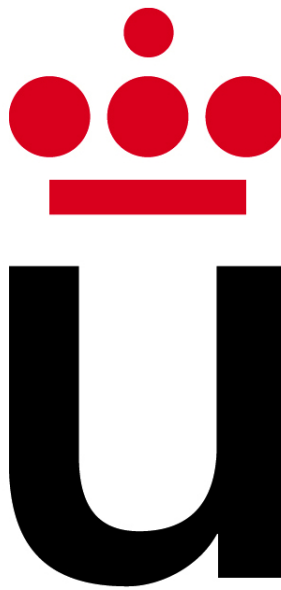


TRABAJO FINAL
Análisis de Datos Avanzados
ANÁLISIS DE SERIES TEMPORALES

José Ignacio Escribano



MÓSTOLES, 19 DE FEBRERO DE 2016

Índice

Índice de figuras	c
Índice de tablas	d
1. Introducción	1
2. Resolución de las series temporales	1
2.1. Índice de empleo de un determinado país	2
2.2. Venta de cigarros puros de una empresa tabacalera	12
3. Código EViews	18
3.1. Índice de empleo	18
3.2. Venta de puros	18
4. Conclusiones	19

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

Índice de tablas

1.	Principales características de la función de autocorrelación y de autocorrelación parcial de los principales modelos estacionarios	1
2.	Análisis de la estimación de los modelos ARIMA(1,1,0) y ARIMA(0,1,2)	6
3.	Predicciones del índice de empleo para el año 1994	9

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)	q 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.

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).

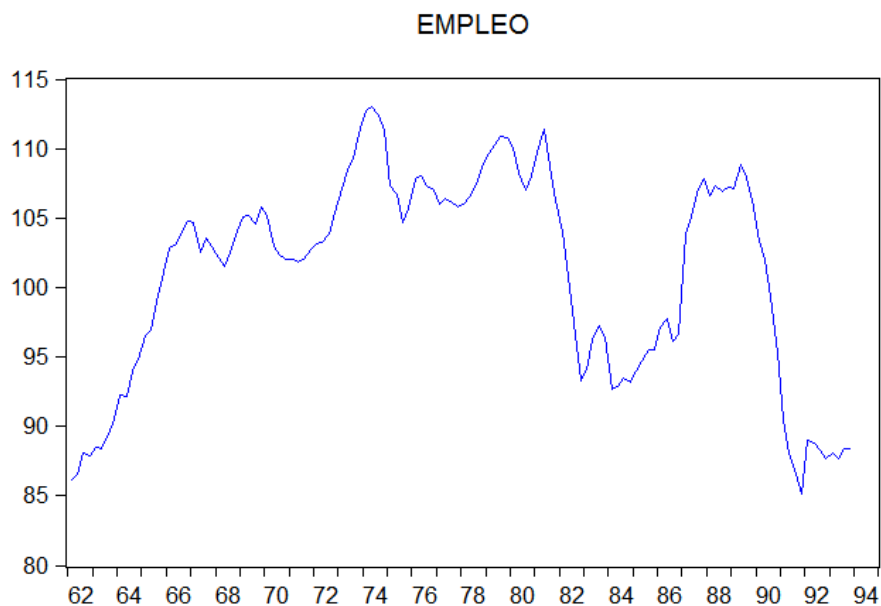


Figura 1: Serie temporal empleo

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).

