.923
ı j ı	<u> </u>	9	0.058	0.059	2.4333	0.932
ı j ı ı		10	0.029	0.028	2.5497	0.959
1 (1	[[11	-0.048	-0.062	2.8692	0.969
<u>п</u>	📮 '	12	-0.144	-0.163	5.8258	0.830
1 ()	III	13	-0.055	-0.058	6.2585	0.856
' = '	'📮 '		-0.127		8.5882	0.738
1 (1)		15	-0.037	0.002	8.7871	0.789
1 🛛 1	'['	16	-0.052	-0.025	9.1866	0.819
' [] '	'[['	17	-0.069	-0.053	9.8906	0.827
' []'	'[]'		-0.066		10.539	0.837
' = '	" '	19	-0.112	-0.122	12.445	0.772
' 🖻 '		20	0.141	0.189	15.470	0.629
' ('		21	-0.029	0.021	15.596	0.684
= '	['	22	-0.187	-0.160	21.044	0.395
	'[['		-0.021		21.110	0.452
יון י	יון י	24	0.097	0.091	22.620	0.423
1 1			-0.012	0.021	22.642	0.482
' ['	'¶'		-0.032		22.805	0.531
' [['	'[['	27	-0.038		23.042	0.575
' 🗗	<u> </u>	28	0.117	0.100	25.305	0.502
' 🗗 '	' <u> </u> '	29	0.112	0.087	27.386	0.443
' <u>P</u> '		30	0.069	0.035	28.185	0.455
' 📭 '	' <u> </u> '	31	0.096	0.075	29.753	0.426
'[['	'[['	32	-0.055		30.284	0.451
' P '	<u> </u>	33	0.124	0.070	32.954	0.372
<u> </u>	<u> </u>	34	0.079	0.055	34.045	0.369
<u> </u>		35	0.008	0.004	34.057	0.417
	י קי	36	0.076	0.049	35.101	0.416

Figura 10: Correlograma de los residuos del modelo ARIMA(0,1,2)

Tabla 3: Predicciones del índice de empleo para el año 1994

Trimestre	Predicción
1/1994	88.34407
2/1994	88.33593
3/1994	88.33219
4/1994	88.33048

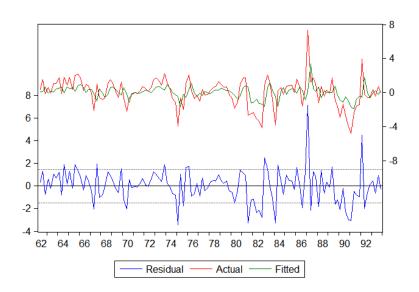


Figura 11: Residuos del modelo ARIMA(0,1,2)

Según estas predicciones, el índice de empleo del 1994 se mantuvo constante, en torno al $88\,\%.$

2.2. Venta de cigarros puros de una empresa tabacalera

La segunda serie temporal es el volumen de ventas mensual de puros de una empresa tabacalera. El período de la serie abarca desde enero de 1989 hasta diciembre de 1996.

Comenzamos representando la serie temporal (Figura 1).

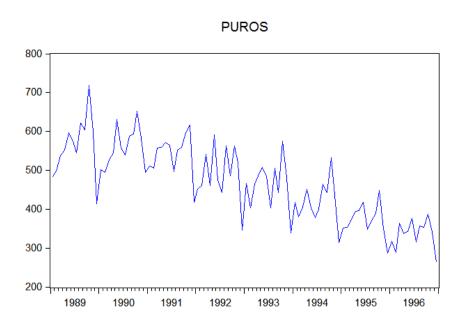


Figura 12: Serie temporal puros

Se puede ver una fuerte tendencia decreciente a lo largo de la serie temporal. Además, se observa una estacionalidad de los datos: la venta aumenta entre los meses de enero y septiembre, y desciende en los meses de octubre a diciembre.

Nos aseguramos de la presencia de tendencia y estacionalidad mirando el correlograma de la serie (Figura 13).

