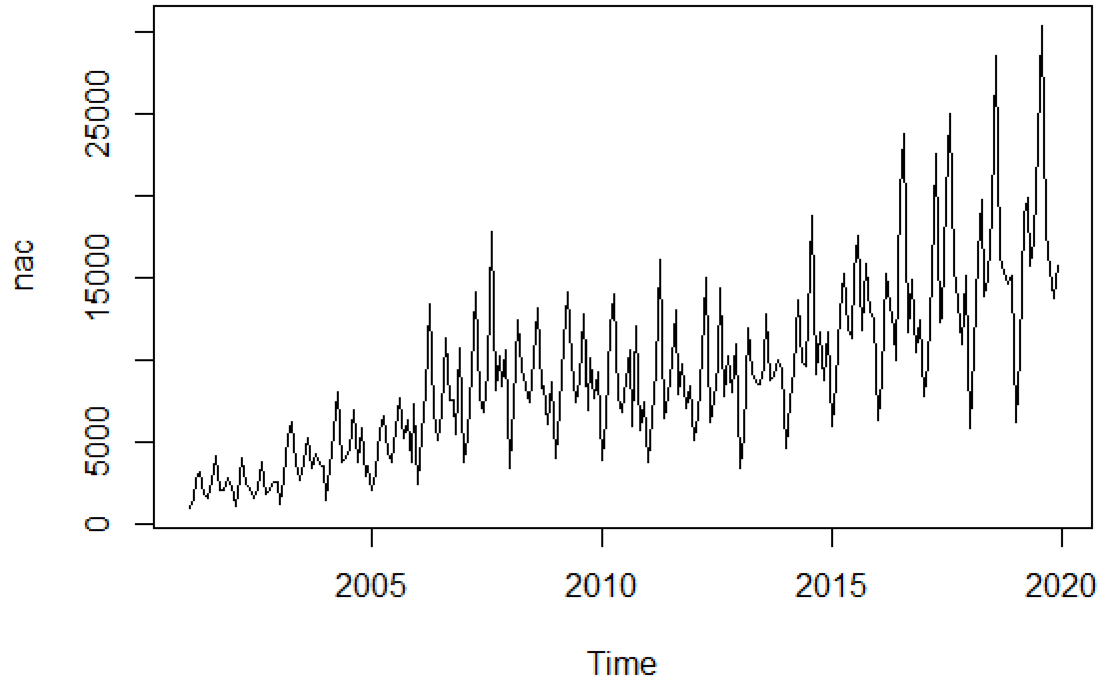
a)



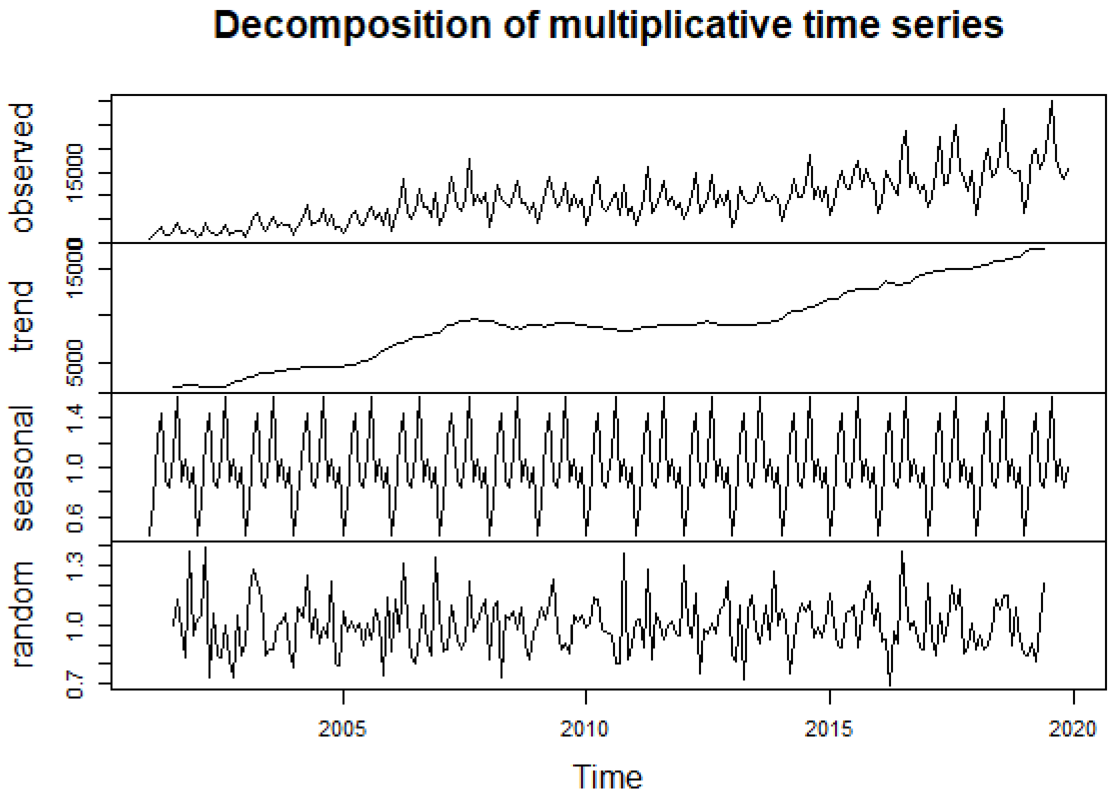
Podemos observar que la serie presenta una tendencia lineal creciente pues a medida que aumenta el tiempo el valor medio de la serie va aumentando.

Por otra parte, se puede dilucidar cierta estacionalidad pues los meses de verano observamos un aumento en la cantidad de turistas, concretamente en los meses de Julio y Agosto, curiosamente en los meses de Marzo y Abril también podemos observar un aumento de turistas. Por otra parte, los meses de enero y febrero son los que presentan menor afluencia.

Por último, podemos observar cierta variabilidad en la dispersión, a medida que vamos avanzando en el tiempo la variabilidad va aumentando.

Al observar que la variabilidad de la serie aumenta a medida que nos desplazamos en el tiempo resulta conveniente descomponer la serie a través de un modelo multiplicativo, donde $X_t=T_t*S_t*I_t$.

Aplicando un método de descomposición clásico obtenemos:



Para confirmar que estamos ante un modelo multiplicativo podemos observar la componente irregular, que oscila alrededor del valor 1 en el tiempo sin grandes variaciones, esto significa que hemos elegido bien el modelo.

Por otra parte, podemos observar que la tendencia sigue un comportamiento lineal creciente, como sospechábamos al observar la serie. Podemos observar que en la mitad de la serie el crecimiento de la tendencia es prácticamente nulo, retomándose en la tercera mitad de la serie.