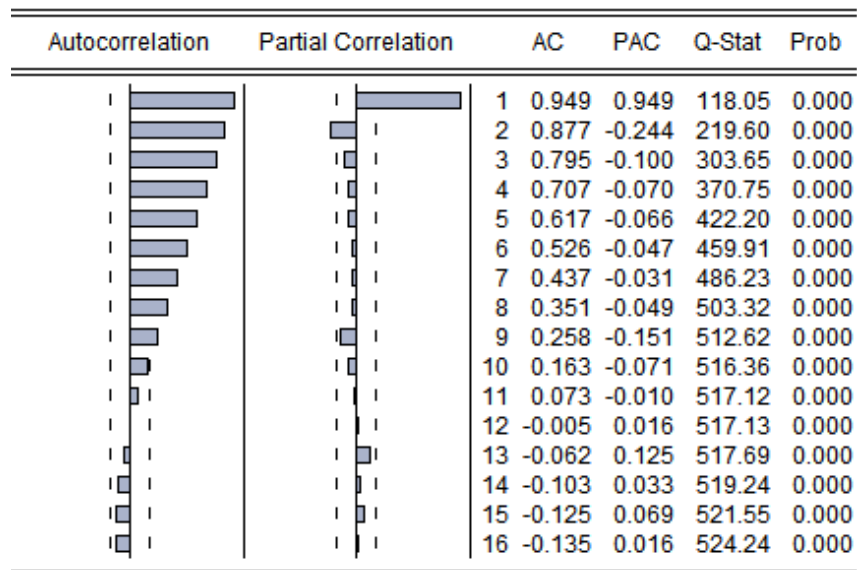


Figura 2: Correlograma de la serie temporal empleo

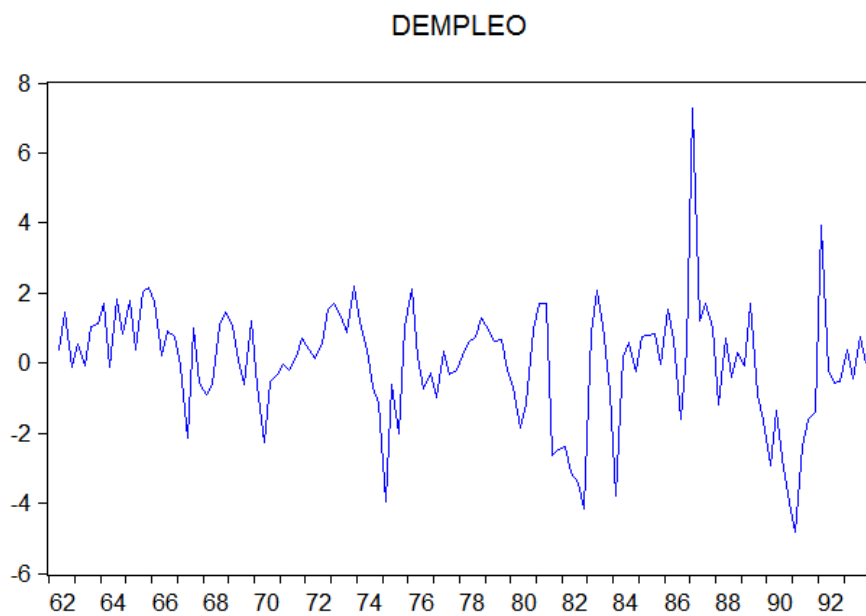


Figura 3: Serie transformada tomando una diferencia regular

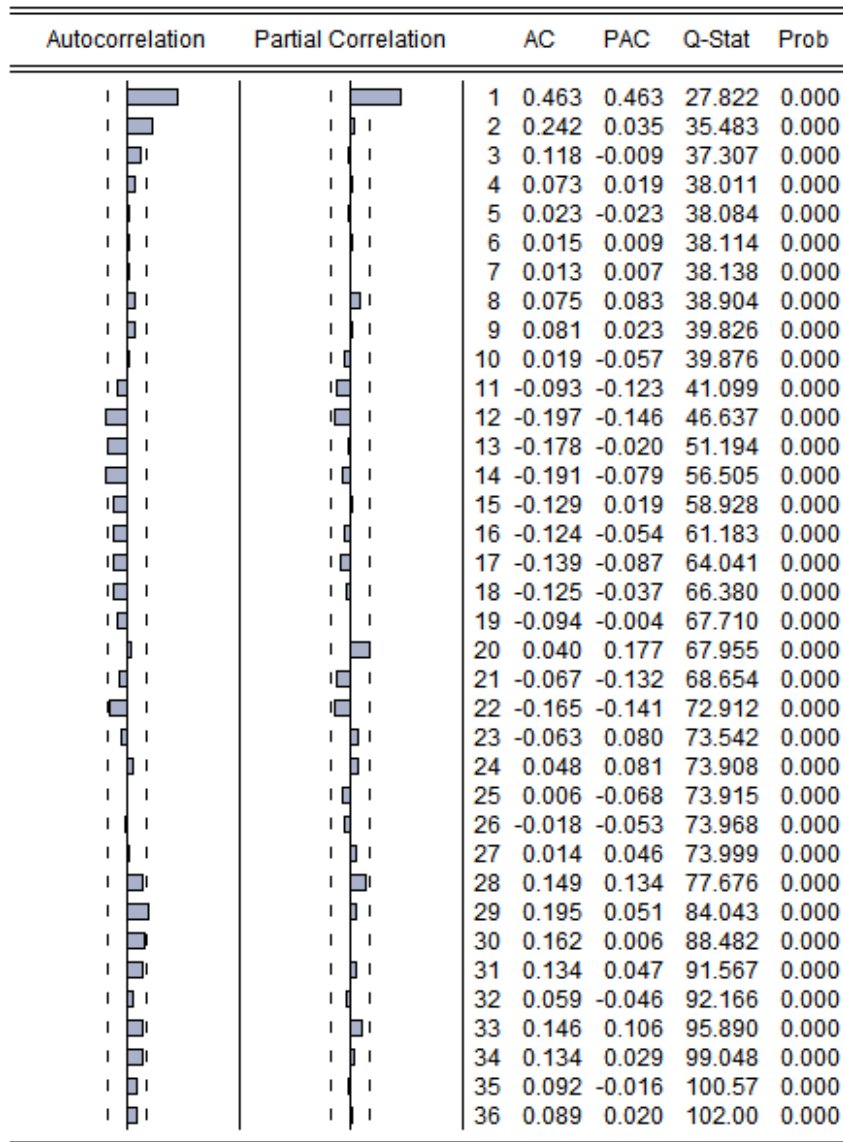


Figura 4: Correlograma de la serie temporal transformada dempleo

Así pues, tenemos que $d = 1$ y $D = 0$, y

$$\text{dempleo} = (1 - B)\text{empleo}$$

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)
2. ARIMA(0,1,2)
3. ARIMA(1,1,2)

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	t-Statistic	Prob.
AR(1)	0.459214	0.064247	7.147611	0.0000
SIGMASQ	2.139284	0.157305	13.59957	0.0000
R-squared	0.214012	Mean dependent var		0.017041
Adjusted R-squared	0.207724	S.D. dependent var		1.656314
S.E. of regression	1.474284	Akaike info criterion		3.631709
Sum squared resid	271.6891	Schwarz criterion		3.676500
Log likelihood	-228.6135	Hannan-Quinn criter.		3.649907
Durbin-Watson stat	2.024943			
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 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.

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 un coeficiente R^2 ajustado de 0.195.

Dependent Variable: D(EMPLOO,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)	0.434809	0.075247	5.778406	0.0000
MA(2)	0.194397	0.093151	2.086909	0.0389
SIGMASQ	2.157131	0.187482	11.50579	0.0000
R-squared	0.207454	Mean dependent var		0.017041
Adjusted R-squared	0.194671	S.D. dependent var		1.656314
S.E. of regression	1.486378	Akaike info criterion		3.655626
Sum squared resid	273.9556	Schwarz criterion		3.722812
Log likelihood	-229.1323	Hannan-Quinn criter.		3.682923
Durbin-Watson stat	1.962206			
Inverted MA Roots	-.22-.38i	-.22+.38i		

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

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.

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
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

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	-0.133146	0.8943
MA(2)	0.017759	0.192464	0.092271	0.9266
SIGMASQ	2.136406	0.185092	11.54242	0.0000
