Predicción económica-empresarial

Trabajo Final:

Modelización y predicción

Carlos José Carrasco Jiménez

DNI: 25615736S

Tabla de contenido

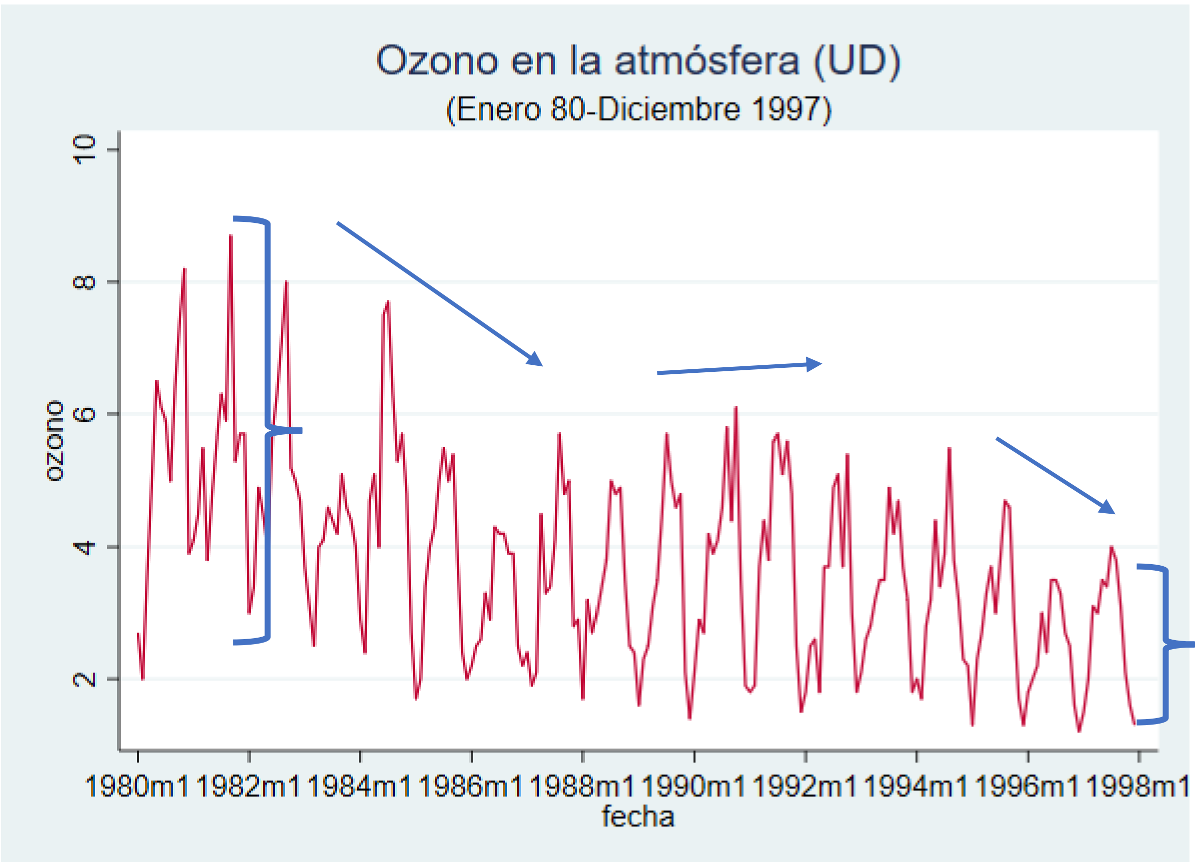
Análisis previo

En este análisis previo de la serie temporal "ozono", que reúne datos mensuales de la cantidad de ozono que existe en la atmósfera entre los periodos de enero de 1980 y diciembre de 1997. A nivel exploratorio resulta relevante señalar que:

- El proceso no es estacionario en media, por tanto, presenta (figura 1) una evolución irregular con una tendencia ascendente en los primeros meses, seguida de un notable descenso hasta la mitad del periodo donde se estabilizan los valores. Finalmente, desde 1992 hasta el final de la serie se observa una reducción de la cantidad de este gas en la atmósfera.

- Así como el proceso no es constante en media, tampoco lo es en varianza y es que se observan fluctuaciones menores en los últimos periodos de la serie, mientras que, al comienzo, las variaciones son mayores. De modo que se interpreta que los valores son más extremos al comienzo de la serie, por lo que se prevé que existía menos control que en los periodos de final de siglo.

Figura 1: Serie Temporal "ozono" (1980m1-1997m12)



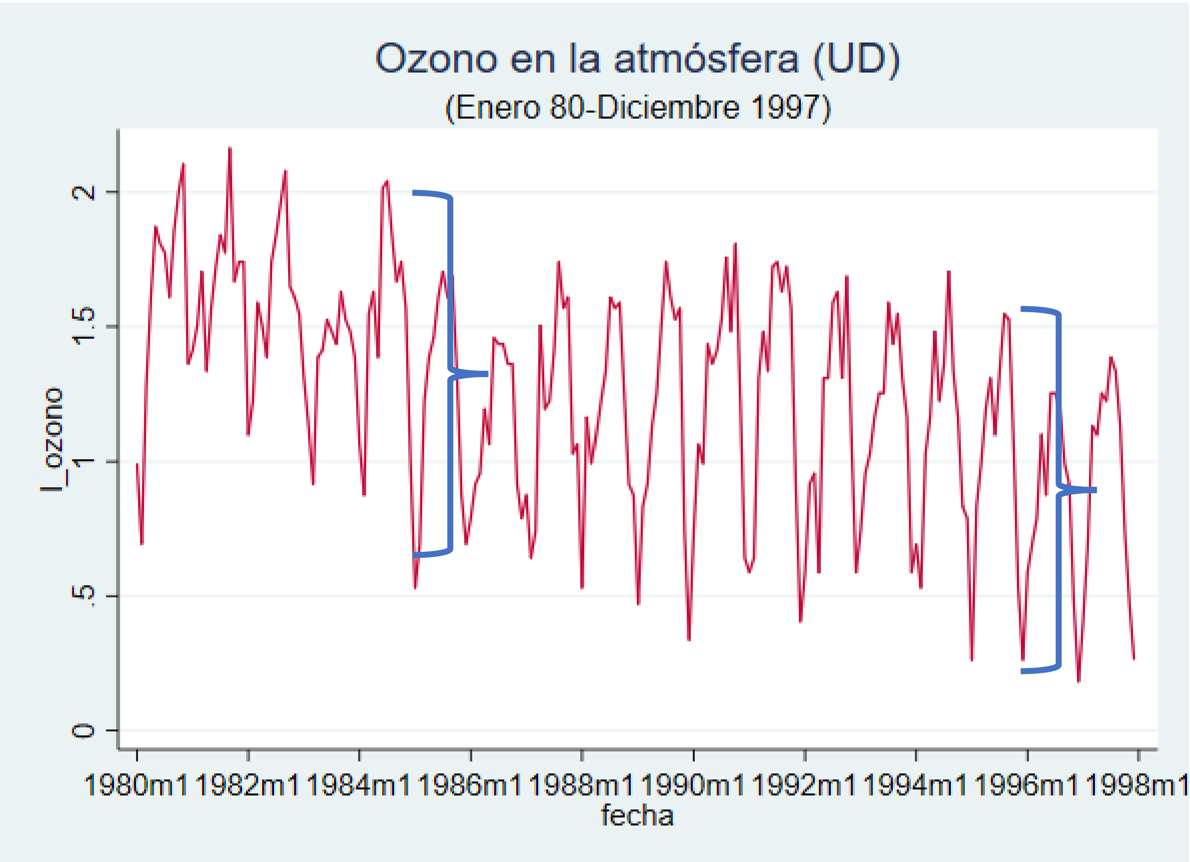
Transformaciones

Logaritmos

Con tal de modelizar la serie y realizar predicciones, se procede a transformar la variable de interés "ozono", y así hacer que algunas de las propiedades que presenta como no estacionarias, pase a serlo. En este caso se aplican logaritmos a la variable de estudio: cantidad de ozono en la atmósfera, es decir, aplicando la transformación Box-Cox a la serie, en la búsqueda de lograr la estacionariedad en varianza.

Los resultados que se consiguen con la transformación son lo que se observan en la figura 2.1, y es que a pesar que aún sigue siendo un proceso no estacionario en media, por la clara evolución descendente de la serie, como refleja la tendencia de la serie a lo largo del tiempo, sí que se ha conseguido estandarizar la varianza, por lo que se procede a tomar diferencias regulares, en términos estacionales (conociendo que la disposición de los datos en el espacio temporal es en meses), de la variable endógena para reducir el efecto de la tendencia y conseguir estandarizar la media del proceso.

Figura 2.1: Serie Temporal: "l_ozono" (1980m1-1997m12)

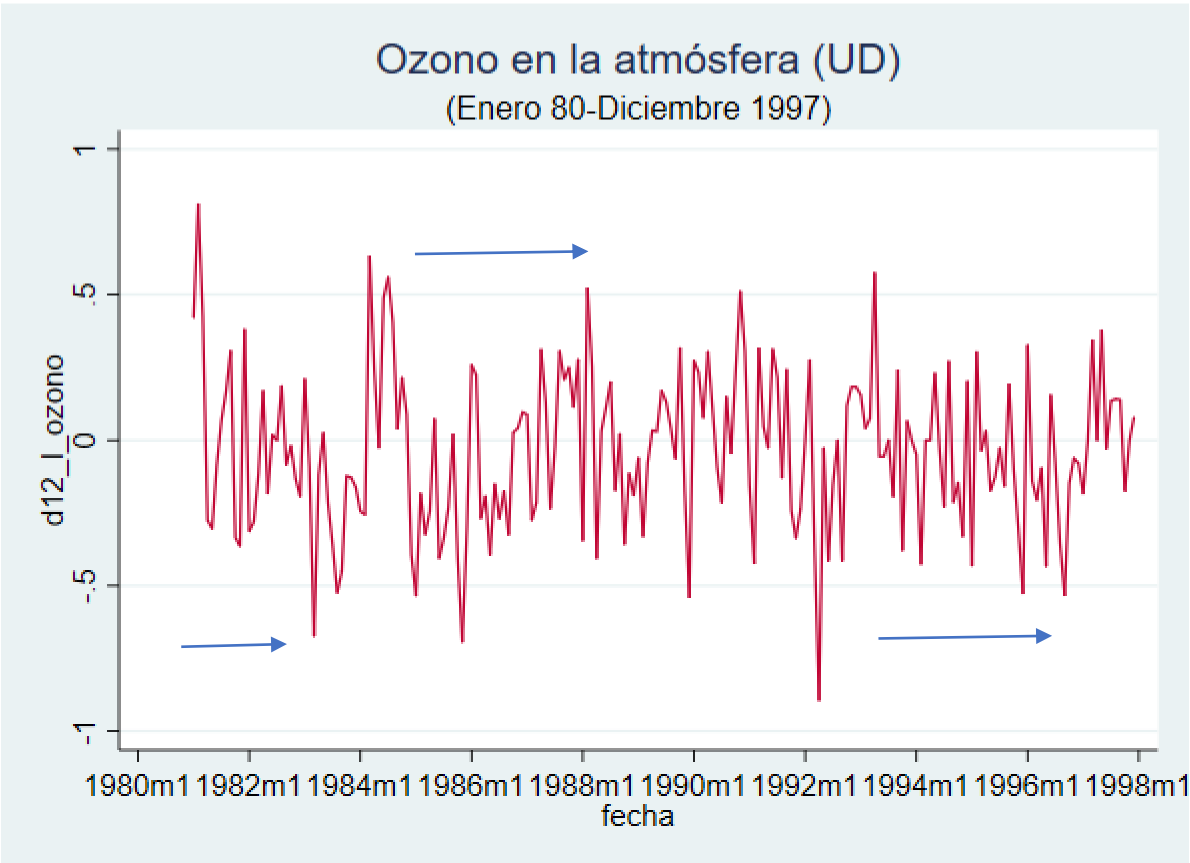


Diferencias

Según el gráfico de nivel, se observa un patrón estacional de duración anual donde en los últimos meses de cada año, y al comienzo del posterior, se produce un decrecimiento de la cantidad de ozono en la atmósfera hasta alcanzar un valor mínimo. Posteriormente se recuperan las cifras, alcanzando en ocasiones las que se registran en los mismos meses del año anterior, tal y como ocurre en la sección media de la serie, donde las variaciones son similares.

A través de la diferencia de carácter estacional entre meses de años diferentes, se estima que se corregirá la no estacionariedad del proceso. De no ser así, se procede a realizar además de esta, diferencias regulares, con tal de conseguir una serie estacionaria en media y varianza, reduciendo los niveles de crecimiento que marcan la tendencia del proceso. El objetivo con estas transformaciones de la variable logarítmica es el de conseguir un proceso estacionario en media. En otras palabras, se conoce que el valor logarítmico de la variable "ozono", estandariza la varianza de la serie. Luego, así surge la variable "d12_l_ozono", que reúne las primeras diferencias de la variable original logarítmica en el periodo t, con respecto al valor del periodo t-12.