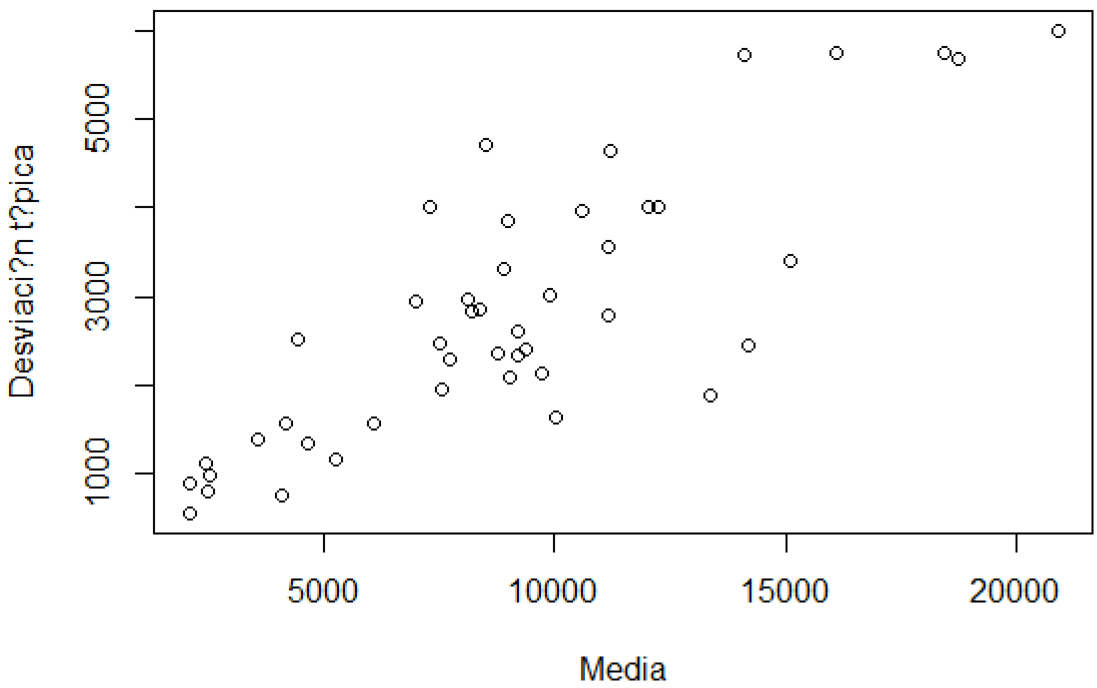
Por otra parte, la estacionalidad presenta 2 picos, el primero de ellos en los meses de Marzo y Abril y el segundo, ligeramente mayor, en Julio y Agosto y muestra un valle en los meses de Enero y Febrero.

b)

Recordemos que para que una serie sea estacionaria esta no debe de presentar tendencia ni estacionalidad ni dispersión en la variabilidad, la serie estudiada presenta tanto tendencia como estacionalidad y dispersión en la varianza, por lo tanto, concluimos que no es estacionaria.

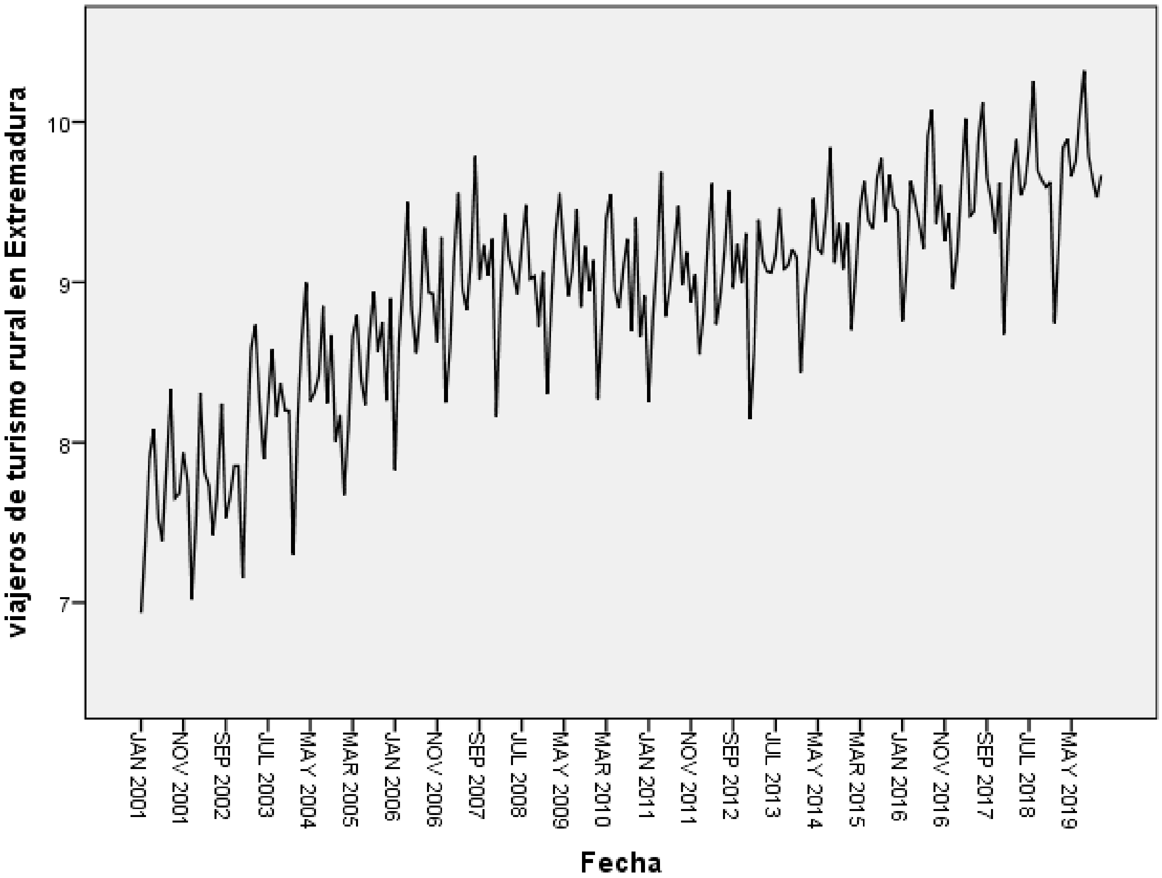
Para convertir la serie en una serie estacionaria lo primero que debemos de tratar es la dispresión en la variabilidad, para ello aplicaremos una transformación de Box-Cox, es decir, supondremos que la media y la varianza de la serie están relacionadas a través de la función $f(x) = k*x^\alpha$ y por tanto al aplicar una transformación logarítmica a la serie obtendremos que $\log(\text{media}) = \log(k) + \alpha*\log(\text{varianza})$.