Se observa el decaimiento tanto en la parte regular como en la parte estacional, es decir, en los retardos múltiplos de 12, por lo que necesitamos eliminar la tendencia y la estacionalidad para que nuestra serie sea estacionaria.

Comenzamos eliminando la tendencia. Para ello, tomamos una diferencia regular de la serie, que llamaremos dpuros. La Figura 14 muestra nuestra nueva serie con una diferencia regular.

Date: 02/14/16 Time: 12:54 Sample: 1989M01 1996M12 Included observations: 96

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
1		1	0.721	0.721	51.465	0.000
		2	0.629	0.227	91.006	0.000
ı		3	0.575	0.134	124.42	0.000
ı		4	0.522	0.060	152.29	0.000
I		5	0.625	0.369	192.73	0.000
ı		6	0.587	0.025	228.77	0.000
I		7	0.568	0.061	262.85	0.000
ı	['	8	0.461	-0.188	285.53	0.000
ı	' -	9	0.448	0.121	307.21	0.000
ı		10	0.476	0.042	331.98	0.000
ı		11	0.537	0.231	363.90	0.000
		12	0.671	0.321	414.36	0.000
ı	🔲 '	13	0.469	-0.399	439.33	0.000
1	" '	14		-0.170	456.83	0.000
' 🗀	'🖣 '	15		-0.116	469.48	0.000
' 🗀	'['	16		-0.025	479.35	0.000
1	' '	17	0.393	0.023	497.74	0.000
' 🗀	'['	18		-0.082	510.79	0.000
' 🗀	ינןי	19	0.316	0.055	522.97	0.000
' 🗖	'['	20		-0.027	529.16	0.000
' 	' '	21	0.193	0.017	533.85	0.000
' 	'['	22		-0.074	539.87	0.000
' 🗀	ינןי	23	0.270	0.053	549.24	0.000
1	' '	24	0.373	0.114	567.38	0.000
' 	' '	25		-0.129	572.63	0.000
' P '	' □ '	26		-0.133	574.43	0.000
י וַן י	' '	27		-0.006	575.08	0.000
י ולן י	' '	28	0.046	0.019	575.37	0.000
' 	'['	29		-0.028	578.00	0.000
י 🏚 י	'Q''	30		-0.057	578.76	0.000
יון י		31	0.059	0.008	579.26	0.000
' ('['	ı		-0.088	579.49	0.000
' [' '		-0.044	0.124	579.78	0.000
1 1	'¶'	34	-0.016		579.82	0.000
1 [1	'['	35		-0.035	579.86	0.000
ı 🗖 ı	1 1	36	0.102	0.005	581.49	0.000

Figura 13: Correlograma de la serie temporal puros

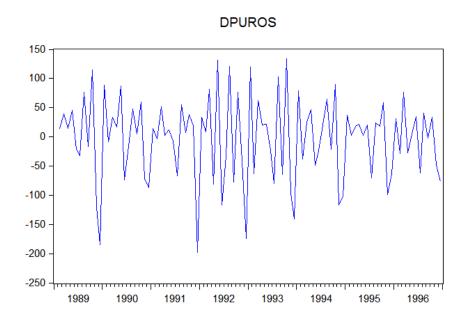


Figura 14: Serie temporal puros

Observamos que se ha eliminado la tendencia, pero no así la estacionalidad. Para asegurarnos observamos de nuevo el correlograma de esta nueva serie temporal (Figura 15).

En el correlograma se puede observar que se ha corregido la tendencia, pero no la estacionalidad en los retardos múltiplos de 12. Así que tomamos una diferencia estacional de período estacional 12. A esta nueva variable la llamamos ddpuros12, que se muestra en la Figura 16.

Se puede ver que esta nueva serie tiene tanto ausencia de tendencia como de estacionalidad. Para asegurarnos, vemos el correlograma de esta serie (Figura 17).

El correlograma no deja duda de que esta nueva serie es estacionaria.