R-squared	0.215069	Mean dependent var		0.017041
Adjusted R-squared	0.195924	S.D. dependent var		1.656314
S.E. of regression	1.485221	Akaike info criterion		3.661878
Sum squared resid	271.3235	Schwarz criterion		3.751459
Log likelihood	-228.5293	Hannan-Quinn criter.		3.698274
Durbin-Watson stat	1.992379			
Inverted AR Roots	.49			
Inverted MA Roots	.02+.13i	.02-.13i		

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

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 Schwarz 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).

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.

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.

La mayoría de los residuos se encuentran dentro de las bandas de confianza, lo que apoya la teoría de autocorrelación. Todo parece confirmar que estamos ante ruido blanco.

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

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 -0.013	-0.013	0.0219	
		2 0.034	0.033	0.1695	0.681
		3 -0.003	-0.002	0.1704	0.918
		4 0.030	0.029	0.2887	0.962
		5 -0.015	-0.015	0.3207	0.988
		6 0.002	-0.001	0.3211	0.997
		7 -0.032	-0.031	0.4601	0.998
		8 0.059	0.058	0.9453	0.996
		9 0.071	0.075	1.6384	0.990
		10 0.036	0.034	1.8194	0.994
		11 -0.039	-0.042	2.0386	0.996
		12 -0.145	-0.155	5.0349	0.929
		13 -0.047	-0.055	5.3560	0.945
		14 -0.116	-0.112	7.2940	0.886
		15 -0.014	-0.008	7.3241	0.921
		16 -0.034	-0.018	7.4949	0.942
		17 -0.068	-0.078	8.1890	0.943
		18 -0.056	-0.071	8.6677	0.950
		19 -0.095	-0.118	10.033	0.931
		20 0.156	0.184	13.758	0.798
		21 -0.031	0.016	13.905	0.835
		22 -0.178	-0.169	18.852	0.595
		23 -0.029	-0.035	18.984	0.646
		24 0.107	0.096	20.817	0.592
		25 -0.008	0.000	20.826	0.649
		26 -0.040	-0.075	21.086	0.688
		27 -0.055	-0.049	21.584	0.711
		28 0.107	0.088	23.476	0.659
		29 0.119	0.088	25.837	0.582
		30 0.057	0.025	26.389	0.605
		31 0.078	0.080	27.430	0.601
		32 -0.073	-0.058	28.351	0.603
		33 0.113	0.066	30.557	0.540
		34 0.067	0.067	31.343	0.550
		35 0.012	0.005	31.369	0.597
		36 0.065	0.059	32.138	0.607