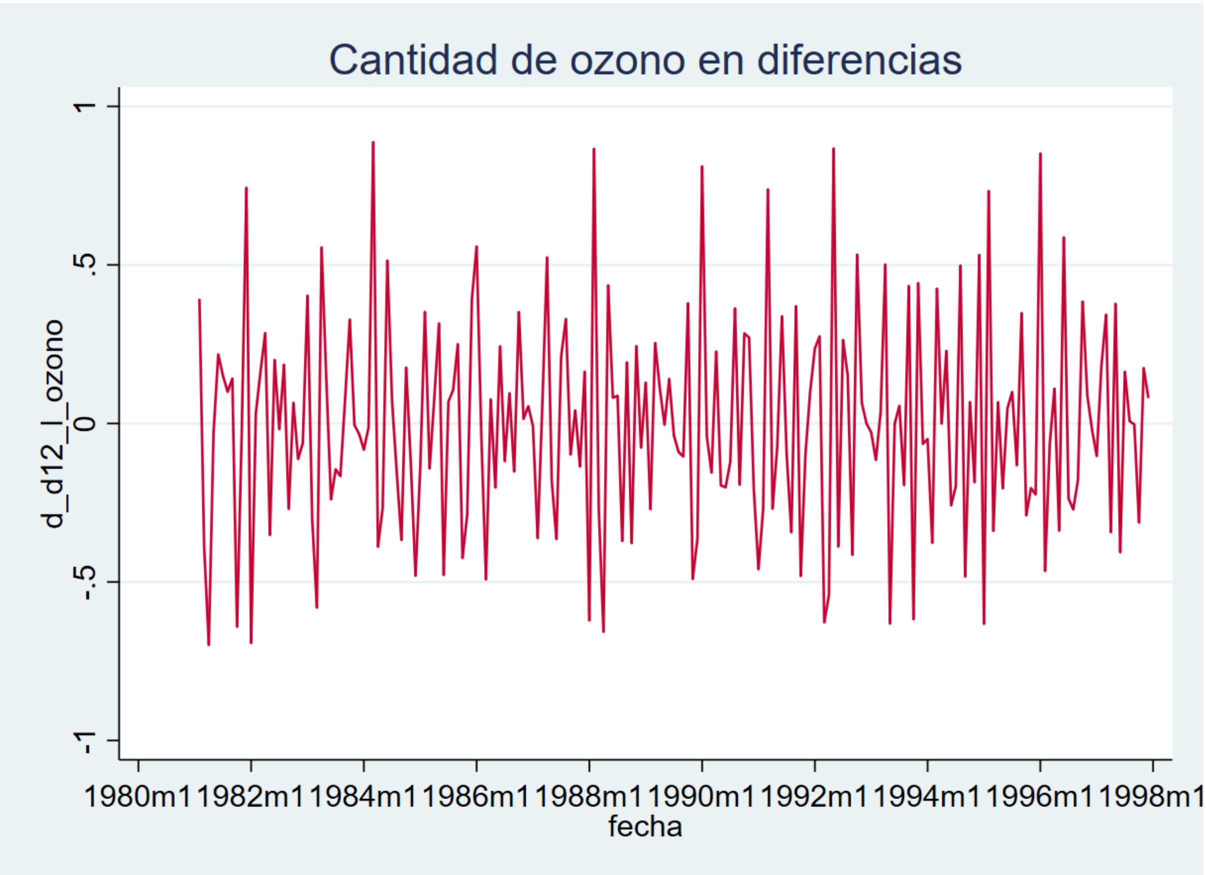
Figura 2.2: Serie Temporal: "d12_l_ozono" (1980m1-1997m12)



Efectivamente, como se intuía, el proceso ha quedado centrado en un valor cercano al cero, por lo que se considera a priori que cumple con estacionariedad en media y varianza. Se trata de una serie que presenta cierto conjunto de variaciones que se sitúan por encima y debajo del valor central alternativamente, por lo que se procede a realizar la diferencia regular y comprobar si el efecto de cambio de nivel se solventa, pese a que la variable a priori es estacionaria.

Con el objetivo de contemplar posibles alternativas, a este modelo estacional puro con diferencias estacionales mensuales se procede a analizar las primeras diferencias de la variable anterior, sobre diferencias estacionales "d12_l_ozono", y crear una nueva variable bajo el nombre de "d_d12_l_ozono", que cuenta con una diferencia estacional y otra regular. Obteniendo la siguiente serie:

Figura 2.3: Serie Temporal: "d_d12_l_ozono" (1980m1-1997m12)



A priori, el análisis gráfico refleja que tanto la diferencia estacional de logaritmos ("d12_l_ozono"), como la variable de primeras diferencias ("d_d12_l_ozono"), cumplen con estacionariedad en varianza, por mantener constante la fluctuación a lo largo del tiempo, y la media, por estar centradas en torno a un valor cercano a cero.

Aun así, para comprobar que efectivamente son variables estacionarias se recurre a realizar inferencia estadística, a través del contraste de Dickey-Fuller, obteniendo los siguientes resultados.

- Para la variable de diferencia estacional "d12_l_ozono": valor p asintótico 3.122e-07. Por lo tanto, la variable se considera estacionaria, pues se descarta la existencia de tendencia, y se ha comprobado que las variaciones se mantienen constante en el gráfico de serie temporal.

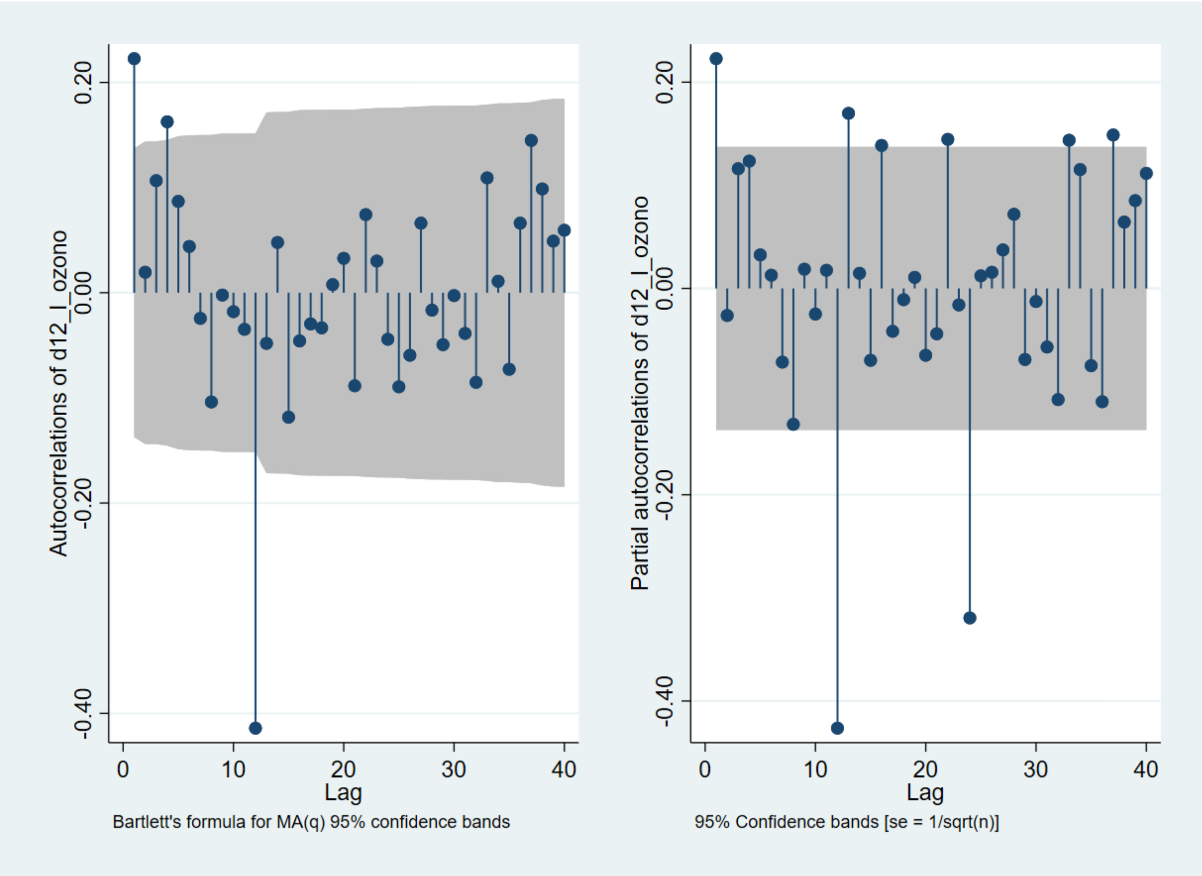
- Para la variable de primeras diferencias, junto con la variable de diferencia estacional "d_d12_l_ozono": valor p asintótico 0.0001. También se confirma su estacionariedad, y por tanto se tendrá en cuenta para la modelización como alternativa a los modelos con diferencia estacional únicamente.

Con los resultados obtenidos, se abren dos vías de investigación, y son por un lado modelizar un proceso ARIMA estacional puro, sin parte diferencia regular, o del otro lado desarrollar un modelo mixto con diferencias regulares y diferencias estacionarias. Se procede a analizar los gráficos de autocorrelación para determinar la estimación de los modelos.

Identificación del modelo

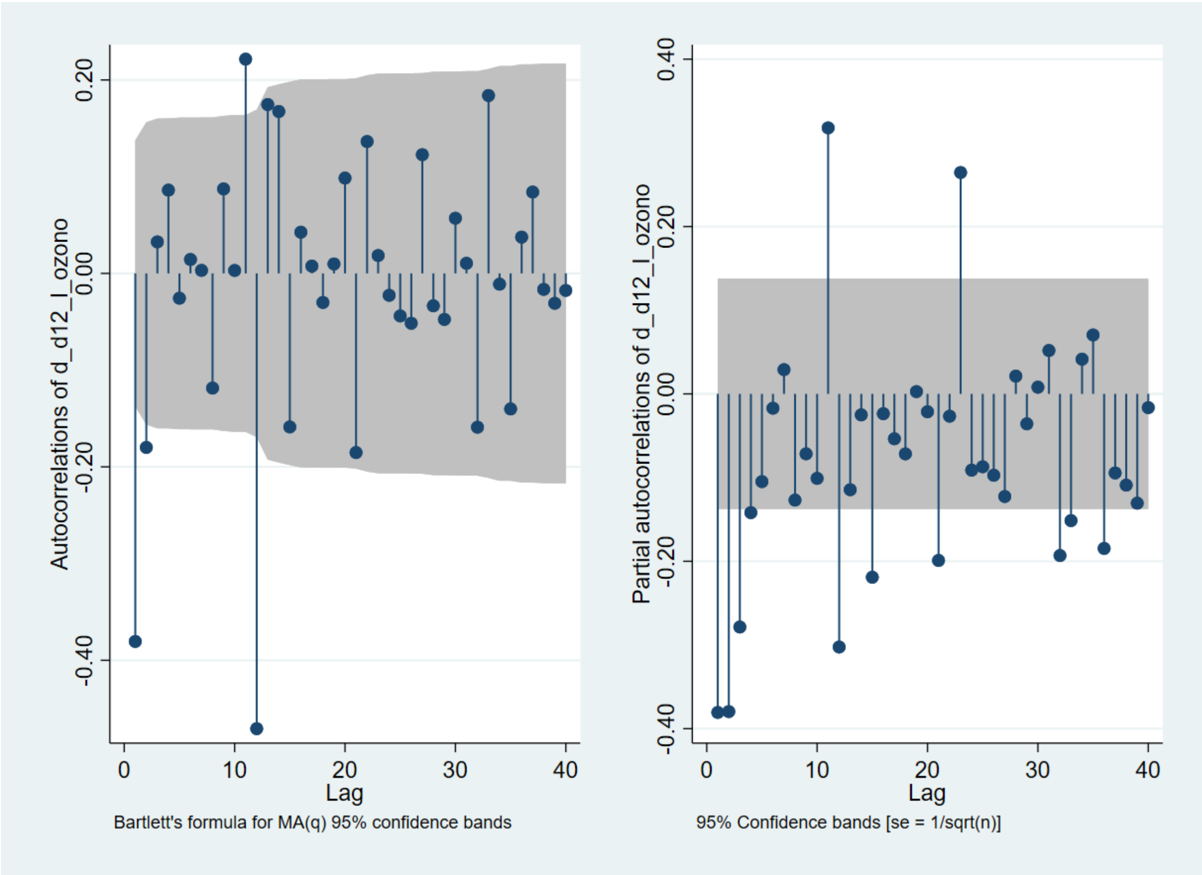
Para estimar el modelo que represente de manera significativa la serie de estudio, y que resulte útil para predecir valores del futuro de la cantidad de ozono en la atmósfera, se calculan las funciones de autocorrelación simple y parcial, para las dos alternativas planteadas.

Figura 3.1. FAC y FACP de "d12_l_ozono"



A tenor de los correlogramas, el modelo podría ser un proceso estacional con medias móviles MA(1), por la autocorrelación de orden 12 y 24 de la FACP donde se aprecia como existe un descenso rápido de la correlación con alternancia de signos, y resultan significativas según la función de autocorrelación FACP, sabiendo que la predominancia de la raíz es negativa. Se conoce que la variable consta de diferencias únicamente estacionales, por lo que el modelo puede seguir un modelo ARIMA$(0,1,1)_{12}$, aunque no se descartan otras opciones como añadir parte autorregresiva al modelo o, alternativamente aplicar una diferencia regular, como refleja el correlograma a continuación.

Figura 3.2. FAC y FACP de “d_d12_l_ozono”



Según los resultados obtenidos, se aprecia como en la diferencia regular, el modelo más claro es un MA(2) con un claro descenso de la autocorrelación de la FACP, mientras que el descenso es mucho más brusco en la FAC, aunque no se descarta la reducción en el apartado de medias móviles a un MA(1), sin sección autorregresiva y con una diferencia integrada.