Por tanto, para convertir la serie en estacionaria hemos tenido que aplicar la siguiente transformación:

$$ddpuros12 = (1 - B)(1 - B^{12}) \cdot puros$$

Es decir, tenemos que d = D = 1.

Debemos establecer el modelo generador de la serie.

El correlograma de la Figura 17 sugiere que el modelo puede estar generado por un

Date: 02/14/16 Time: 12:57 Sample: 1989M01 1996M12 Included observations: 95

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
	 	1	-0.333	-0.333	10.871	0.001
1 (" '	2	-0.056		11.186	0.004
1 🚺 1	'📮 '	3	-0.019		11.220	0.011
<u> </u>		4		-0.436	20.766	0.000
' 🗖	'[]'	5		-0.080	27.589	0.000
1 1	'[]'	6		-0.074	27.598	0.000
'		7	0.153	0.149	30.036	0.000
' -		8		-0.199	33.774	0.000
' 🗓 '	<u>'</u>		-0.074		34.357	0.000
' 🖺 '	'		-0.096		35.353	0.000
' -		11	-0.124		37.043	0.000
	' =	12	0.639	0.320	82.378	0.000
- '	l ' ₽ '	13	-0.213	0.168	87.492	0.000
'] '	<u>"</u> "	14	0.006	0.133	87.497	0.000
<u>'</u>				0.038	88.019	0.000
	<u>'</u> ['			-0.033	98.029	0.000
'	<u> </u>	17	0.318	0.086	109.96	0.000
<u> </u>	'¶'	18	-0.069		110.53	0.000
' -	!!!	19	0.143	0.022	112.99	0.000
<u> </u>	'¶'	20		-0.028	114.86	0.000
<u> </u>		21	-0.121	0.034	116.68	0.000
<u> </u>			-0.060	0.006	117.14	0.000
<u> </u>	'¶.'		-0.100		118.41	0.000
	l : E:	24	0.523	0.106	153.96	0.000
<u> </u>			-0.133	0.148	156.28	0.000
1 1			-0.022	0.035	156.35	0.000
'∄ :	1 : L :	28	-0.085 -0.217	-0.022 0.080	157.34 163.81	0.000
	' '	29	0.274	0.063		0.000
	l '_" '				174.31	
,	'¶	30	-0.104	-0.100	175.86 180.10	0.000
' <u>-</u> -'			-0.164		184.05	0.000
<u>'</u>					185.12	0.000
; 4 ;			-0.003	0.000	185.19	0.000
	; ;	35	-0.022	0.046	186.28	0.000
; 4 ·	; ; ;	36	0.441	0.054	216.68	0.000
	· P ·	1 30	0.441	0.054	210.00	0.000

Figura 15: Correlograma de la serie temporal diferenciada dpuros

DD12PUROS 150 100 50 -50 -100 1990 1991 1992 1993 1994 1995 1996

Figura 16: Serie temporal ddpuros12

MA(1) estacional ya que la función de autocorrelación presenta un valor no nulo en parte negativa del eje X en los múltiplos de 12 retardos, y la función de autocorrelación parcial muestra un decrecimiento lento en la parte negativa del eje X.

En la parte regular, el modelo puede estar dado por un AR(2), ya que la función de autocorrelación muestra un decrecimiento lento a lo largo de todos los retardos, y la función de autocorrelación parcial, dos valores no nulos negativos en la parte del eje X. También puede ser generado por un MA(1) ya que la función de autocorrelación presenta un valor no nulo negativo en el eje X, y la función de autocorrelación parcial presenta un decrecimiento lento en todos los retardos.

Así, los modelos candidatos a generar la serie son:

1. ARIMA $(0,1,1) \times (0,1,1)_{12}$, que escrito en forma de ecuación es

$$(1-B)(1-B^{12})$$
puros = $(1-\theta_1B)(1-\Theta_1B^{12})a_t$

2. ARIMA $(2,1,0) \times (0,1,1)_{12}$, que escrito en forma de ecuación es

$$(1 - \phi_1 B - \phi_2 B^2)(1 - B)(1 - B^{12})$$
puros = $(1 - \theta_1 B^{12})a_t$

Necesitamos estimar ambos modelos y validarlos.