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

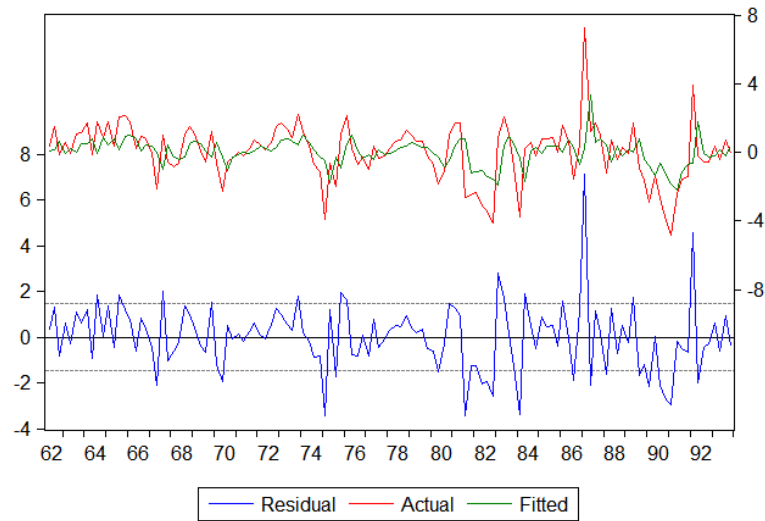


Figura 9: Residuos del modelo ARIMA(1,1,0)

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 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.

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

Según estas predicciones, el índice de empleo del 1994 se mantuvo constante, en torno

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 0.018	0.018	0.0442	
		2 0.040	0.040	0.2550	
		3 0.088	0.087	1.2866	0.257
		4 0.031	0.027	1.4173	0.492
		5 -0.006	-0.013	1.4216	0.700
		6 0.014	0.004	1.4481	0.836
		7 -0.024	-0.029	1.5299	0.910
		8 0.056	0.058	1.9640	0.923
		9 0.058	0.059	2.4333	0.932
		10 0.029	0.028	2.5497	0.959
		11 -0.048	-0.062	2.8692	0.969
		12 -0.144	-0.163	5.8258	0.830
		13 -0.055	-0.058	6.2585	0.856
		14 -0.127	-0.111	8.5882	0.738
		15 -0.037	0.002	8.7871	0.789
		16 -0.052	-0.025	9.1866	0.819
		17 -0.069	-0.053	9.8906	0.827
		18 -0.066	-0.068	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
		23 -0.021	-0.049	21.110	0.452
		24 0.097	0.091	22.620	0.423
		25 -0.012	0.021	22.642	0.482
		26 -0.032	-0.065	22.805	0.531
		27 -0.038	-0.051	23.042	0.575
		28 0.117	0.100	25.305	0.502
		29 0.112	0.087	27.386	0.443
		30 0.069	0.035	28.185	0.455
		31 0.096	0.075	29.753	0.426
		32 -0.055	-0.065	30.284	0.451
		33 0.124	0.070	32.954	0.372
		34 0.079	0.055	34.045	0.369
		35 0.008	0.004	34.057	0.417
		36 0.076	0.049	35.101	0.416