También podemos estudiar la relación entre la media y la varianza a través de un gráfico de dispersión:



En el podemos observar que la media y la desviación típica muestran una relación lineal directa, de hecho, el valor de $\lambda = 1 - \alpha = 0.87$, es decir, el valor de α es próximo a 0, por lo que resulta necesario una transformación logarítmica.

Al aplicar el logaritmo a la serie la serie resultante es:

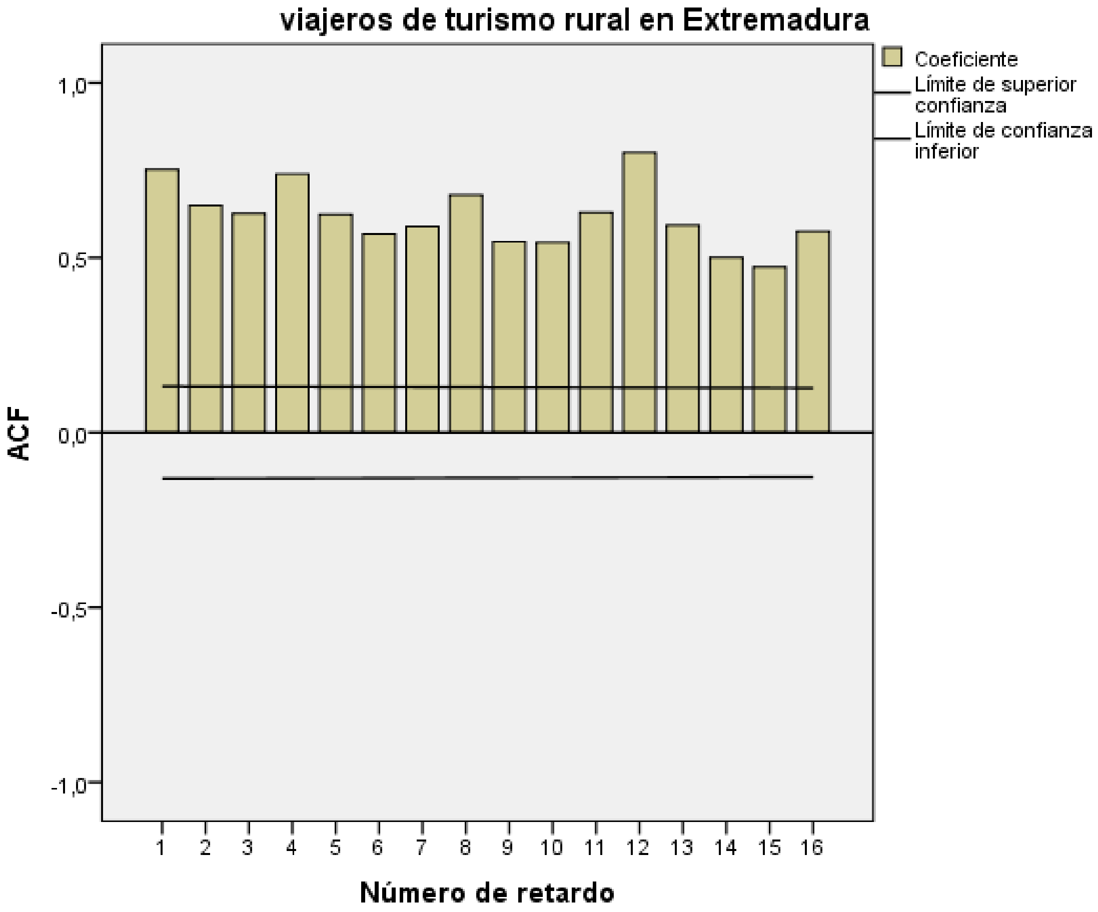


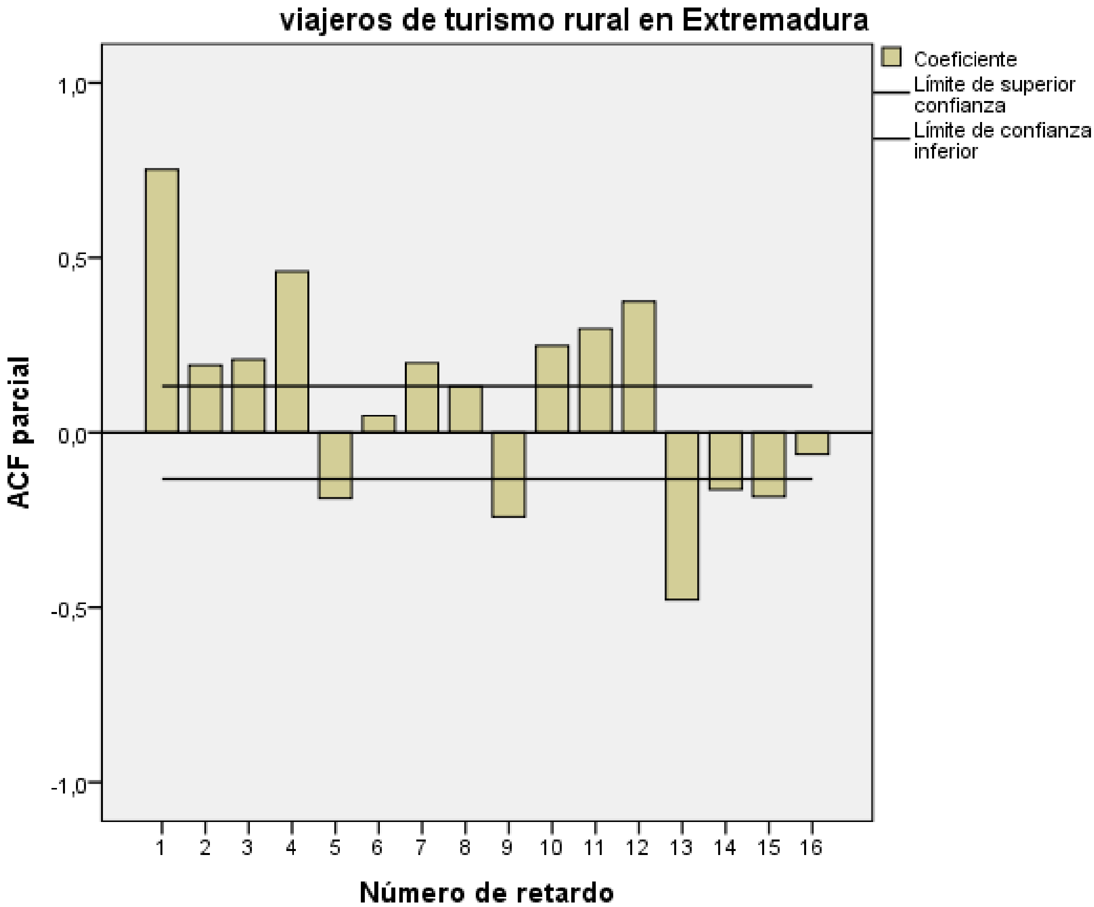
Transformaciones: logaritmo natural

La cual podemos observar que es estable en varianza.