Por otro lado, para la sección estacional, se observa como las autocorrelaciones de los periodos 12, 24 y 36 resultan significativas, según la FACP, mientras que la FAC no aclara la modelización estacional del proceso, por ello se estima un modelo ARIMA$(0,1,1)_{12}$.

Como conclusiones a priori, el modelo estacional único parece más apto para describir el proceso, mientras que el mixto, genera más confusión por la alternancia de signos y la presencia de nula significación en la FAC. Así los modelos que se estiman convenientes con el objetivo de predecir la cantidad de ozono en el futuro, y que por tanto permiten descartar aquellas componentes o modelos completos que no resulten significativos ni explicativos del fenómeno.

Se ha de destacar que el descarte de la modelización de la variable original, así como la de su primera diferencia, sus valores logarítmicos y la diferencia los logaritmos, se ha visto adecuado por diversos motivos, y es que no presentaban estacionariedad en media, sí en varianza para el caso de los logaritmos, pero, por otro lado, sus correlogramas presentaban sinuosidad con un patrón que no corresponde a modelos autorregresivos con medias móviles ni diferencias integradas.

De este modo los modelos que se han identificado y procede su estimación son:

ARIMA$(0,1,1)_{12}$

ARIMA$(2,1,0)_{12}$

ARIMA$(0,1,2)\,x(0,1,1)_{12}$

ARIMA$(0,1,1)\,x(0,1,1)_{12}$

ARIMA$(2,1,0)\,x(0,1,1)_{12}$

Se espera que con el desarrollo de la modelización se refine el modelo predictivo, permitiendo ajustarse a las observaciones de manera correcta, con tal de ofrecer una óptima capacidad de ajuste y de predicción.

Estimación

Se procede a la estimación del modelo ARIMA $(0,1,1)_{12}$, o el modelo de medias móviles sobre la variable endógena definida con una diferencia estacional entre periodos con diferencia 12, es decir con diferencia anual del mismo mes. La variable a modelizar por tanto es "l_ozono" que muestra los valores logarítmicos de la serie original sobre la cantidad de ozono en la atmósfera. Para comparar los resultados con otros posibles modelos explicativos de las variaciones de la variable en cuestión, se estiman junto con el modelo que se ha descrito anteriormente los siguientes: ARIMA$(0,1,2)\,x(0,1,1)_{12}$; ARIMA$(0,1,1)\,x(0,1,1)_{12}$; ARIMA $(2,1,0)\,x(0,1,1)_{12}$ y la alternativa autorregresiva de ARIMA$(2,1,0)_{12}$. Así los resultados obtenidos quedan representados en las siguientes tablas:

ARIMA$(0,1,1)_{12}$

```
ARIMA regression

   le:  1981m1 - 1997m12                   Number of obs     =        204
                                           Wald chi2(1)      =     112.82
Log likelihood =  10.93408                 Prob > chi2       =     0.0000
```

S12.l_ozono	Coef.	OPG Std. Err.	z	P>|z|	[95% Conf.	Interval]
l_ozono						
_cons	-.0413764	.0062809	-6.59	0.000	-.0536868	-.029066
ARMA12						
ma L1.	-.6756124	.0636065	-10.62	0.000	-.8002789	-.5509459
/sigma	.2252674	.0113673	19.82	0.000	.2029878	.247547

$\text{ARIMA}(2, 1, 0)_{12}$

```
ARIMA regression

Sample:  1981m1 - 1997m12                        Number of obs     =        204
                                                 Wald chi2(2)      =      62.69
Log likelihood =   8.96607                       Prob > chi2       =     0.0000
```

S12.l_ozono	Coef.	OPG Std. Err.	z	P>|z|	[95% Conf.	Interval]
l_ozono						
_cons	-.0444299	.0094262	-4.71	0.000	-.0629049	-.0259549
ARMA12						
ar						
L1.	-.6015832	.0791733	-7.60	0.000	-.7567599	-.4464065
L2.	-.3101058	.0716858	-4.33	0.000	-.4506075	-.1696042
/sigma	.228595	.0117228	19.50	0.000	.2056187	.2515713

$\text{ARIMA}(0, 1, 2)x(0, 1, 1)_{12}$

```
ARIMA regression

Sample:  1981m2 - 1997m12                        Number of obs     =        203
                                                 Wald chi2(3)      =     506.39
Log likelihood =  20.39936                       Prob > chi2       =     0.0000
```

DS12.l_ozono	Coef.	OPG Std. Err.	z	P>|z|	[95% Conf.	Interval]
l_ozono						
_cons	-.0001267	.0005962	-0.21	0.832	-.0012952	.0010419
ARMA						
ma						
L1.	-.6732147	.0723084	-9.31	0.000	-.8149364	-.5314929
L2.	-.2219728	.0651466	-3.41	0.001	-.3496578	-.0942877
ARMA12						
ma						
L1.	-.7471954	.0632625	-11.81	0.000	-.8711876	-.6232032
/sigma	.2125231	.0103792	20.48	0.000	.1921803	.232866

$\text{ARIMA}(0,1,1)x(0,1,1)_{12}$

ARIMA regression

Sample: 1981m2 - 1997m12 | Number of obs = 203 | Wald chi2(2) = 411.62

Log likelihood = 15.79634 | Prob > chi2 = 0.0000

DS12.l_ozono	Coef.	OPG Std. Err.	z	P>|z|	[95% Conf.	Interval]
l_ozono						
_cons	-.0003584	.0010141	-0.35	0.724	-.002346	.0016291
ARMA						
ma						
L1.	-.819537	.0494013	-16.59	0.000	-.9163619	-.7227122
ARMA12						
ma						
L1.	-.7305353	.060651	-12.04	0.000	-.8494091	-.6116615
/sigma	.2181907	.0104762	20.83	0.000	.1976576	.2387237

$\text{ARIMA}(2,1,0)x(0,1,1)_{12}$

ARIMA regression

Sample: 1981m2 - 1997m12 | Number of obs = 203 | Wald chi2(3) = 243.07

Log likelihood = 6.319501 | Prob > chi2 = 0.0000

DS12.l_ozono	Coef.	OPG Std. Err.	z	P>|z|	[95% Conf.	Interval]
l_ozono						
_cons	-.0008596	.002987	-0.29	0.774	-.0067139	.0049947
ARMA						
ar						
L1.	-.4787519	.0637959	-7.50	0.000	-.6037895	-.3537143
L2.	-.3188821	.0696027	-4.58	0.000	-.4553009	-.1824633
ARMA12						
ma						
L1.	-.7491982	.0587948	-12.74	0.000	-.8644339	-.6339626
/sigma	.2287115	.0116739	19.59	0.000	.2058311	.2515919

Los resultados obtenidos en las estimaciones de cada uno de los modelos, reflejan que la mayoría no cuenta con término independiente, siendo el ARIMA$(0,1,1)_{12}$ el único que lo hace. Centrando el análisis en aquellos que cuenta únicamente con diferencia estacional, es destacable que para el primero de ellos, que cuenta con término independiente, presenta una mayor probabilidad de ocurrencia según las observaciones, pues su Log Likelihood es superior al segundo modelo de esta categoría, algo que lo sitúa como la mejor opción, siendo además un proceso con menos parámetros.

Del otro lado, los modelos mixtos con diferencia regular y diferencia estacional de la variable original expresada en valores logarítmicos, se seleccionan el primero y segundo de ellos: ARIMA$(0,1,2)x(0,1,1)_{12}$ y ARIMA$(0,1,1)x(0,1,1)_{12}$, respectivamente. Ninguno presenta término independiente, por lo que en caso que los modelos estacionales puros estimados resulten no válidos en los contrastes de los residuos, se aplican estas alternativas para predecir el fenómeno atmosférico en cuestión.

Validación

El foco de este apartado se centra en el estudio de los residuos de los modelos estudiados hasta este punto, en particular la validación irá dedicada a analizar el modelo con diferencia estacional ARIMA$(0,1,1)_{12}$, tal y como se ha descrito en el apartado anterior.

El análisis del ruido blanco reúne sus descriptivos principales, valor medio y variabilidad con respecto a la media al cuadrado (varianza), así como la correlación entre valores no contemporáneos, y por último su ajuste a la distribución normal.

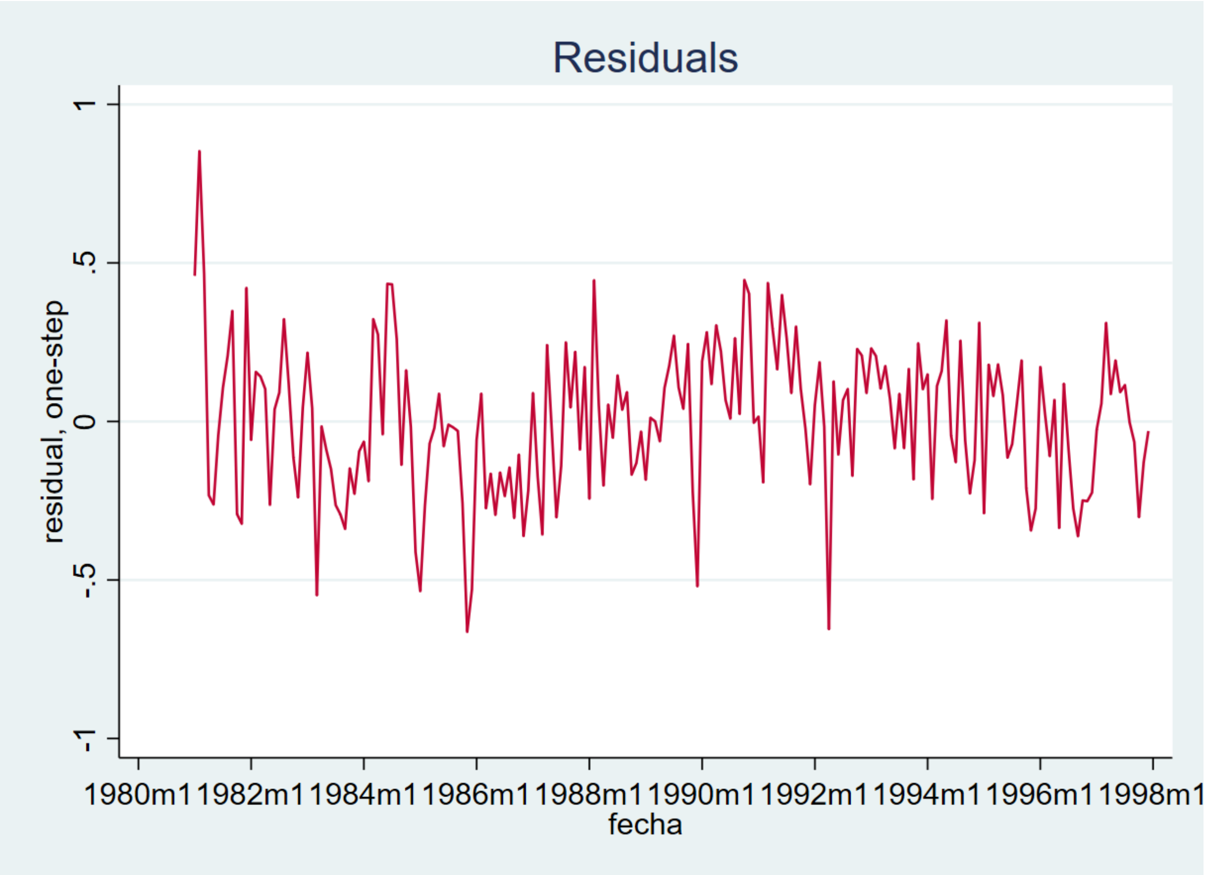
Media

```
. ttest residuals == 0

One-sample t test
------------------------------------------------------------------------------
Variable |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
residu~s |     204    .0014061    .0163286    .2332188   -.0307893    .0336014
------------------------------------------------------------------------------
    mean = mean(residuals)                                        t =   0.0861
Ho: mean = 0                                     degrees of freedom =      203

    Ha: mean < 0                 Ha: mean != 0                 Ha: mean > 0
 Pr(T < t) = 0.5343         Pr(|T| > |t|) = 0.9315          Pr(T > t) = 0.4657
```

El test de contraste para la hipótesis nula que confirma la inferencia estadística para conocer si efectivamente la media de los residuos generados por el modelo es igual a cero, o por el contrario es distinta de cero, según el nivel de significación obtenido, no se puede rechazar la hipótesis de que la media es nula.

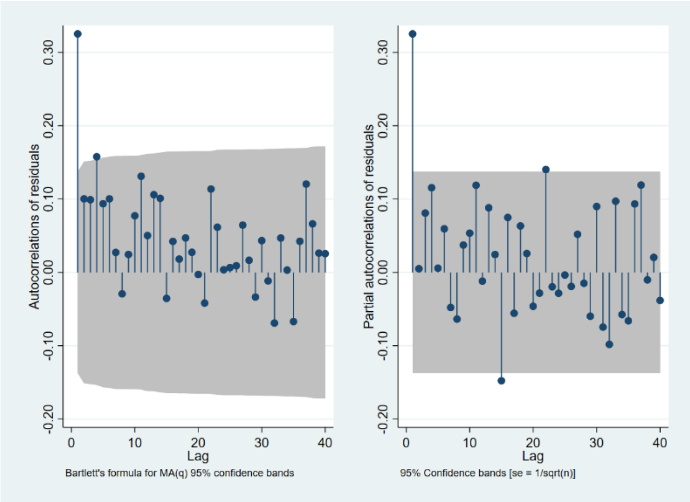
Esta premisa confirma que a priori, los residuos corresponden a la propiedad del valor medio del ruido blanco, sin ser aún confirmadas el resto de propiedades para concluir con esta idea. Para ello será necesaria analizar el resto de propiedades relacionadas con la varianza, autocorrelación o distribución normal.