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

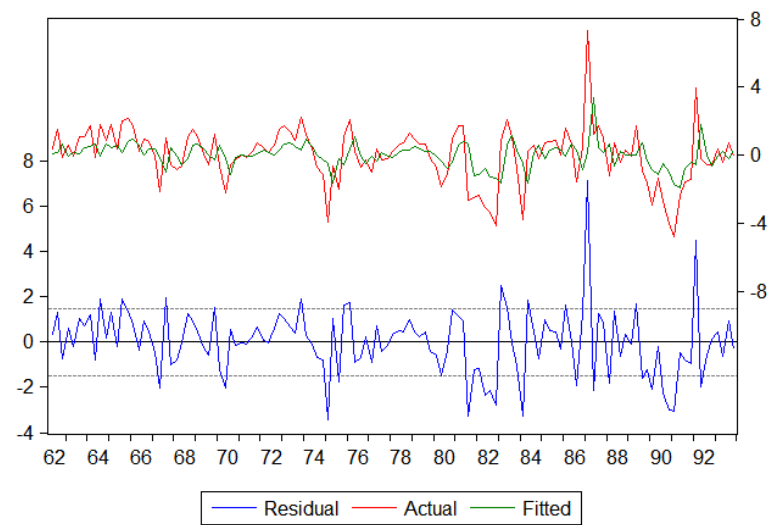


Figura 11: Residuos del modelo ARIMA(0,1,2)

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).

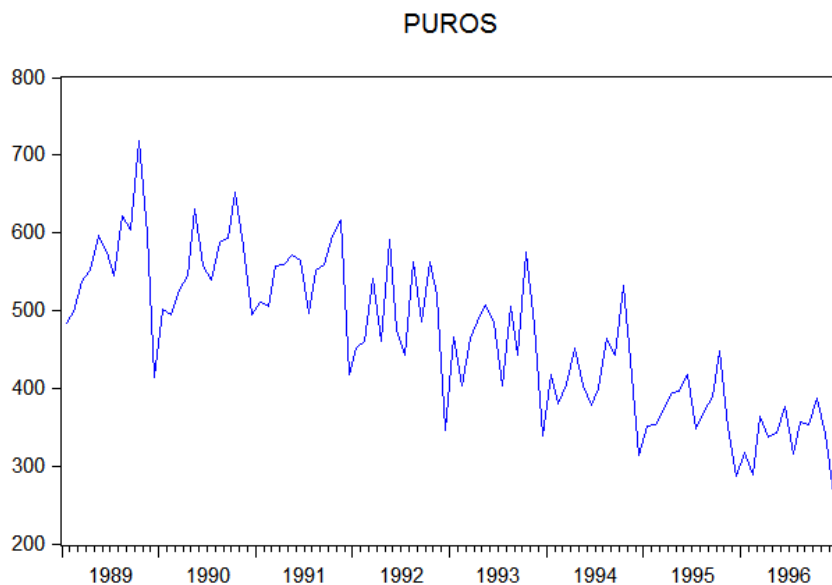


Figura 12: Serie temporal puros

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 descende 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