Date: 02/14/16 Time: 13:02 Sample: 1989M01 1996M12 Included observations: 83

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
		1	-0.571	-0.571	28.029	0.000
ı j ı	·	2	0.028	-0.442	28.097	0.000
· 🗀 ·	' '	3	0.164	-0.120	30.470	0.000
' = '	' □ '	4	-0.166	-0.142	32.937	0.000
1 1		5	0.000	-0.231	32.937	0.000
ı 🗀 ı	'd''	6	0.134	-0.073	34.589	0.000
'	📮 '	7	-0.187	-0.197	37.823	0.000
ı 🗖 ı	' = '	8	0.111	-0.165	38.989	0.000
1 j i 1	'['	9	0.046	-0.053	39.195	0.000
' = '	' = '	10		-0.157	41.421	0.000
' 🗀		11	0.308	0.272	50.696	0.000
-	ינן י	12	-0.312	0.047	60.362	0.000
	" '	13	0.010	-0.181	60.372	0.000
' 	🗖 '	14	0.119	-0.262	61.815	0.000
1 j 1	יום י	15	0.026	0.073	61.887	0.000
'□ '	'['	16	-0.162		64.647	0.000
' 🖭 '	' □ '	17	0.153	-0.144	67.151	0.000
1 1	ינוי	18	-0.017	0.027	67.184	0.000
' 🗓 '		19	-0.070	-0.007	67.723	0.000
' P '	י 🖪 י	20	0.139	0.100	69.883	0.000
' " '	' '	21	-0.097	0.138	70.950	0.000
' [] '	' □ '	22	-0.090	-0.116	71.889	0.000
' P '	יוןי	23	0.142	0.072	74.266	0.000
'■'	' '	24	-0.103	0.011	75.547	0.000
' I II '	' '	25		-0.004	76.071	0.000
'_ I I'	'['	26		-0.037	76.426	0.000
' = _'		27	-0.159	0.011	79.612	0.000
' P '		28	0.127	0.014	81.679	0.000
1) 1	'['	29	0.012	-0.076	81.698	0.000
' [] '	' '	30	-0.054	0.155	82.085	0.000
1 1		31	-0.022	0.014	82.150	0.000
1 1	'['	32		-0.089	82.150	0.000
-	י וון י	33	0.023	0.082	82.224	0.000
· • • • • • • • • • • • • • • • • • • •	'['	34		-0.057	82.261	0.000
י ון י	' '	35	0.049	0.151	82.613	0.000
' □ '	'4'	36	-0.128	-0.075	85.086	0.000

Figura 17: Correlograma de la serie temporal ddpuros12

Comenzamos con el modelo ARIMA $(0,1,1) \times (0,1,1)_{12}$. En la Figura 18 se muestra el resultado de la estimación de este modelo en EViews.

Dependent Variable: D(PUROS,1,12)

Method: ARMA Maximum Likelihood (OPG - BHHH)

Date: 02/19/16 Time: 21:32 Sample: 1990M02 1996M12 Included observations: 83

Convergence achieved after 21 iterations

Coefficient covariance computed using outer product of gradients

Variable	Coefficient	Std. Error t-Statistic		Prob.		
MA(1) SMA(12) SIGMASQ	-0.842796 -0.616497 1045.706	0.065574 -12.85262 0.156795 -3.931867 178.3908 5.861882		0.156795 -3.931867		0.0000 0.0002 0.0000
R-squared Adjusted R-squared S.E. of regression Sum squared resid Log likelihood Durbin-Watson stat	0.620837 0.611358 32.93812 86793.61 -409.8661 2.276776	Mean depen S.D. depend Akaike info d Schwarz crit Hannan-Qui	-0.481928 52.83531 9.948582 10.03601 9.983705			
Inverted MA Roots	.96 .48+.83i 48+.83i 96	.84 .4883i 4883i		.8348i 0096i 83+.48i		

Figura 18: Estimación del modelo ARIMA $(0,1,1) \times (0,1,1)_{12}$

Se puede observar que todos los términos son significativos, y que tiene un coeficiente \mathbb{R}^2 ajustado de 0.620.