Por otra parte, para eliminar la tendencia lineal de la serie aplicaremos una diferenciación ordinaria de primer orden a la serie Log(X_t), es decir, consideraremos la nueva serie ∇ Log(X_t) = Log(X_t) - Log(X_t-1).

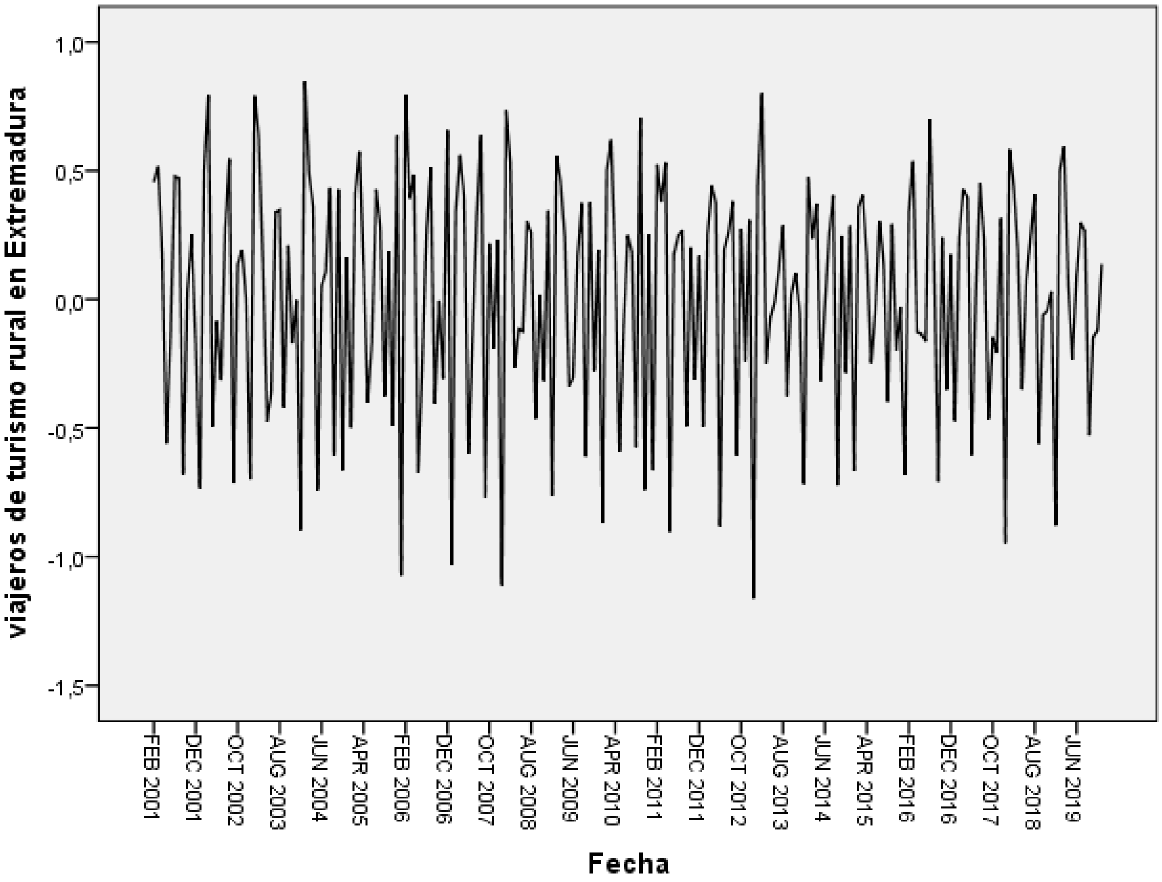
Podemos observar la existencia de tendencia en la serie de varias maneras, la primera y la mas común es a través de la representación gráfica de la serie, en la que suele ser relativamente fácil identificar la existencia de tendencia, otra forma es representar la FAS y FAP del modelo:


viajeros de turismo rural en Extremadura
Coeficiente
Límite de superior confianza
Límite de confianza inferior
ACF
1,0
0,5
0,0
-0,5
-1,0
1 2 3 4 5 6 7 8 9 10 11 12 13 14 15 16
Número de retardo



Si en la FAS observamos un decaimiento lineal de los retardos y en la FAP observamos que el primer retardo es próximo a 1 tenemos claros indicios de la existencia de tendencia.