Varianza



Según el gráfico de serie temporal de los residuos, se interpreta que la varianza, pese al salto que experimenta en los primeros periodos del proceso, y las irregulares fluctuaciones en la mitad de la serie, se observa una varianza constante a lo largo del periodo.

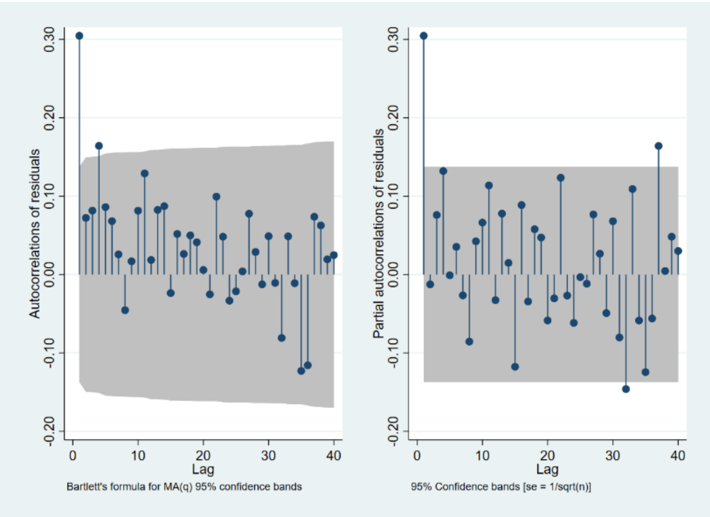
Autocorrelación




```
. corrgram residuals
```

LAG	AC	PAC	Q	Prob>Q	-1 0 1 [Autocorrelation]	-1 0 1 [Partial Autocor]
1	0.3252	0.3252	21.893	0.0000		
2	0.1003	0.0050	23.987	0.0000		
3	0.0991	0.0810	26.04	0.0000		
4	0.1577	0.1155	31.265	0.0000		
5	0.0936	0.0057	33.116	0.0000		
6	0.1003	0.0595	35.252	0.0000		
7	0.0272	-0.0479	35.41	0.0000		
8	-0.0292	-0.0637	35.593	0.0000		
9	0.0244	0.0373	35.721	0.0000		
10	0.0772	0.0536	37.011	0.0001		
11	0.1310	0.1189	40.75	0.0000		
12	0.0502	-0.0119	41.302	0.0000		
13	0.1060	0.0881	43.773	0.0000		
14	0.1010	0.0244	46.03	0.0000		
15	-0.0354	-0.1477	46.31	0.0000		
16	0.0423	0.0749	46.709	0.0001		
17	0.0181	-0.0557	46.783	0.0001		
18	0.0470	0.0634	47.282	0.0002		
19	0.0274	0.0257	47.453	0.0003		
20	-0.0029	-0.0462	47.455	0.0005		
21	-0.0418	-0.0283	47.856	0.0007		
22	0.1137	0.1403	50.839	0.0005		
23	0.0616	-0.0196	51.719	0.0005		
24	0.0035	-0.0285	51.722	0.0009		
25	0.0066	-0.0036	51.732	0.0013		
26	0.0091	-0.0191	51.752	0.0019		
27	0.0645	0.0520	52.739	0.0022		
28	0.0165	-0.0148	52.805	0.0031		
29	-0.0336	-0.0597	53.075	0.0041		
30	0.0433	0.0899	53.528	0.0052		
31	-0.0116	-0.0746	53.561	0.0072		
32	-0.0690	-0.0981	54.726	0.0074		
33	0.0469	0.0970	55.266	0.0089		
34	0.0032	-0.0574	55.268	0.0120		
35	-0.0671	-0.0660	56.386	0.0124		
36	0.0425	0.0934	56.837	0.0149		
37	0.1205	0.1191	60.49	0.0087		
38	0.0661	-0.0102	61.598	0.0091		
39	0.0263	0.0205	61.773	0.0116		
40	0.0255	-0.0382	61.94	0.0146		

El modelo ARIMA$(0,1,1)_{12}$ con término constante, y sin él, presenta autocorrelación a nivel gráfico y estadístico, de manera que no cumple con los requisitos necesarios para considerar los residuos como ruido blanco. Así, se realiza el mismo análisis para el modelo con diferencia estacional de ARIMA$(0,1,2)_{12}$ con término independiente y sin él, sabiendo que cumple con el resto de premisas propuestas hasta ahora, tal y como se puede comprobar en el código adjunto al presente documento.



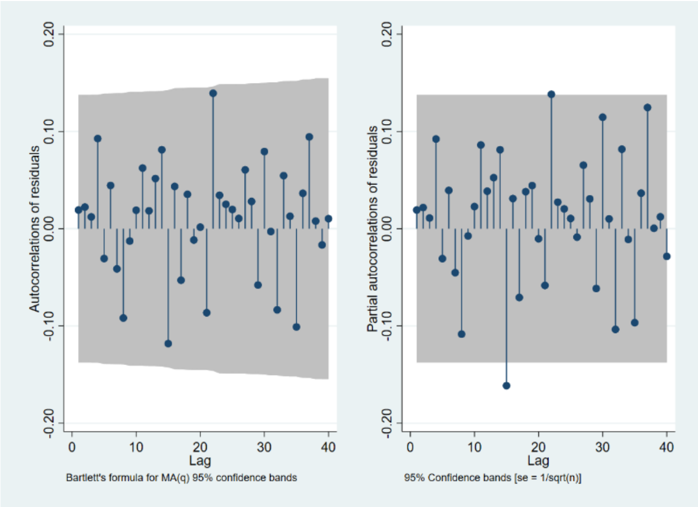
. corrgram residuals

LAG	AC	PAC	Q	Prob>Q	-1 0 1 [Autocorrelation]	-1 0 1 [Partial Autocor]
1	0.3046	0.3046	19.202	0.0000		
2	0.0722	-0.0126	20.288	0.0000		
3	0.0815	0.0759	21.676	0.0001		
4	0.1641	0.1321	27.337	0.0000		
5	0.0860	-0.0010	28.9	0.0000		
6	0.0681	0.0353	29.886	0.0000		
7	0.0257	-0.0266	30.026	0.0001		
8	-0.0455	-0.0856	30.469	0.0002		
9	0.0168	0.0424	30.53	0.0004		
10	0.0815	0.0662	31.967	0.0004		
11	0.1290	0.1137	35.59	0.0002		
12	0.0185	-0.0327	35.665	0.0004		
13	0.0824	0.0777	37.159	0.0004		
14	0.0871	0.0148	38.838	0.0004		
15	-0.0237	-0.1176	38.963	0.0006		
16	0.0518	0.0887	39.563	0.0009		
17	0.0262	-0.0344	39.717	0.0014		
18	0.0499	0.0578	40.28	0.0019		
19	0.0411	0.0472	40.664	0.0027		
20	0.0058	-0.0586	40.672	0.0041		
21	-0.0253	-0.0304	40.819	0.0059		
22	0.0994	0.1235	43.099	0.0046		
23	0.0482	-0.0268	43.638	0.0058		
24	-0.0334	-0.0617	43.899	0.0078		
25	-0.0215	-0.0034	44.007	0.0108		
26	0.0040	-0.0117	44.011	0.0151		
27	0.0775	0.0766	45.438	0.0146		
28	0.0287	0.0266	45.635	0.0190		
29	-0.0126	-0.0492	45.673	0.0253		
30	0.0488	0.0681	46.249	0.0294		
31	-0.0108	-0.0804	46.278	0.0382		
32	-0.0809	-0.1462	47.878	0.0353		
33	0.0487	0.1092	48.46	0.0403		
34	-0.0111	-0.0586	48.49	0.0511		
35	-0.1230	-0.1245	52.25	0.0305		
36	-0.1158	-0.0561	55.604	0.0195		
37	0.0736	0.1641	56.968	0.0190		
38	0.0626	0.0045	57.959	0.0200		
39	0.0193	0.0483	58.054	0.0253		
40	0.0248	0.0301	58.212	0.0313		

Según los valores del correlograma, se aprecia cierta autocorrelación a nivel gráfico, que es confirmada en el análisis de Ljung-Box, que se adjunta a continuación. Así se concluye que, según los modelos estimados considerados como modelos estacionales puros, no cumplen con los contrastes necesarios para que se considere a los residuos generados como ruido blanco, pues a pesar de presentar valor medio cero, y varianza constante, están auto correlacionados, según el contraste analítico y gráfico. De este modo, se procede a analizar el otro grupo de modelos planteados, que presentan diferencias estacionales y regulares.

Se procede a analizar el modelo ARIMA$(0,1,2)x(0,1,1)_{12}$ con término constante en primer lugar, obteniendo los siguientes resultados. Se conoce que el resto de premisas planteadas anteriormente[1], en referencia a los contrastes sobre la media y la varianza con tal de comprobar si efectivamente cumple las propiedades adecuadas, para que los residuos sean considerados como ruido blanco, son aceptadas. Y es que a priori, no es significativo el término constante[2] por lo que descartamos su existencia y el análisis consecuente dado su nivel de significación, por tanto, se procede a mostrar únicamente el análisis del modelo ARIMA{(0,1,2)x(0,1,1)}_{12} sin término constante.

[1] En el apartado anexo al presente documento, se puede comprobar que las premisas son correctamente analizadas.
[2] La estimación del modelo con y sin término constante quedan adjuntas al final del presente documento en el apartado anexo.




```
. corrgram residuals
```

LAG	AC	PAC	Q	Prob>Q	-1 0 1 [Autocorrelation]	-1 0 1 [Partial Autocor]
1	0.0192	0.0192	.07624	0.7825		
2	0.0223	0.0217	.1791	0.9143		
3	0.0122	0.0110	.2099	0.9760		
4	0.0927	0.0922	2.0062	0.7346		
5	-0.0307	-0.0308	2.2038	0.8203		
6	0.0445	0.0394	2.6221	0.8546		
7	-0.0413	-0.0452	2.9837	0.8865		
8	-0.0918	-0.1084	4.7807	0.7807		
9	-0.0126	-0.0074	4.815	0.8501		
10	0.0190	0.0228	4.8931	0.8982		
11	0.0624	0.0861	5.7368	0.8903		
12	0.0184	0.0386	5.8107	0.9253		
13	0.0516	0.0525	6.394	0.9306		
14	0.0813	0.0812	7.8477	0.8971		
15	-0.1181	-0.1614	10.935	0.7572		
16	0.0435	0.0310	11.356	0.7870		
17	-0.0530	-0.0707	11.984	0.8011		
18	0.0353	0.0381	12.264	0.8333		
19	-0.0116	0.0444	12.295	0.8726		
20	0.0016	-0.0103	12.295	0.9055		
21	-0.0864	-0.0584	14.002	0.8695		
22	0.1393	0.1383	18.463	0.6782		
23	0.0344	0.0274	18.737	0.7164		
24	0.0251	0.0204	18.884	0.7582		
25	0.0198	0.0105	18.975	0.7983		
26	0.0105	-0.0086	19.001	0.8364		
27	0.0605	0.0654	19.868	0.8363		
28	0.0280	0.0307	20.054	0.8625		
29	-0.0579	-0.0615	20.856	0.8643		
30	0.0794	0.1146	22.372	0.8402		
31	-0.0029	0.0101	22.374	0.8709		
32	-0.0834	-0.1035	24.067	0.8420		
33	0.0545	0.0818	24.794	0.8472		
34	0.0128	-0.0110	24.835	0.8745		
35	-0.1010	-0.0966	27.36	0.8181		
36	0.0364	0.0366	27.691	0.8380		
37	0.0944	0.1247	29.927	0.7889		
38	0.0078	0.0004	29.942	0.8215		
39	-0.0167	0.0122	30.013	0.8488		
40	0.0104	-0.0284	30.041	0.8741		

Según los resultados del correlograma, se puede confirmar que no existe correlación entre los residuos no contemporáneos, de manera que se confirman las hipótesis para creer que los residuos se corresponden con ruido blanco.

En este punto del estudio, se conoce que el modelo ARIMA$(0,1,2)x(0,1,1)_{12}$ sin término constante, presenta la mejor ratio de verosimilitud de entre los modelos estudiados hasta el momento, y además sus residuos muestran una media cercana a cero con una varianza constante, según la inferencia estadística visible en el anexo del presente documento, así como unos valores no autocorrelacionados. Se procede pues a estimar el contraste de distribución normal para comprobar si finalmente sus valores pertenecen a ruido blanco, y por tanto la inferencia aplicada es válida, o de lo contrario será necesario formular y contrastar otros modelos para predecir los valores en los periodos sucesivos.