Si ahora estimamos el modelo 2, es decir, el modelo ARIMA $(2,1,0) \times (0,1,1)_{12}$, se tiene el resultado de la Figura 19.

Se puede observar que todos los coeficientes de este modelo son significativos, y que tiene un coeficiente R^2 ajustado de 0.623.

A continuación, validaremos el modelo comparando tanto los resultados de la estimación como los residuos de ambos modelos. En la Tabla 4 se puede ver una comparativa entre ambos modelos.

Method: ARMA Maximum Likelihood (OPG - BHHH)

Date: 02/19/16 Time: 21:41 Sample: 1990M02 1996M12 Included observations: 83

Convergence achieved after 22 iterations

Coefficient covariance computed using outer product of gradients

Variable	Coefficient	Std. Error t-Statistic		Prob.
AR(1)	-0.862410	0.107217	-8.043558	0.0000
AR(2)	-0.493444	0.111440	-4.427903	0.0000
MA(12)	-0.703127	0.148611	-4.731318	0.0000
SIGMASQ	1000.483	177.9938	5.620885	0.0000
R-squared	0.637235	Mean depen	-0.481928	
Adjusted R-squared	0.623459	S.D. depend	52.83531	
S.E. of regression	32.42129	Akaike info c	riterion	9.952185
Sum squared resid	83040.08	Schwarz crite	erion	10.06876
Log likelihood	-409.0157	Hannan-Quii	nn criter.	9.999017
Durbin-Watson stat	2.025038			
Inverted AR Roots	43+.55i	4355i		
Inverted MA Roots	.97	.84+.49i	.8449i	.49+.84i
	.4984i	.00+.97i	0097i	4984i
	49+.84i	8449i	84+.49i	97

Figura 19: Estimación del modelo ARIMA $(2,1,0) \times (0,1,1)_{12}$

Tabla 4: Análisis de la estimación de los modelos ARIMA $(0,1,1) \times (0,1,1)_{12}$ y ARIMA $(2,1,0) \times (0,1,1)_{12}$

	ARIMA $(2,1,0) \times (0,1,1)_{12}$	ARIMA $(0,1,1) \times (0,1,1)_{12}$
R^2	0.620	0.637
\mathbb{R}^2 ajustado	0.611	0.623
Akaike Info Criterion	9.949	9.952
Schwarz Criterion	10.036	10.068
Error de regresión	32.938	32.421

Se puede observar que el valor de los coeficientes R^2 y R^2 ajustado son mayores en el modelo ARIMA $(2,1,0)\times(0,1,1)_{12}$ que en el modelo ARIMA $(2,1,0)\times(0,1,1)_{12}$. Además, también es menor tanto el Akaike Info Criterion como el Schwarz Criterion en el modelo ARIMA $(2,1,0)\times(0,1,1)_{12}$. Sin embargo, el modelo ARIMA $(2,1,0)\times(0,1,1)_{12}$ tiene un menor error de regresión que el modelo ARIMA $(0,1,1)\times(0,1,1)_{12}$.

A continuación, pasamos a analizar los residuos de los modelos. Comenzamos con el modelo ARIMA $(0,1,1)\times(0,1,1)_{12}$. Las Figuras 20 y 21 muestra los residuos y el correlograma de estos. Se puede observar que la mayoría caen dentro de las bandas de confianza, por lo que es de suponer que los residuos se comportan como ruido blanco. El correlograma parece también contrastarlo.