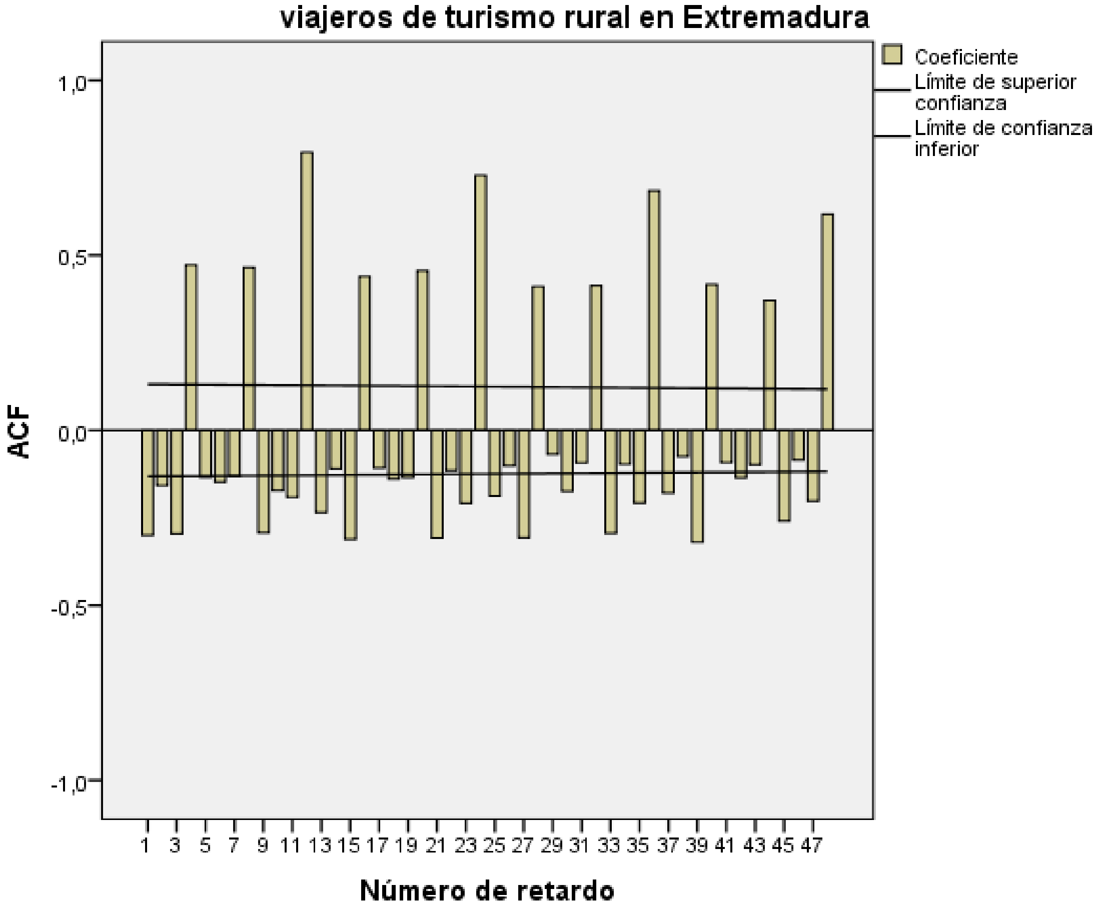
Al aplicar la transformación ordinaria de primer orden la serie resultante es:



Transformaciones: logaritmo natural, diferencia(1)

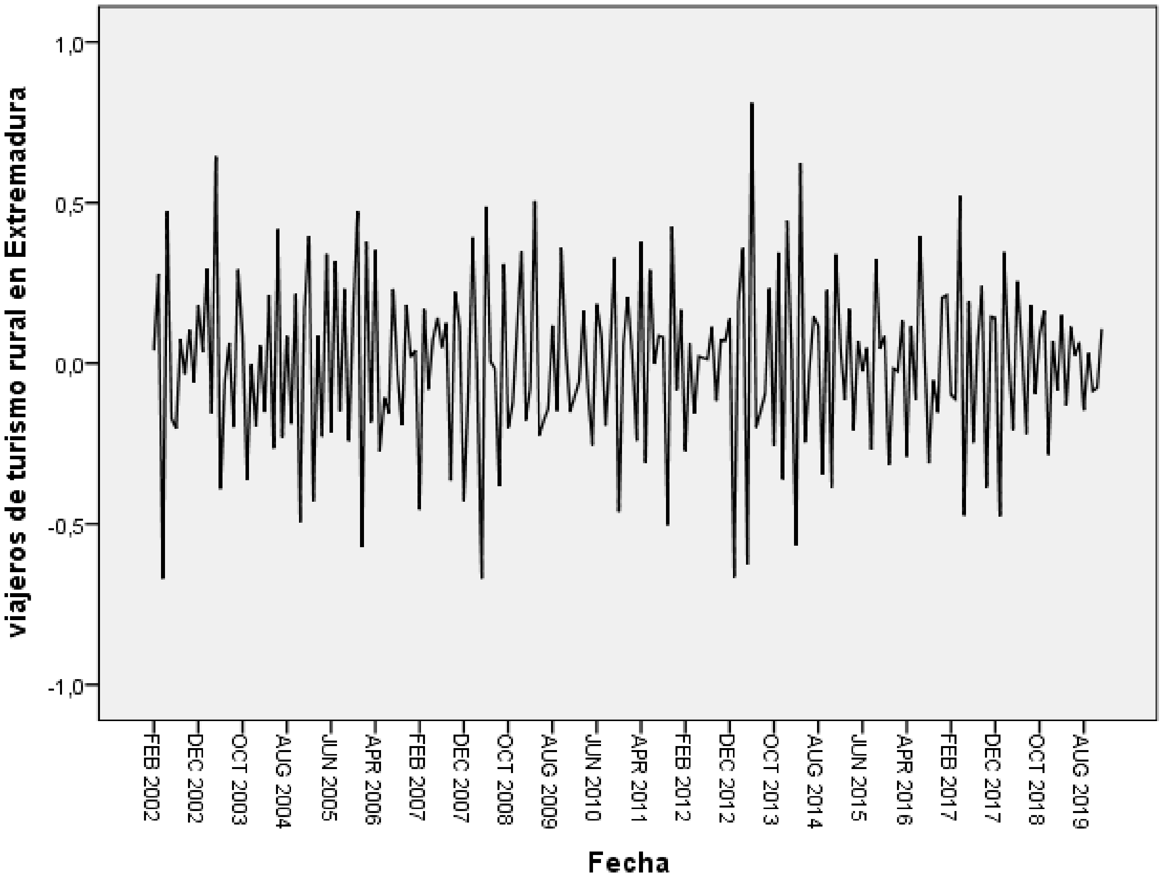
Por último, para eliminar la estacionalidad aplicaremos una diferenciación estacional de periodo s= 12 a la serie Log(X_t), consideramos la nueva serie $\nabla^{12}\nabla$ Log(X_t)

Para estudiar la existencia de estacionalidad en la serie podemos buscar la existencia de cierta variación periódica en la serie original o podemos estudiar la FAS:



En ella podemos observar un decaimiento lineal en los retardos del periodo (12,24,36...), un claro indicio de estacionalidad.