Normalidad

```
. ***CONTRASTES DE NORMALIDAD
. sktest residuals

                   Skewness/Kurtosis tests for Normality
                                                       ------ joint ------
    Variable |        Obs  Pr(Skewness)  Pr(Kurtosis) adj chi2(2)   Prob>chi2
-------------+---------------------------------------------------------------
   residuals |        203     0.0828        0.1345        5.25         0.0724
```

Dado un p valor superior al 0.05, del estadístico correspondiente al contraste de Kolmogorov-Smirnov, se concluye que los residuos se ajustan a una distribución normal, y por tanto no se puede rechazar la hipótesis nula, a un nivel de confianza del 95 por ciento.

Predicción

Una vez descartados los modelos puramente estacionales, por la existencia de autocorrelación de los residuos, el que mejor se ajusta a los datos observados es el que se ha analizado en último lugar, que se corresponde con un ARIMA$(0,1,2)x(0,1,1)_{12}$ sin término constante. Es un modelo que se ha estimado con todas las observaciones disponibles, desde el primer periodo del año 1980, hasta el último de 1997.

Comentar que para comprobar que las conclusiones obtenidas hasta este punto de la investigación son correctas, se ha realizado la estimación de los modelos analizados previamente tomando únicamente las observaciones comprendidas entre los años 1980-1996, obteniendo las mismas conclusiones[3]. De modo que con el objetivo de predecir la serie para los periodos correspondientes al último año 1997, se estima el modelo tomando el rango de datos de 1980-1996[4], cuya predicción permitirá comparar las estimaciones realizadas por el modelo con las observaciones de la serie real.

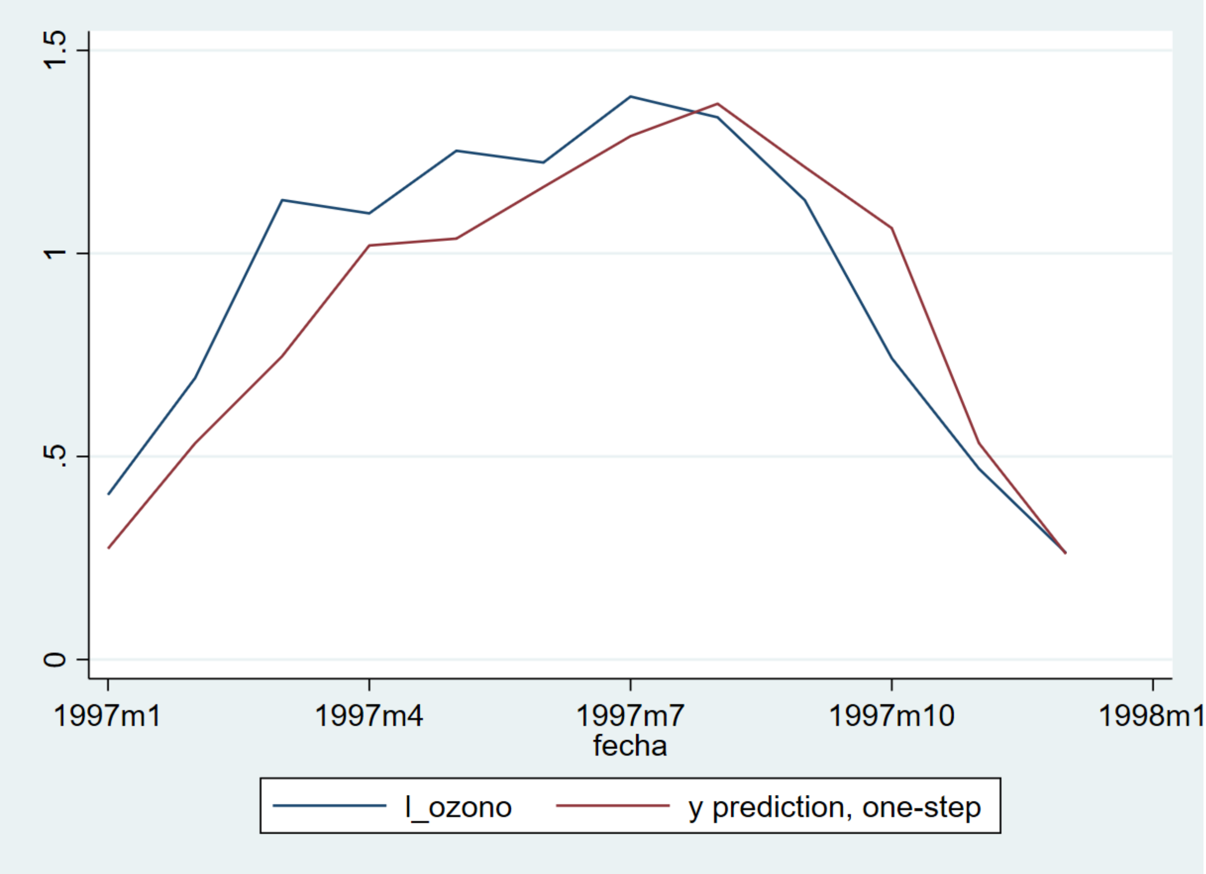
Así los resultados obtenidos en la predicción son los que figuran a continuación:

```
. list l_ozono l_ozono_pred if year>1996
```

	l_ozono	l_ozon~d
205.	.4054651	.2730466
206.	.6931472	.5325164
207.	1.131402	.7469166
208.	1.098612	1.019488
209.	1.252763	1.036389
210.	1.223776	1.163515
211.	1.386294	1.288805
212.	1.335001	1.368495
213.	1.131402	1.212812
214.	.7419373	1.062103
215.	.4700036	.5328661
216.	.2623642	.2600381

[3] Quedan representadas y con opción de ser chequeadas en el archivo de código adjunto.

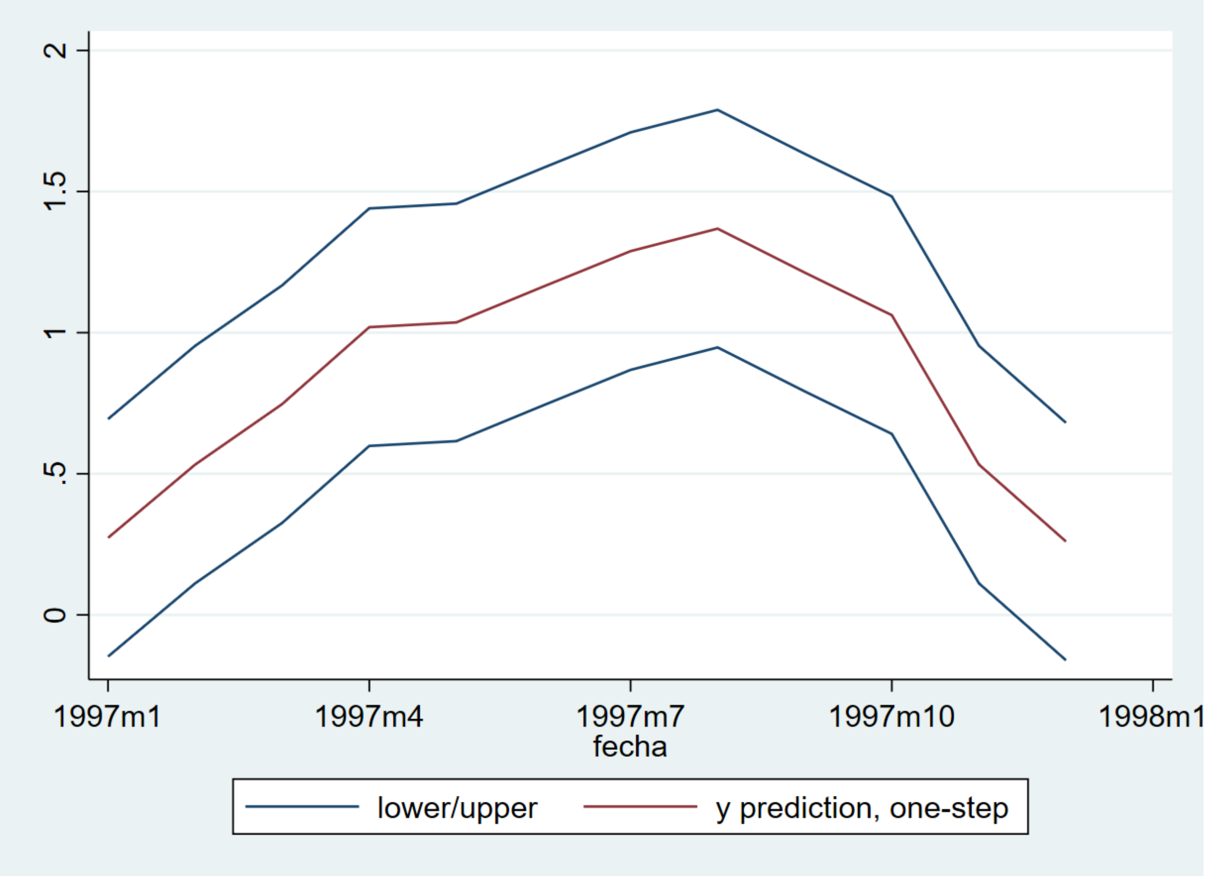
[4] Los resultados obtenidos de la estimación y correspondientes contrastes, queda visible en el apartado anexo del presente documento.



La comparativa entre la serie estimada y la observada, como parte del proceso de predicción, refleja que en los valores medios y finales del periodo 1997, muestran un mejor ajuste a la serie real, que los valores iniciales.

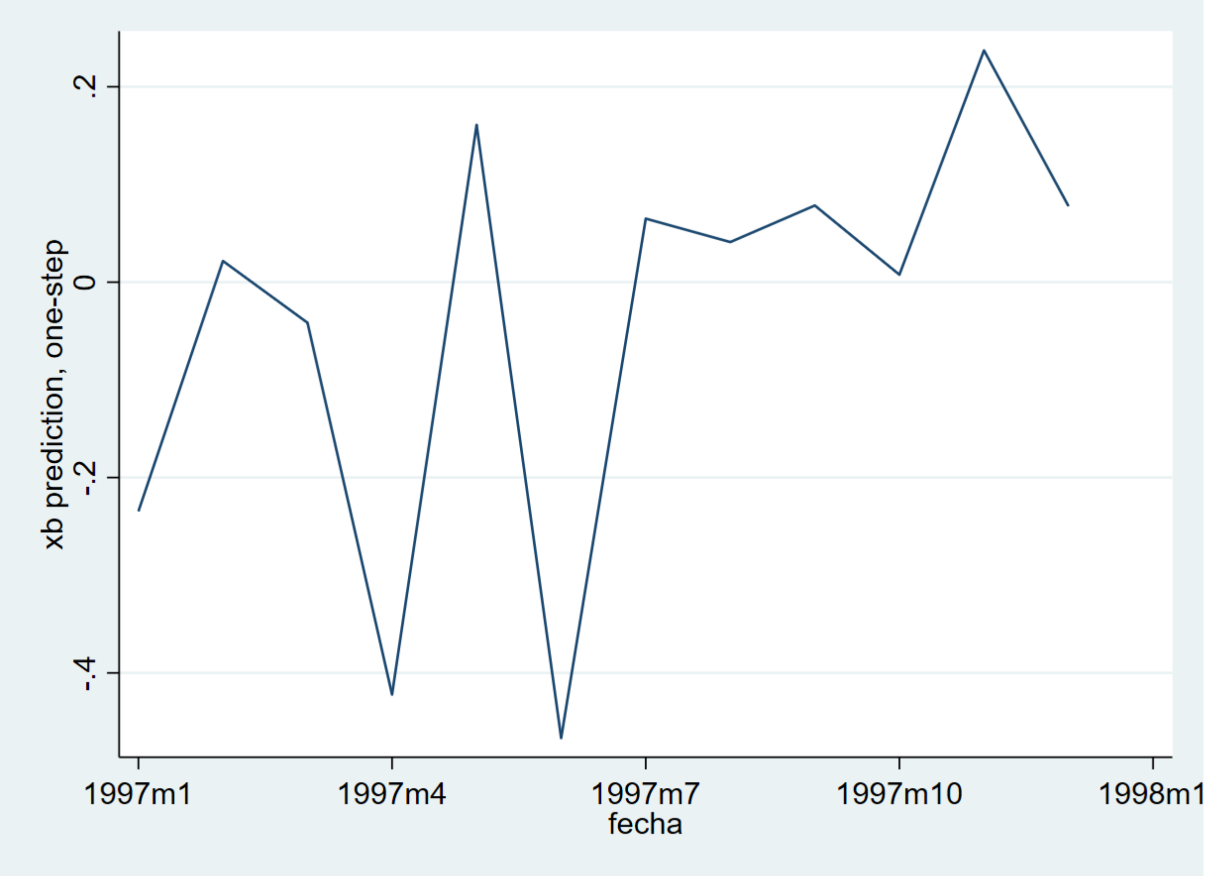
Es decir, particularmente, el modelo predice de manera correcta los meses de agosto y diciembre, mientras que para el resto de meses, las estimaciones de la predicción superan las observaciones entre los meses de agosto y diciembre, y, del otro lado, entre enero y julio, los valores se sitúan por debajo de los valores reales de la serie observada para los periodos recogidos en el año 1997.

Además, se han estimado los valores para cada mes dentro de un intervalo de confianza para el 95 por ciento, que quedan representados en el apartado adjunto como un listado con las estimaciones de cada uno de los valores para cada mes, y también de forma visual, en el gráfico siguiente:



Por último, se ha calculado el error cuadrático medio que mide la diferencia que existe entre la predicción y la observación de la variable, de modo que es posible conocer en términos estadístico cuanto se desvía el modelo de la realidad, y conocer por tanto el error medio al cuadrado que presente en cada uno de los periodos.