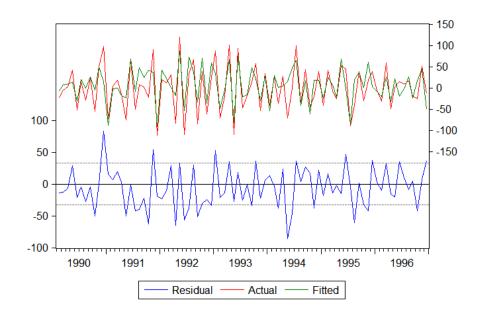


Figura 20: Residuos del modelo ARIMA $(0,1,1) \times (0,1,1)_{12}$

De forma similar, comprobamos que los residuos del modelo ARIMA $(2, 1, 0) \times (0, 1, 1)_{12}$ (Figuras 22 y 23).

A tenor de los resultados, estamos ante dos modelos muy parecidos, pero nos decantaremos por el modelo ARIMA $(0,1,1) \times (0,1,1)_{12}$, ya que es el que menor valor tiene en los Akaike Info Criterion y Schwarz Criterion, aunque tiene un mayor error de regresión (cercano a 0.5).

Por tanto, la venta de puros viene dada por el modelo ARIMA $(2,1,0) \times (0,1,1)_{12}$.

Usaremos este modelo para predecir las ventas de puros en el año 1997. Estos datos se pueden ver en la Tabla 5.

Date: 02/19/16 Time: 21:38 Sample: 1989M01 1996M12 Included observations: 83

Q-statistic probabilities adjusted for 2 ARMA terms

Autocorrelation	Partial Correlation	AC	PAC	Q-Stat	Prob
	 	1 -0.199	-0.199	3.4180	
1 1	'['	2 -0.006	-0.048	3.4214	
ı 🗀 ı		3 0.167	0.163	5.8769	0.015
-	" '	4 -0.238	-0.184	10.934	0.004
' [' □ '	l	-0.168	11.470	0.009
' 	י 🗗 י	6 0.143	0.083	13.338	0.010
– '	" '	l	-0.173	19.494	0.002
י נון י	'['	l	-0.045	19.796	0.003
' P '	' '	9 0.116	0.057	21.077	0.004
' = '	' '	10 -0.124		22.566	0.004
' 	י וון י	11 0.155	0.072	24.928	0.003
י נן י	' '	12 0.037	0.010	25.066	0.005
-	'🖣 '	13 -0.189	-0.115	28.654	0.003
' 🗗 '	' '	14 0.142	0.030	30.712	0.002
ı j ı	' '	15 0.036	0.098	30.846	0.004
' -	'[] '	16 -0.180		34.277	0.002
' 🗗	' '	17 0.166	0.026	37.226	0.001
1] 1	' '	18 0.043	0.129	37.428	0.002
' 🖣 '	' '	l	-0.023	39.305	0.002
' 🗗 '	' '	20 0.139	-0.022	41.456	0.001
' [] '	'Q '	21 -0.110	-0.073	42.826	0.001
' 二 '	' [] '		-0.107	45.757	0.001
' 	' '	23 0.208	0.100	50.847	0.000
' [] '	' '	24 -0.107	0.009	52.211	0.000
1 j 1	' '	25 0.028	0.046	52.304	0.000
' 	'['	l	-0.062	53.224	0.001
'■'	'['	l	-0.037	55.097	0.000
ı j ı	' '	28 0.038	0.012	55.280	0.001
' '	' '	l	-0.013	56.434	0.001
'□ '	' '	l	-0.018	59.369	0.000
' ['	' '	l	-0.137	59.880	0.001
' j i '	' '	32 0.075	0.009	60.665	0.001
1 1	' '	33 -0.020	0.126	60.721	0.001
' <u> </u>	' '	34 0.072	0.024	61.474	0.001
'_	'] '	35 0.110	0.039	63.265	0.001
' 二 '	'[['	36 -0.133	-0.068	65.930	0.001

Figura 21: Correlograma de los residuos del modelo ARIMA $(0,1,1)\times(0,1,1)_{12}$