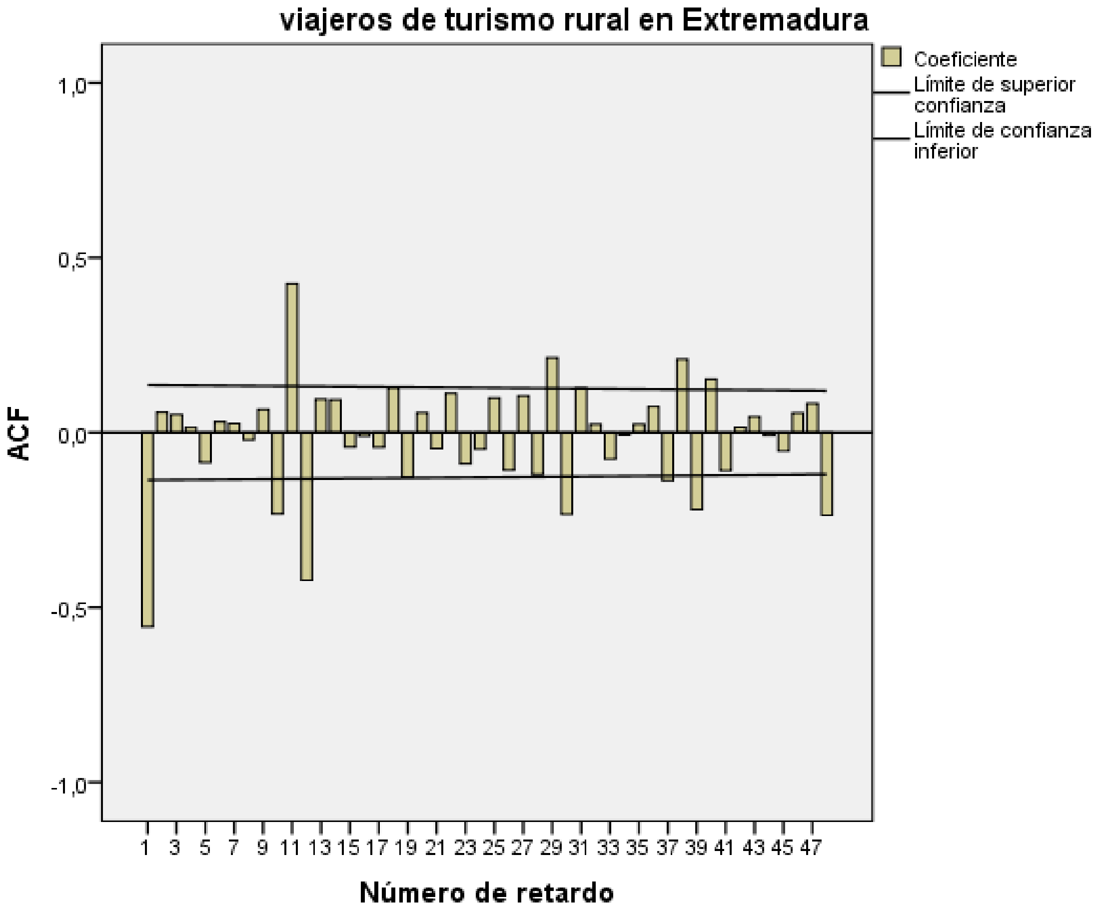
Tras aplicar la diferenciación estacional obtenemos la serie:

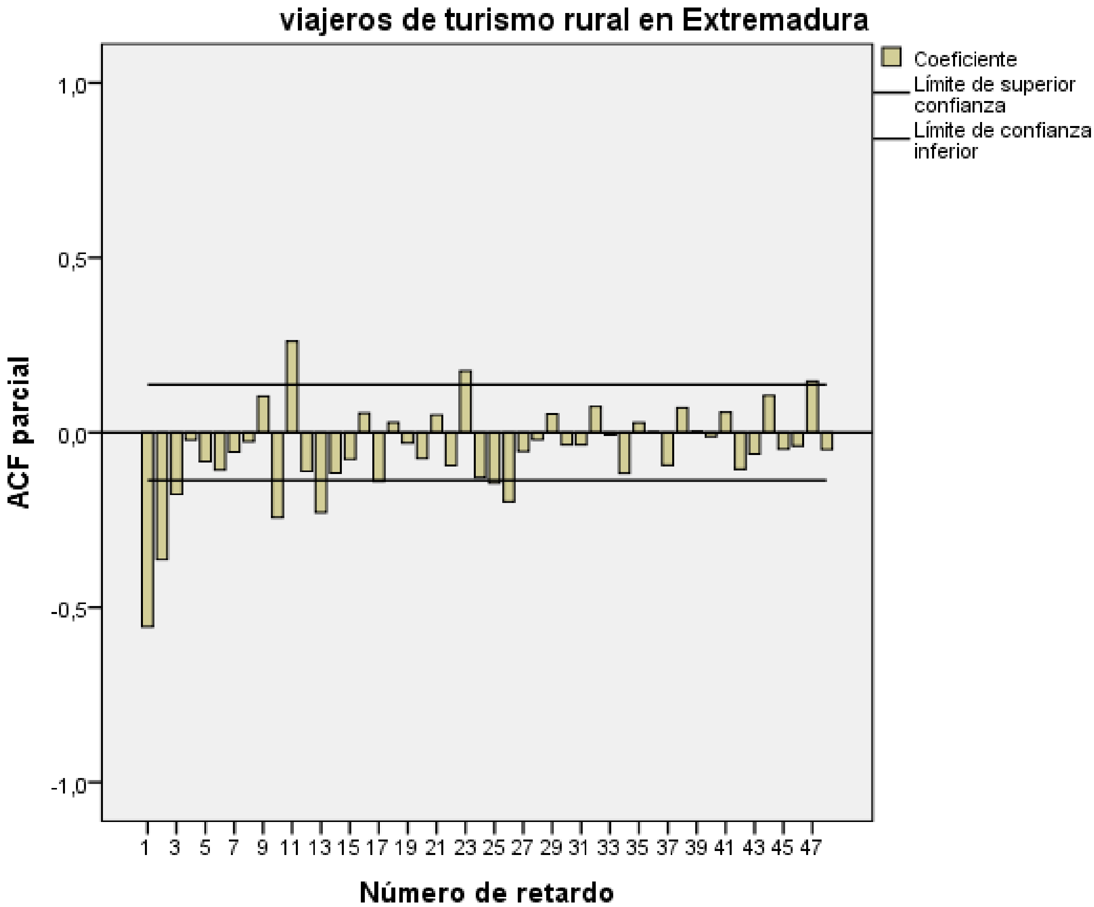


Transformaciones: logaritmo natural, diferencia(1), diferencia estacional(1, periodo 12)

c)