Los valores específicos para cada valor de la predicción quedan visibles en el apartado anexo del presente documento, y a continuación se muestra una gráfica que representa la desviación al cuadrado de las predicciones con respecto a los valores observados. Se aprecia como en los valores medios y finales de la gráfica, el error se aproxima más a cero, y por tanto coincide con la comparativa realizada en la gráfica anterior en la que analiza a nivel gráfico la serie de predicción con la original "l_ozono", coincidiendo en esta idea de mejor ajuste en los periodos medios y finales de la serie.



Conclusiones

Dado los resultados obtenidos el proceso que mejor se adapta atendiendo a los criterios de selección de modelos, es el ARIMA$(0,1,2)x(0,1,1)_{12}$ sin término constante, de modo que el ajuste a los valores reales de la observación es mejor en este modelo [5]que los demás planteados.

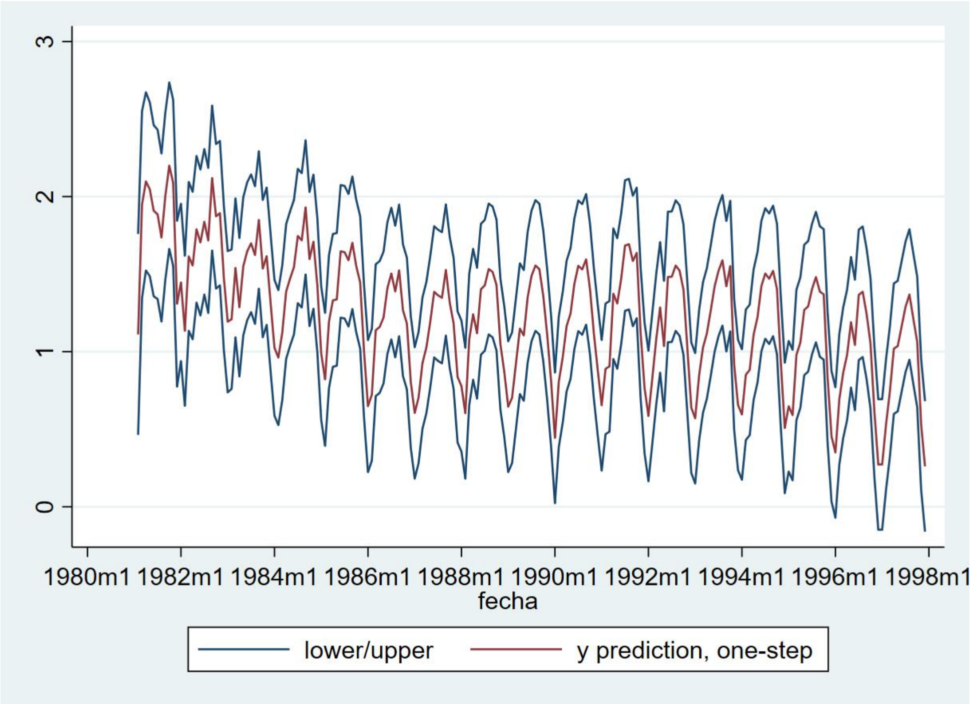
Por tanto, el modelo final queda definido como:

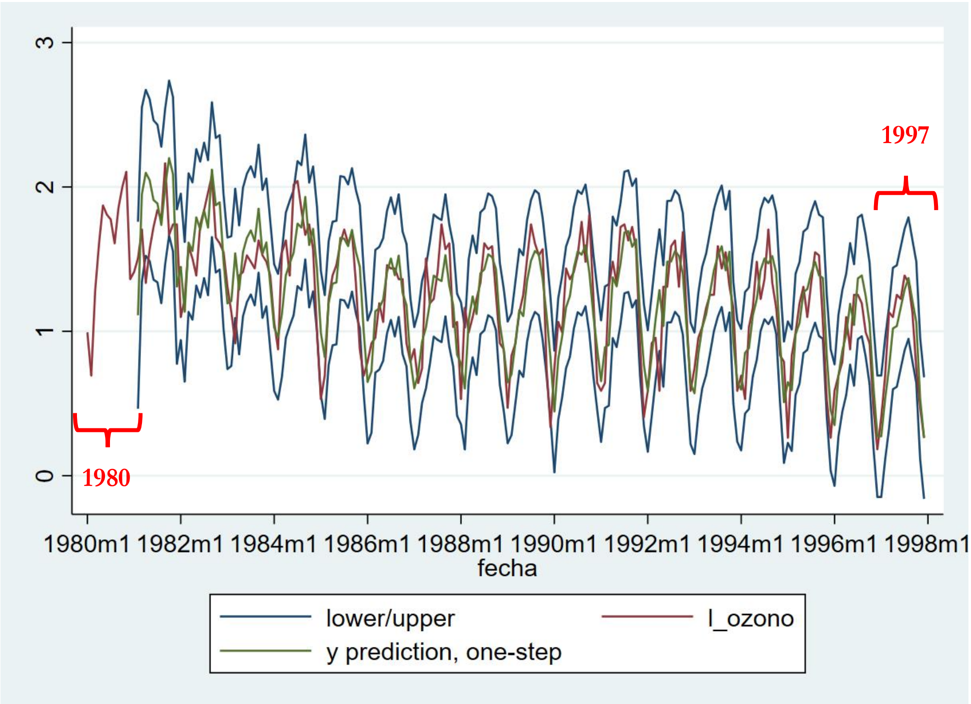
$$(1-L)(1-L^{12})\ln(ozono_t) = (1-0.679L-0.195L^2)(1-0.762L^{12})\varepsilon_t;$$
$$\Delta\Delta^{12}\ln(ozono_t) = (1-0.679L-0.195L^2)(1-0.762L^{12})\varepsilon_t;$$
$$\Delta\Delta^{12}\ln(ozono_t) = -0.679\varepsilon_{t-1} - 0.195\varepsilon_{t-2} - 0.762\varepsilon_{t-12} + \varepsilon_t$$

Como conclusión, las predicciones que realiza el modelo estimado para todo el periodo de la serie quedan recogidas en el siguiente gráfico. Al tratarse de un modelo para series temporales a base de

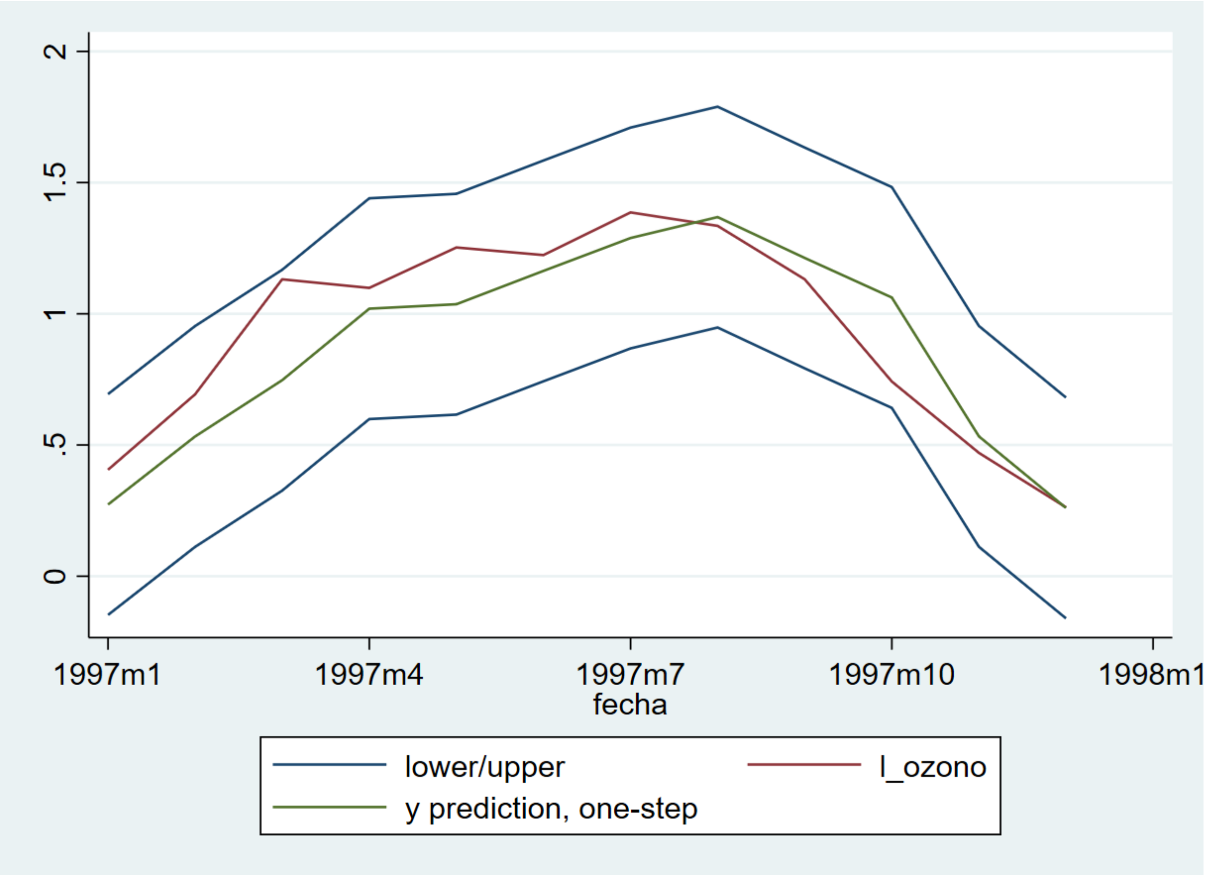
[5] Los coeficientes del modelo, corresponden a la última estimación realizada, que recoge en exclusiva los periodos correspondientes a los años 1980-1996.

diferencia regular y estacional, se predice a partir de la obtención de los primeros resultados, es decir en el periodo t+12, siendo en ese momento, en el que se produce la primera diferencia estacional, tal y como reflejan los dos gráficos adjuntos a continuación. El primero de ellos, únicamente las predicciones y sus intervalos de confianza son graficados, mientras que en el segundo se incorpora la serie original con tal de visualizar las desviaciones de la serie que predice.





Finalmente se grafica el año referencia para la predicción (1997) junto a los valores observados, y los correspondientes al intervalo de confianza con el objetivo de conocer si pese a las desviaciones más notables del comienzo del año 1997, los valores observados se encuentran dentro del intervalo de confianza que establece el modelo estimado al 95 por ciento.



Anexo

ARIMA$(0,1,2)x(0,1,1)_{12}$ con término constante

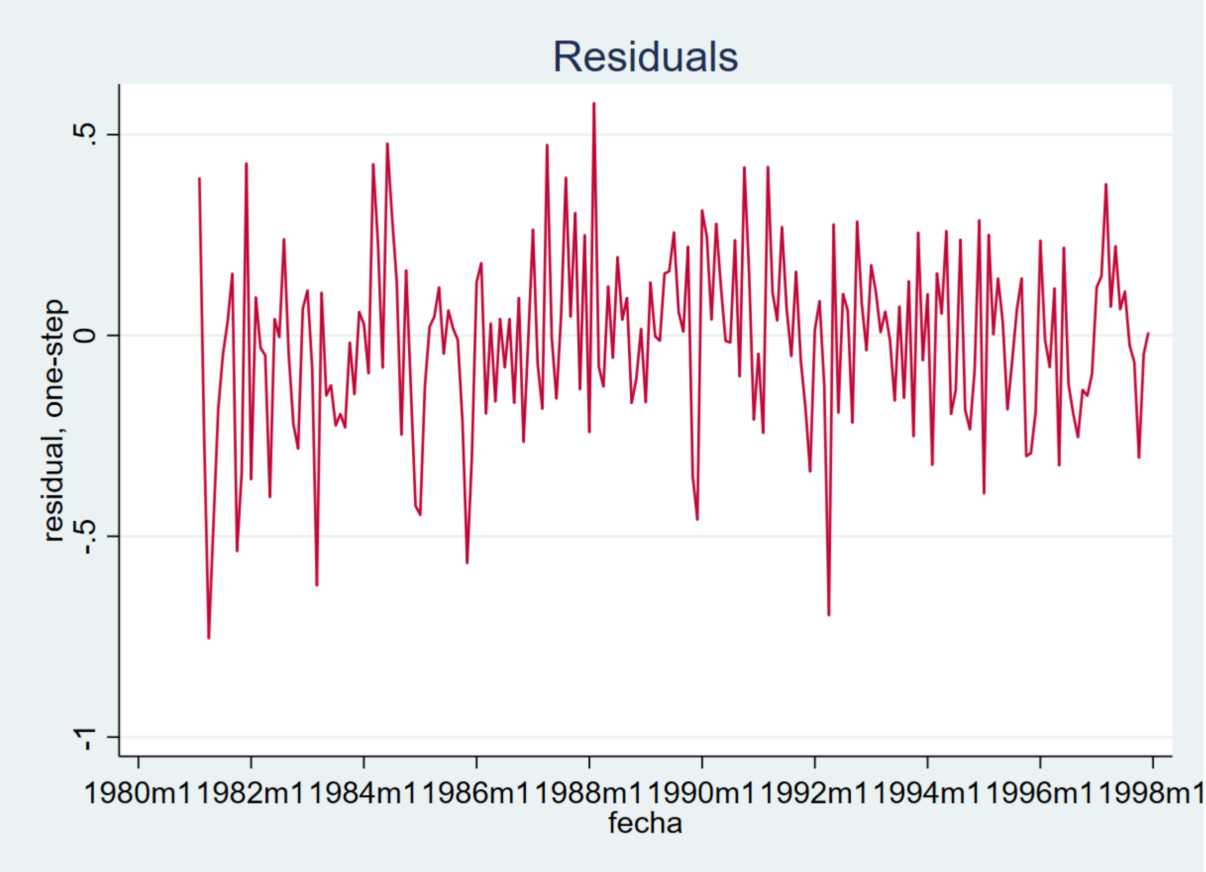
```
ARIMA regression

Sample:  1981m2 - 1997m12                       Number of obs     =        203
                                                Wald chi2(3)      =     506.39
Log likelihood =  20.39936                      Prob > chi2       =     0.0000
```