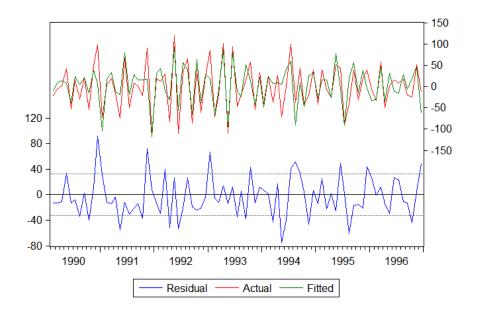


Figura 22: Residuos del modelo ARIMA $(2,1,0)\times(0,1,1)_{12}$

Tabla 5: Predicciones de ventas de puros para el año 1997

Mes	Predicción
Enero	278.693
Febrero	252.391
Marzo	328.391
Abril	301.391
Mayo	306.391
Junio	341.391
Julio	279.391
Agosto	320.391
Septiembre	318.391
Octubre	352.391
Noviembre	304.391
Diciembre	228.391

De forma más gráfica, se puede ver en la Figura 24. Si comparamos la serie temporal inicial junto con las predicciones, observamos la clara tendencia descendente que se mantiene en el año 1997 (Figura 25).

Date: 02/19/16 Time: 21:44 Sample: 1989M01 1996M12 Included observations: 83

Q-statistic probabilities adjusted for 3 ARMA terms

Autocorrelation	Partial Correlation		AC	PAC	Q-Stat	Prob
1 (1		1	-0.041	-0.041	0.1427	
1 [] 1	'[]'	2	-0.073	-0.075	0.6060	
-	🗖 '	3	-0.196	-0.204	4.0042	
' □ '	' = '	4	-0.118	-0.152	5.2551	0.022
' " '	' = '	5		-0.166	6.2220	0.045
1 j i 1	'('	6	0.077	-0.016	6.7673	0.080
' = '	🗏 '	7		-0.267	9.2409	0.055
- 	'📮 '	8		-0.119	9.2670	0.099
ı اا ا	'['	9		-0.037	9.6956	0.138
' '	' '	10		-0.115	9.7451	0.203
' P '	'] '	11	0.115	0.033	11.034	0.200
1] 1	'['	12		-0.041	11.069	0.271
' - '	│ ' □ '	13	-0.131	-0.136	12.793	0.235
' [] '	' '	14	0.070	0.051	13.289	0.275
' 	' '	15	0.015	0.006	13.312	0.347
'■_'	'¶'	16		-0.122	14.527	0.338
' P '	l ' 🖳 '	17	0.102	0.088	15.640	0.336
<u> </u>	! ₽ !	18	0.083	0.138	16.382	0.357
' [['	' ['	19	-0.061	-0.020	16.793	0.399
' j i '		20	0.025	0.035	16.865	0.464
' = '	'['	21		-0.040	18.412	0.429
'■ '	' '	22		-0.109	20.264	0.379
' P '	' '	23	0.149	0.104	22.859	0.296
1 j 1	ינוי	24	0.025	0.025	22.933	0.348
1 j i 1	ינוי	25	0.047	0.041	23.207	0.390
1 1	' '		-0.001		23.208	0.449
' [] '	' '	27		-0.005	23.698	0.479
1 j i 1	יון י	28	0.049	0.073	24.003	0.519
י 🏻 י	' '	29	0.078	0.003	24.790	0.531
'■'	'['	30	-0.128		26.979	0.465
'■ '	'🖣 '	31	-0.115		28.780	0.424
1 j i 1	' '	32	0.042	0.009	29.023	0.464
1 j i 1	י וַבּן י	33	0.052	0.075	29.399	0.497
1 j i 1	'['	34		-0.053	29.918	0.522
' 	י וַן י	35	0.099	0.061	31.353	0.499
' 🗖 '		36	-0.094	-0.013	32.677	0.483

Figura 23: Correlograma de los residuos del modelo ARIMA $(2,1,0) \times (0,1,1)_{12}$