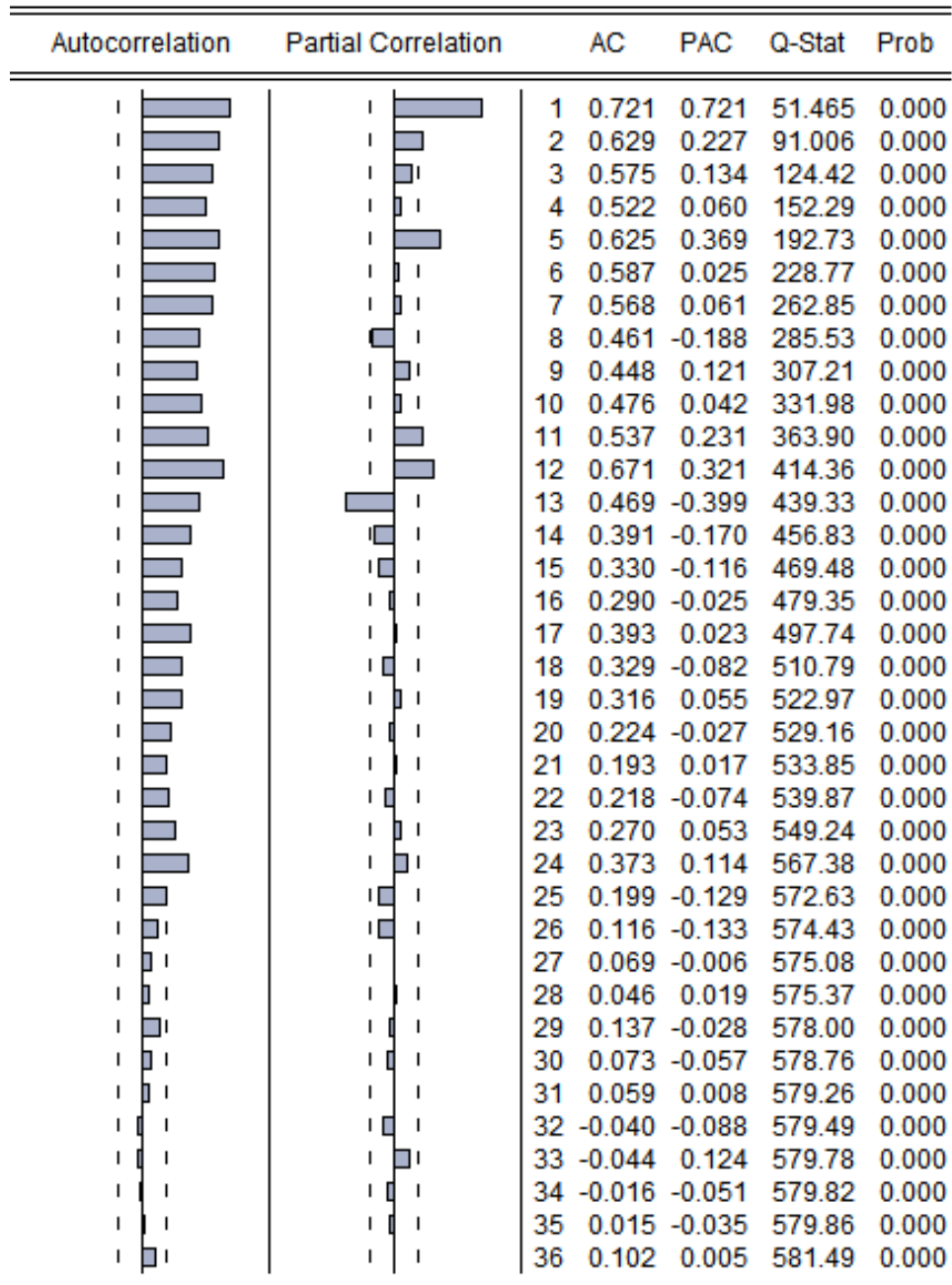


Figura 13: Correlograma de la serie temporal puros

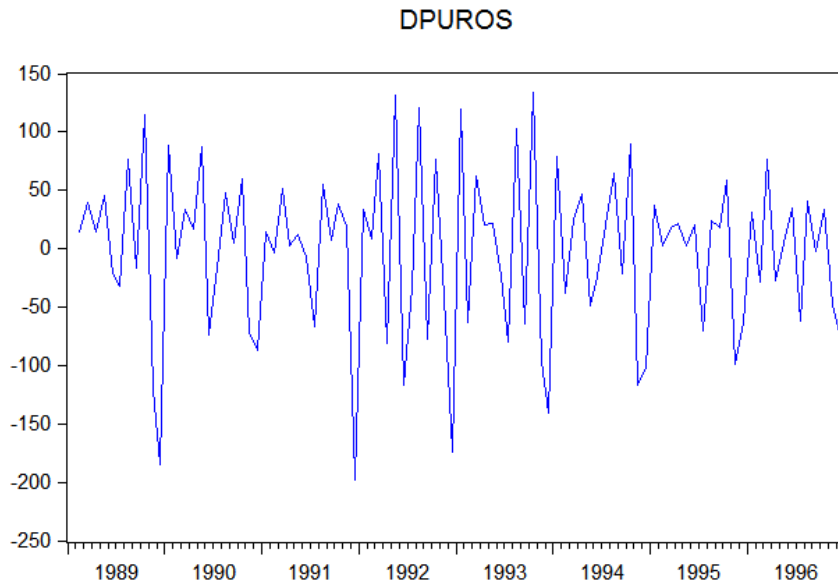


Figura 14: Serie temporal puros

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:

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

Es decir, tenemos que $d = D = 1$.

Debemos establecer el modelo generador de la serie.