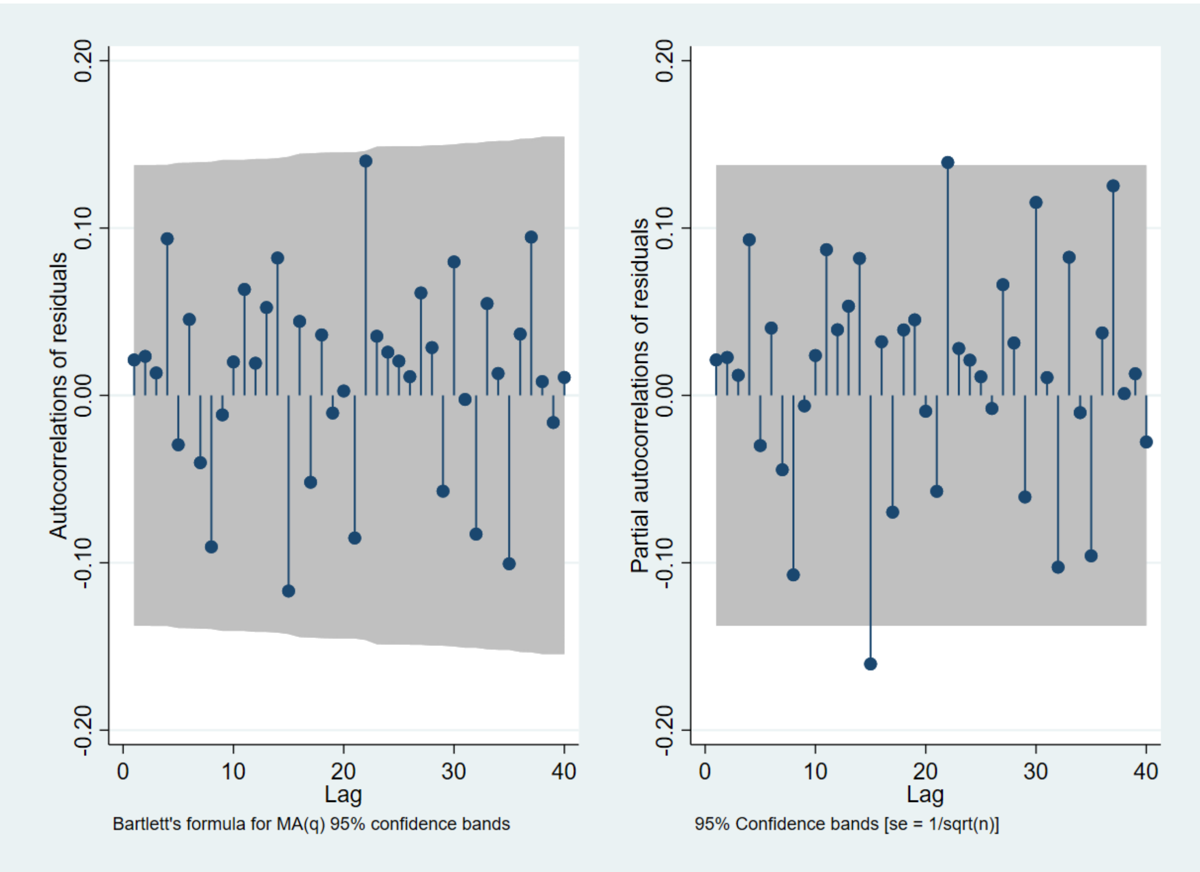
DS12.l_ozono	Coef.	OPG Std. Err.	z	P>|z|	[95% Conf.	Interval]
l_ozono						
_cons	-.0001267	.0005962	-0.21	0.832	-.0012952	.0010419
ARMA						
ma						
L1.	-.6732147	.0723084	-9.31	0.000	-.8149364	-.5314929
L2.	-.2219728	.0651466	-3.41	0.001	-.3496578	-.0942877
ARMA12						
ma						
L1.	-.7471954	.0632625	-11.81	0.000	-.8711876	-.6232032
/sigma	.2125231	.0103792	20.48	0.000	.1921803	.232866

```
. ****CONTRASTE MEDIA RESIDUOS CERO
. ttest residuals == 0

One-sample t test
```

Variable	Obs	Mean	Std. Err.	Std. Dev.	[95% Conf.	Interval]
residu~s	203	-.0153566	.0157015	.2237113	-.0463164	.0156032

```
    mean = mean(residuals)                                        t =  -0.9780
Ho: mean = 0                                     degrees of freedom =      202

    Ha: mean < 0                 Ha: mean != 0                 Ha: mean > 0
 Pr(T < t) = 0.1646         Pr(|T| > |t|) = 0.3292          Pr(T > t) = 0.8354
```


Residuals
residual, one-step
.5
0
-.5
-1
1980m1 1982m1 1984m1 1986m1 1988m1 1990m1 1992m1 1994m1 1996m1 1998m1
fecha

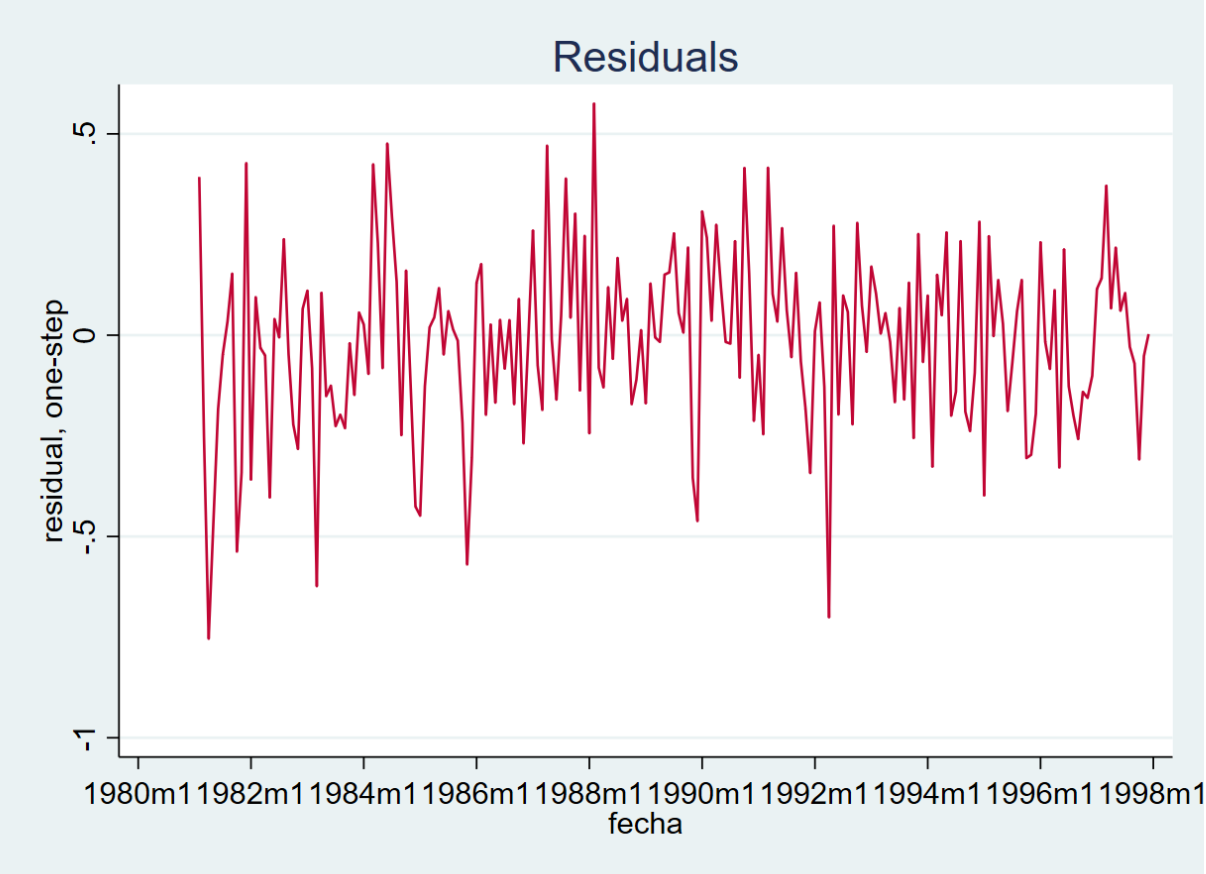

Autocorrelations of residuals
0.20
0.10
0.00
-0.10
-0.20
0 10 20 30 40
Lag
Bartlett's formula for MA(q) 95% confidence bands
Partial autocorrelations of residuals
0.20
0.10
0.00
-0.10
-0.20
0 10 20 30 40
Lag
95% Confidence bands [se = 1/sqrt(n)]

```
. ***TEST LJUNG-BOX
. corrgram residuals

                                          -1       0       1 -1       0       1
 LAG       AC       PAC      Q     Prob>Q  [Autocorrelation]  [Partial Autocor]
-------------------------------------------------------------------------------
1        0.0212   0.0212  .09266  0.7608
2        0.0233   0.0227  .20537  0.9024
3        0.0134   0.0121  .24286  0.9704
4        0.0936   0.0930  2.0758  0.7218
5       -0.0295  -0.0299  2.2586  0.8123
6        0.0454   0.0402  2.6939  0.8462
7       -0.0402  -0.0444  3.0363  0.8816
8       -0.0905  -0.1072  4.7833  0.7805
9       -0.0116  -0.0063  4.8122  0.8504
10       0.0200   0.0239  4.8985  0.8979
11       0.0634   0.0871  5.7684  0.8884
12       0.0192   0.0393  5.8491  0.9235
13       0.0525   0.0533  6.4532  0.9281
14       0.0821   0.0819  7.9372  0.8926
15      -0.1169  -0.1604   10.96  0.7554
16       0.0442   0.0321  11.395  0.7845
17      -0.0519  -0.0697  11.997  0.8003
18       0.0362   0.0391  12.291  0.8318
19      -0.0105   0.0452  12.316  0.8717
20       0.0026  -0.0095  12.318  0.9047
21      -0.0852  -0.0573  13.978  0.8706
22       0.1400   0.1392  18.486  0.6768
23       0.0354   0.0280  18.775  0.7142
24       0.0258   0.0211   18.93  0.7557
25       0.0205   0.0111  19.028  0.7958
26       0.0112  -0.0078  19.057  0.8340
27       0.0612   0.0661  19.944  0.8331
28       0.0286   0.0314  20.138  0.8594
29      -0.0572  -0.0607  20.921  0.8620
30       0.0798   0.1153  22.451  0.8372
31      -0.0024   0.0107  22.452  0.8683
32      -0.0829  -0.1026  24.123  0.8399
33       0.0549   0.0825   24.86  0.8449
34       0.0131  -0.0103  24.902  0.8724
35      -0.1006  -0.0958  27.408  0.8163
36       0.0367   0.0373  27.744  0.8362
37       0.0946   0.1252  29.986  0.7866
38       0.0083   0.0011  30.003  0.8194
39      -0.0162   0.0130   30.07  0.8470
40       0.0108  -0.0278  30.099  0.8724
```

```
. ***CONTRASTES DE NORMALIDAD
. sktest residuals

                    Skewness/Kurtosis tests for Normality
                                                          ------ joint ------
    Variable |        Obs  Pr(Skewness)  Pr(Kurtosis) adj chi2(2)   Prob>chi2
-------------+---------------------------------------------------------------
   residuals |        203     0.0769        0.1319        5.38         0.0680
```