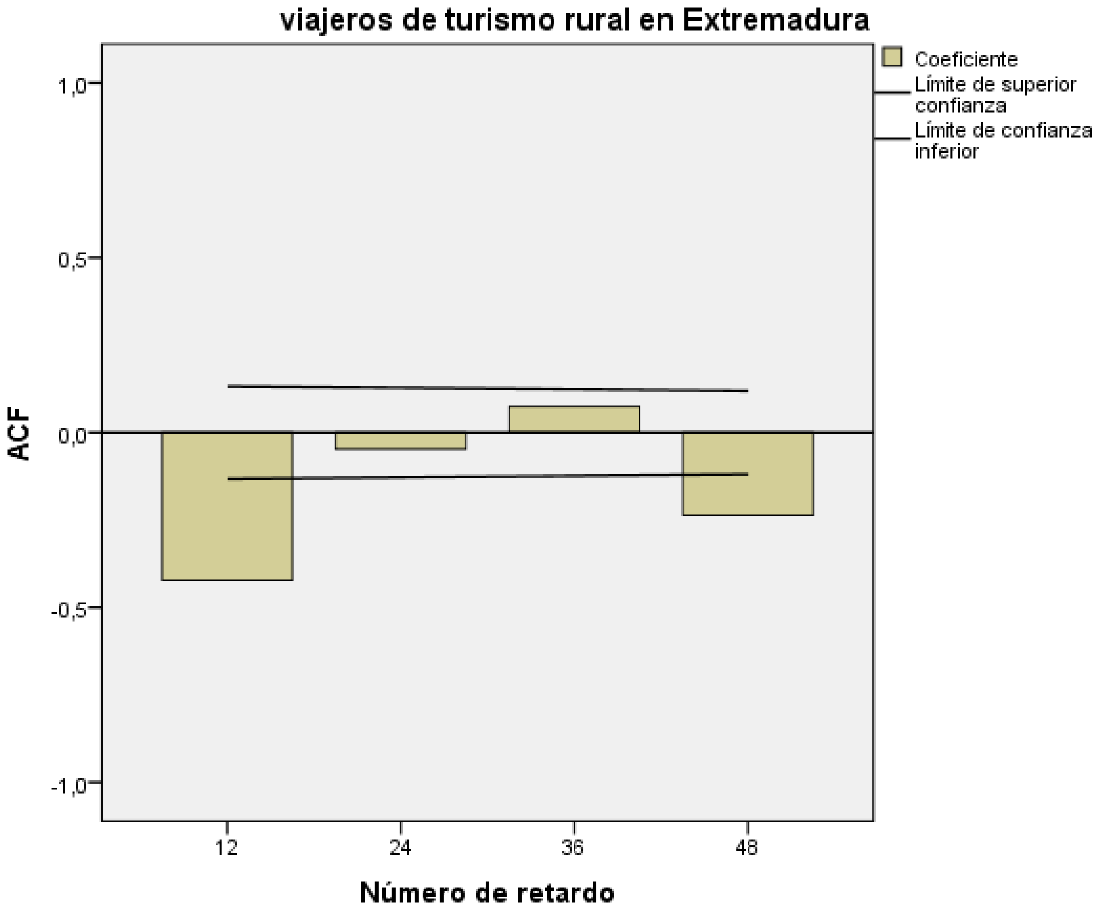
Para determinar el modelo ARIMA más verosímil nos ayudaremos de la FAS y FAP muestrales:

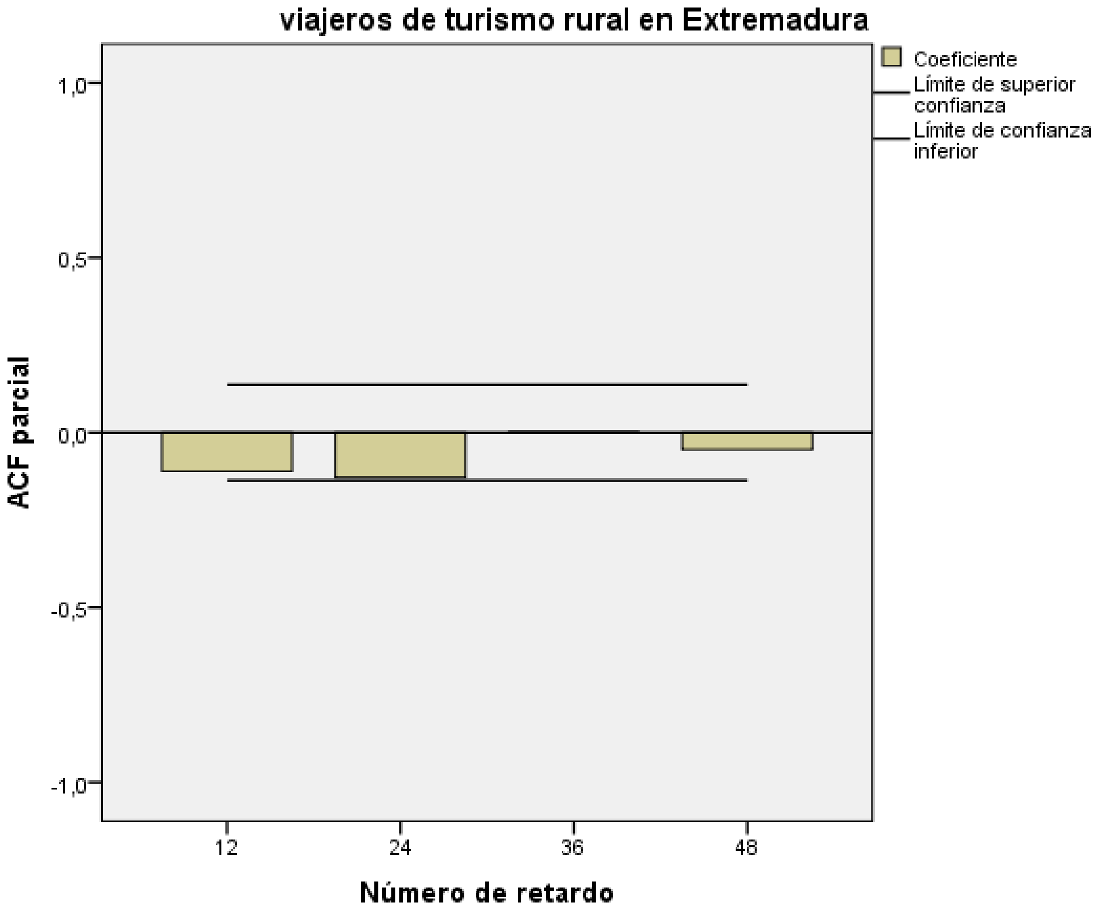

viajeros de turismo rural en Extremadura
Coeficiente
Límite de superior confianza
Límite de confianza inferior
1,0
0,5
0,0
-0,5
-1,0
ACF
1 3 5 7 9 11 13 15 17 19 21 23 25 27 29 31 33 35 37 39 41 43 45 47
Número de retardo



Parte NO estacional:

Para estudiar la parte no estacional del modelo nos fijaremos únicamente en los retardos bajos de las FAS y FAP, en la FAS podemos observar que el primer retardo es significativo, mientras que el segundo y tercero son no significativos, esto nos invita a pensar que se trata de un MA(1) para corroborar nuestras sospecha observamos la FAP, vemos un decrecimiento exponencial en los primeros retardos, lo cual concuerda con un modelo MA(1), al ser los primeros 3 retardos no significativos también podemos sospechar de un AR(3), pero al observar la FAS no vemos un decaimiento exponencial o sinusoidal en los primeros retardos, lo cual debilita la hipótesis del AR(3)


viajeros de turismo rural en Extremadura
Coeficiente
Límite de superior confianza
Límite de confianza inferior
ACF
1,0
0,5
0,0
-0,5
-1,0
12
24
36
48
Número de retardo