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

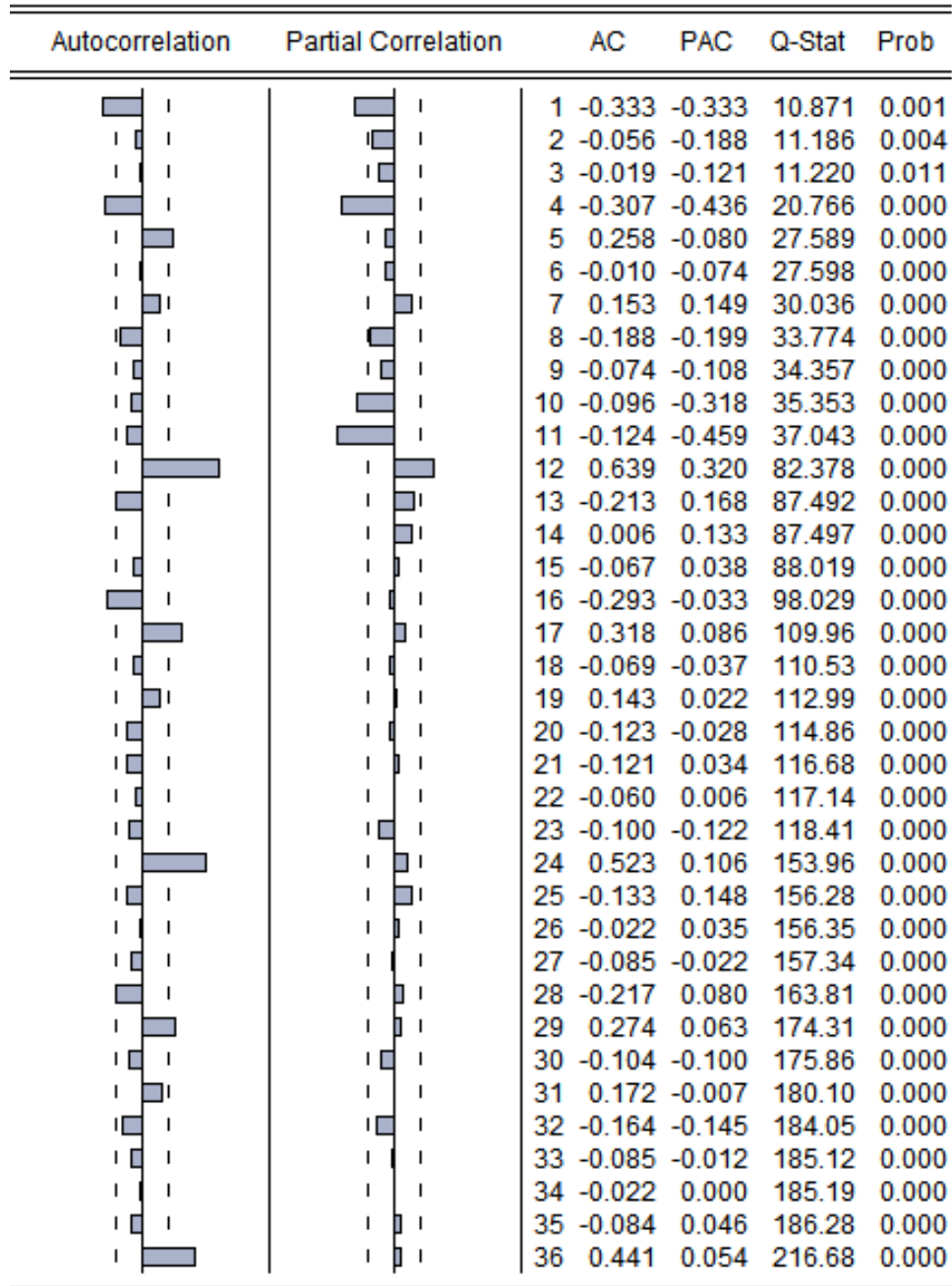


Figura 15: Correlograma de la serie temporal diferenciada dpuros

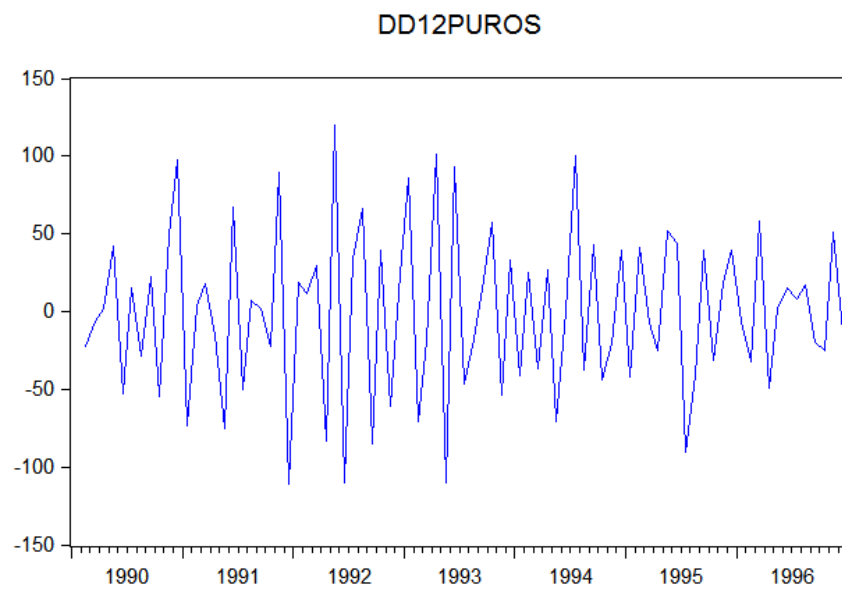


Figura 16: Serie temporal ddpuros12

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