$\text{ARIMA}(0,1,2)x(0,1,1)_{12}$ sin término constante

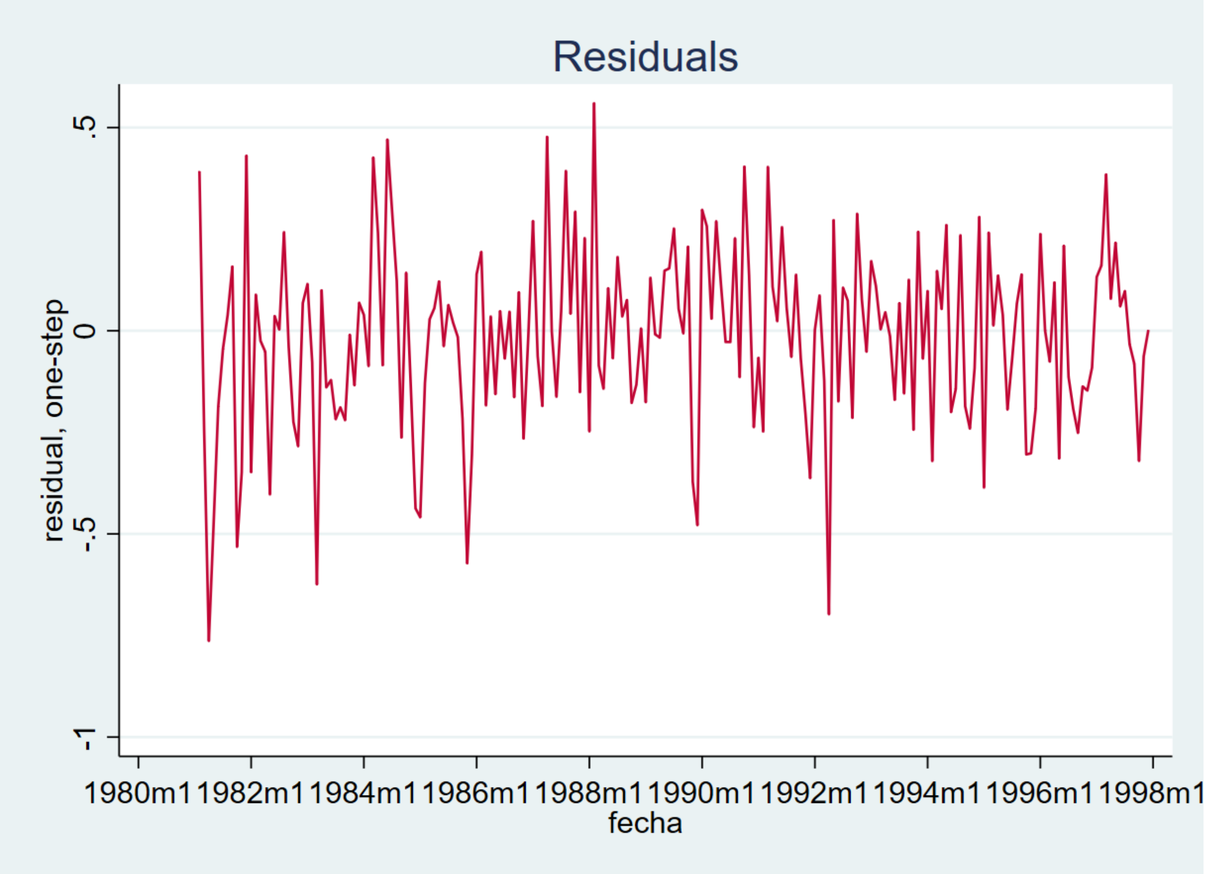
```
ARIMA regression

Sample:  1981m2 - 1997m12                       Number of obs     =        203
                                                Wald chi2(3)      =     515.70
Log likelihood =  20.37131                      Prob > chi2       =     0.0000

------------------------------------------------------------------------------
             |                 OPG
DS12.l_ozono |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
ARMA         |
          ma |
         L1. |  -.6733404   .0704767    -9.55   0.000    -.8114721   -.5352086
         L2. |  -.2229956   .0621202    -3.59   0.000     -.344749   -.1012422
-------------+----------------------------------------------------------------
ARMA12       |
          ma |
         L1. |  -.7472483   .0630166   -11.86   0.000    -.8707586    -.623738
-------------+----------------------------------------------------------------
      /sigma |   .2125416    .010357    20.52   0.000     .1922421     .232841
------------------------------------------------------------------------------


. ****CONTRASTE MEDIA RESIDUOS CERO
. ttest residuals == 0

One-sample t test
------------------------------------------------------------------------------
Variable |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
residu~s |     203   -.0186356    .0156867    .2235012   -.0495663    .0122951
------------------------------------------------------------------------------
    mean = mean(residuals)                                        t =  -1.1880
Ho: mean = 0                                     degrees of freedom =      202

    Ha: mean < 0                 Ha: mean != 0                 Ha: mean > 0
 Pr(T < t) = 0.1181         Pr(|T| > |t|) = 0.2362          Pr(T > t) = 0.8819
```

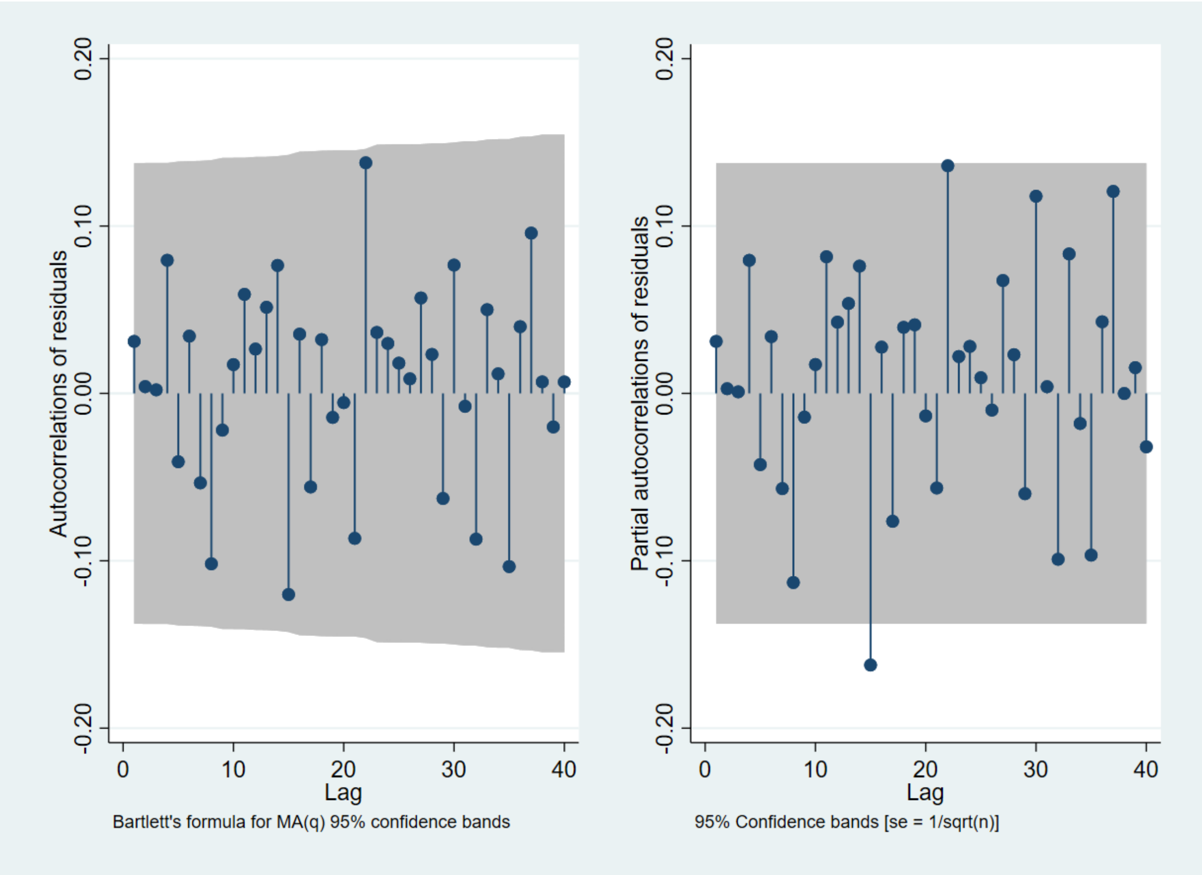

Residuals
residual, one-step
.5
0
-.5
-1
1980m1 1982m1 1984m1 1986m1 1988m1 1990m1 1992m1 1994m1 1996m1 1998m1
fecha

ARIMA$(0,1,2)x(0,1,1)_{12}$ sin término constante, con el rango de datos 1980-1996

```
ARIMA regression

Sample:  1981m2 - 1996m12                       Number of obs     =        191
                                                Wald chi2(3)      =     401.78
Log likelihood =  16.77706                      Prob > chi2       =     0.0000

------------------------------------------------------------------------------
             |                 OPG
DS12.l_ozono |      Coef.   Std. Err.      z    P>|z|     [95% Conf. Interval]
-------------+----------------------------------------------------------------
ARMA         |
          ma |
         L1. |  -.6793992   .0722794    -9.40   0.000    -.8210642   -.5377342
         L2. |   -.194693   .0651152    -2.99   0.003    -.3223165   -.0670695
-------------+----------------------------------------------------------------
ARMA12       |
          ma |
         L1. |  -.7618375   .0629149   -12.11   0.000    -.8851485   -.6385265
-------------+----------------------------------------------------------------
      /sigma |   .2146558   .0108324    19.82   0.000     .1934246     .235887
------------------------------------------------------------------------------
```

```
. ****CONTRASTE MEDIA RESIDUOS CERO
. ttest residuals == 0

One-sample t test
------------------------------------------------------------------------------
Variable |     Obs        Mean    Std. Err.   Std. Dev.   [95% Conf. Interval]
---------+--------------------------------------------------------------------
residu~s |     203   -.0186054    .0157075    .2237969   -.0495771    .0123662
------------------------------------------------------------------------------
    mean = mean(residuals)                                        t =  -1.1845
Ho: mean = 0                                     degrees of freedom =      202

    Ha: mean < 0                 Ha: mean != 0                 Ha: mean > 0
 Pr(T < t) = 0.1188         Pr(|T| > |t|) = 0.2376          Pr(T > t) = 0.8812
```


Residuals
residual, one-step
.5
0
-.5
-1
1980m1 1982m1 1984m1 1986m1 1988m1 1990m1 1992m1 1994m1 1996m1 1998m1
fecha


Autocorrelations of residuals
0.20
0.10
0.00
-0.10
-0.20
0 10 20 30 40
Lag
Bartlett's formula for MA(q) 95% confidence bands
Partial autocorrelations of residuals
0.20
0.10
0.00
-0.10
-0.20
0 10 20 30 40
Lag
95% Confidence bands [se = 1/sqrt(n)]

```
. corrgram residuals

                                          -1       0       1 -1       0       1
 LAG       AC       PAC      Q     Prob>Q  [Autocorrelation]  [Partial Autocor]

1        0.0311   0.0311    .199  0.6555
2        0.0041   0.0028  .20243  0.9037
3        0.0021   0.0009  .20338  0.9770
4        0.0796   0.0795  1.5272  0.8218
5       -0.0408  -0.0425  1.8774  0.8658
6        0.0342   0.0339  2.1248  0.9079
7       -0.0535  -0.0569  2.7316  0.9087
8       -0.1018  -0.1130  4.9437  0.7636
9       -0.0219  -0.0142  5.0469  0.8302
10       0.0172   0.0173  5.1108  0.8837
11       0.0592   0.0817  5.8703  0.8819
12       0.0265   0.0426  6.0233  0.9149
13       0.0515   0.0537  6.6032  0.9214
14       0.0765   0.0761  7.8913  0.8949
15      -0.1201  -0.1623  11.087  0.7464
16       0.0354   0.0276  11.366  0.7864
17      -0.0559  -0.0764  12.065  0.7962
18       0.0322   0.0395  12.297  0.8315
19      -0.0143   0.0410  12.344  0.8704
20      -0.0055  -0.0135  12.351  0.9035
21      -0.0866  -0.0565  14.065  0.8668
22       0.1378   0.1360   18.43  0.6802
23       0.0364   0.0221  18.737  0.7164
24       0.0299   0.0281  18.945  0.7549
25       0.0181   0.0094  19.021  0.7961
26       0.0087  -0.0100  19.039  0.8348
27       0.0570   0.0675  19.808  0.8387
28       0.0233   0.0232  19.937  0.8667
29      -0.0628  -0.0599   20.88  0.8635
30       0.0767   0.1178  22.294  0.8431
31      -0.0077   0.0040  22.309  0.8731
32      -0.0871  -0.0991  24.154  0.8388
33       0.0500   0.0833  24.767  0.8481
34       0.0117  -0.0180  24.801  0.8755
35      -0.1034  -0.0966  27.452  0.8147
36       0.0399   0.0428  27.848  0.8326
37       0.0958   0.1207  30.148  0.7802
38       0.0069  -0.0001   30.16  0.8138
39      -0.0200   0.0154  30.261  0.8408
40       0.0069  -0.0319  30.273  0.8675
```

```
. ***CONTRASTES DE NORMALIDAD
. sktest residuals

                    Skewness/Kurtosis tests for Normality
                                                         ------ joint ------
    Variable |        Obs  Pr(Skewness)  Pr(Kurtosis) adj chi2(2)   Prob>chi2
-------------+---------------------------------------------------------------
   residuals |        203     0.0583        0.1293        5.77        0.0558
```

Valores del IC y del ECM

```
. list upper lower

     +-----------------------+
     |    upper        lower |
     |-----------------------|
205. | .6937841    -.1476909 |
     |-----------------------|
206. | .9532523     .1117805 |
207. | 1.167652      .326181 |
208. | 1.440223      .598753 |
209. | 1.457124     .6156537 |
210. |  1.58425     .7427801 |
     |-----------------------|
211. | 1.709539     .8680701 |
212. |  1.78923     .9477607 |
213. | 1.633546     .7920775 |
214. | 1.482837     .6413687 |
215. | .9536002     .1121319 |
     |-----------------------|
216. | .6807722     -.160696 |
```

```
     | (ECM) mse |
     |-----------|
205. | -.2346974 |
     |-----------|
206. |  .0216907 |
207. | -.0415408 |
208. | -.4220688 |
209. |  .1609198 |
210. | -.4665422 |
     |-----------|
211. |  .0650292 |
212. |  .0410413 |
213. |  .0784813 |
214. |  .0076618 |
215. |  .2372159 |
     |-----------|
216. |  .0777165 |
```