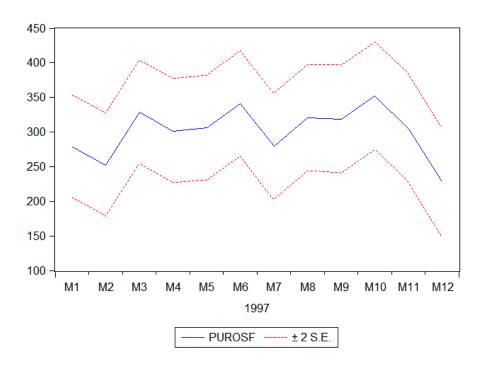


Figura 24: Predicciones (e intervalo de confianza) de ventas de puros para el año 1997

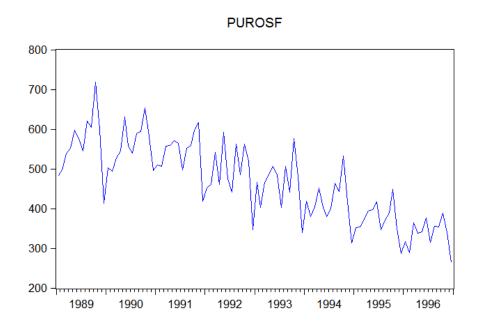


Figura 25: Serie temporal de puros junto a las predicciones para el año 1997

3. Conclusiones

En este trabajo hemos visto cómo aplicar la metodología Box-Jenkins para modelizar distintas series temporales, que trataban distintos temas socioeconómicos: la primera serie era sobre el índice de paro de un país, y la segunda, la venta de puros. En ambos casos, hemos utilizado el software EViews, que gracias a su potencia y sus modelos de series temporales ya implementados, ha permitido ahorrar mucho tiempo estimando, analizando y prediciendo nuestros modelos propuestos.

4. Código EViews

A continuación se muestra el código EViews utilizado para resolver cada una de las series temporales planteadas.

4.1. Índice de empleo

```
empleo.sheet
{%graph}.line
empleo.correl(16)
series d
series dempleo = empleo - empleo (-1)
dempleo.sheet
{%graph}.line
dempleo.correl
{%equation}.ls(optmethod=opg) d(empleo,1) ar(1)
{%equation}.resids(g)
{%equation}.correl
{%equation}.ls(optmethod=opg) d(empleo,1) ma(1) ma(2)
{%equation}.resids(g)
{%equation}.results
{%equation}.correl
{%equation}.resids(g)
dempleo.hist
{%graph}.line
{%equation}.ls(optmethod=opg) d(empleo,1) ar(1)
smpl 1994q1 1994q4
{%equation}.forecast(e, g) empleof
smpl 1962q1 1993q4
{%equation}.forecast
{%equation}.results
empleof.sheet
{%graph}.line
empleo.correl(16)
empleo.sheet
empleof.sheet
```

4.2. Venta de puros

```
puros.sheet
{%graph}.line
puros.correl
series dpuros = puros - puros(-1)
```

```
dpuros.sheet
{%graph}.line
dpuros.correl
series dd12puros = dpuros(puros,1,12)
series dd12puros = d(puros, 1, 12)
dd12puros.sheet
{%graph}.line
dd12puros.correl
{%equation}.ls(optmethod=opg) dd12puros ma(1) sma(12)
{%equation}.ls(optmethod=opg) d(puros,1,12) ma(1) sma(12)
{%equation}.correl
{%equation}.resids(g)
{%equation}.correl
{%equation}.ls(optmethod=opg) d(puros,1,12) ar(1) ar(2) ma(12)
{%equation}.resids(g)
{%equation}.correl
{%equation}.ls(optmethod=opg) d(puros,1,12) ma(1) sma(1)
dd12puros.sheet
puros.sheet
pagestruct (end=1997M12)
smpl 1997m01 1997m12
{%equation}.forecast(e, g) purosf
smpl 1989m01 1996m12
purosf.sheet
{%graph}.line
```