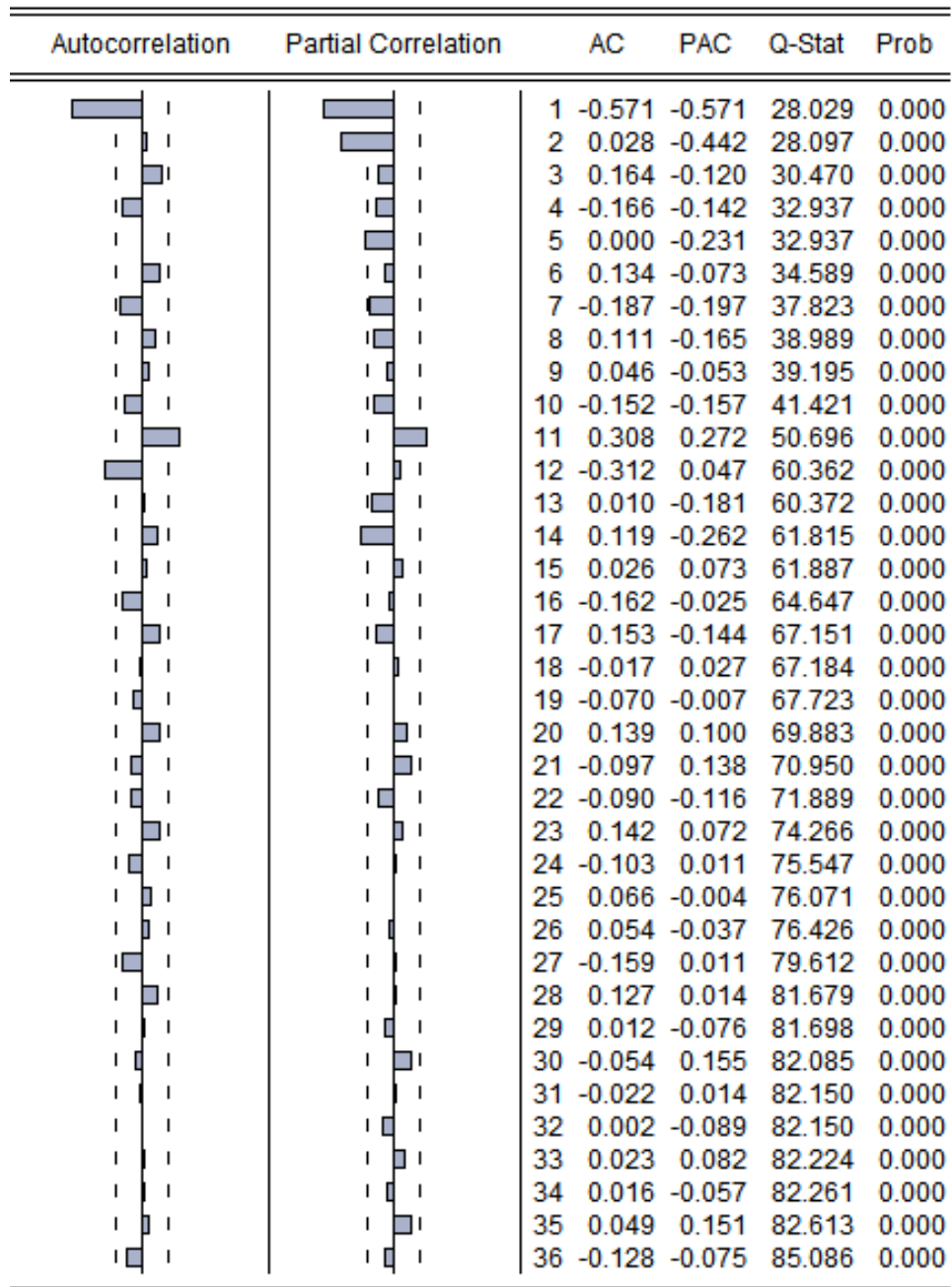


Figura 17: Correlograma de la serie temporal ddpuros12

3. Código EViews

3.1. 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
```

3.2. 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
```

4. Conclusiones