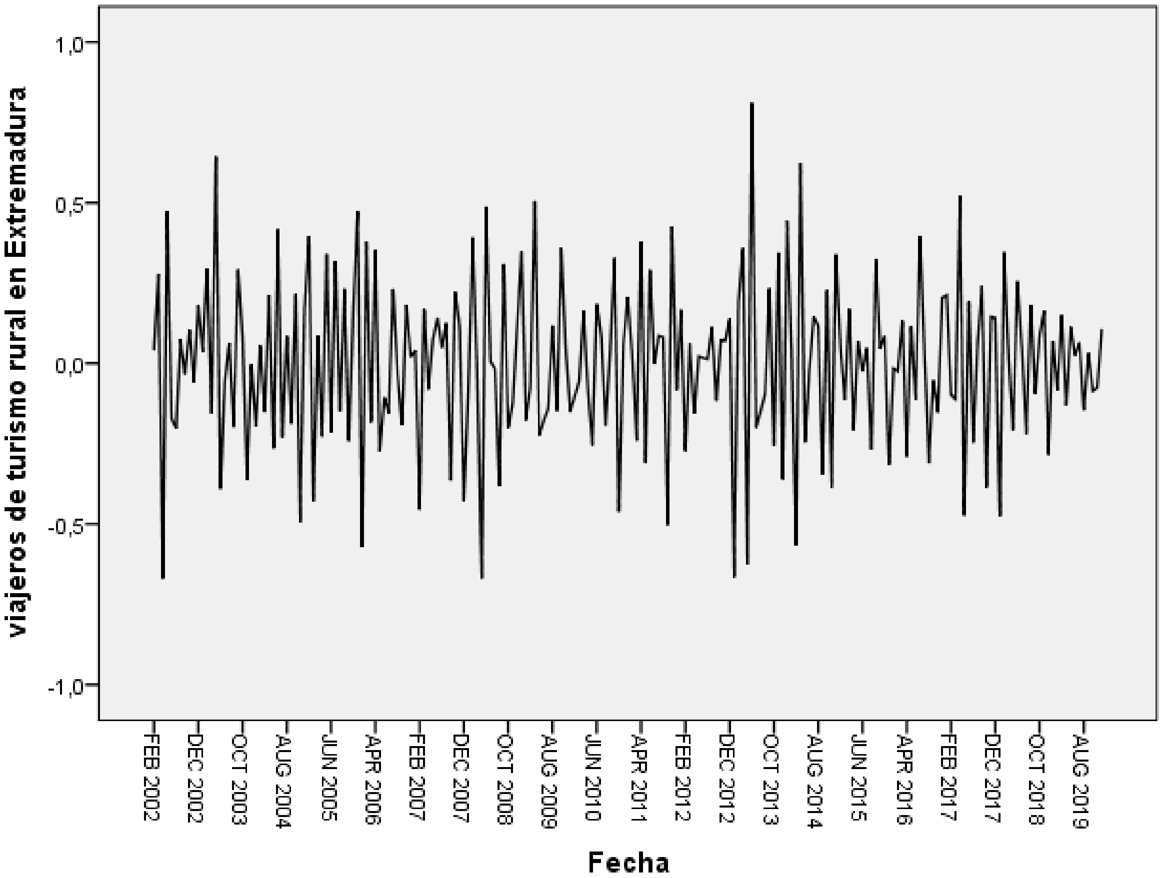


Parte estacional:

Para estudiar la parte estacional solo nos fijaremos en los retardos del periodo, 12,24,36,48…. En la FAS podemos observar que el retardo 12 es significativo mientras que el 24 y 36 no, esto nos hace sospechar de un MA(1)_12, por otra parte, en la FAP no observamos ningún retardo significativo ni ninguna tendencia clara.

Valores atípicos:

Para estudiar los valores atípicos debemos de fiarnos en la serie transformada:



Transformaciones: logaritmo natural, diferencia(1), diferencia estacional(1, periodo 12)

En ella podemos observar un pico entre 2012 y 2013, el cual puede deberse a un valor atípico aditivo, también pueden observarse más picos a lo largo de la serie, pero no son tan notorios como el señalado anteriormente.

Pasamos al estudio de los posibles modelos:

Primero valoramos el modelo que mejor se adapta a los datos, el ARIMA(0,1,1)xARIMA(0,1,1), los resultados se muestran a continuación, notar que se ha eliminado la constante del modelo al resultar no significativa.

Estadísticos del modelo

Modelo	Número de predictores	Estadísticos de ajuste del modelo		Ljung-Box Q(18)			Número de valores atípicos
		R cuadrado estacionaria	BIC normalizado	Estadísticos	DF	Sig.	
viajeros de turismo rural en Extremadura-Modelo_1	0	,602	14,868	22,189	16	,137	1

Parámetros del modelo ARIMA

					Estimación	SE	t	Sig.
viajeros de turismo rural en Extremadura-Modelo_1	viajeros de turismo rural en Extremadura	Logaritmo natural	Diferencia		1			
			MA	Retardo 1	,794	,043	18,361	,000
			Diferencia estacional		1			
			MA, estacional	Retardo 1	,758	,059	12,764	,000

Valores atípicos

			Estimación	SE	t	Sig.
viajeros de turismo rural en Extremadura-Modelo_1	Feb 2003	Cambio de nivel	,368	,096	3,835	,000

En la primera tabla observamos el valor del contraste Ljung-Box, que resulta significativo, es decir, podemos suponer que los residuos provienen de un ruido blanco. Por otra parte, en los coeficientes del modelo podemos observar que todos resultan significativos, indicio de que hemos ajustado bien el modelo. Por último, observamos un valor atípico en Febrero de 2003, un cambio de nivel, parece que no detecta el valor atípico aditivo señalado anteriormente.

Planteamos también el modelo ARIMA(3,1,0)xARIMA(0,1,1) y obtenemos:

Estadísticos del modelo

Modelo	Número de predictores	Estadísticos de ajuste del modelo		Ljung-Box Q(18)			Número de valores atípicos
		R cuadrado estacionaria	BIC normalizado	Estadísticos	DF	Sig.	
viajeros de turismo rural en Extremadura-Modelo_1	0	,573	14,936	21,187	14	,097	0

Parámetros del modelo ARIMA

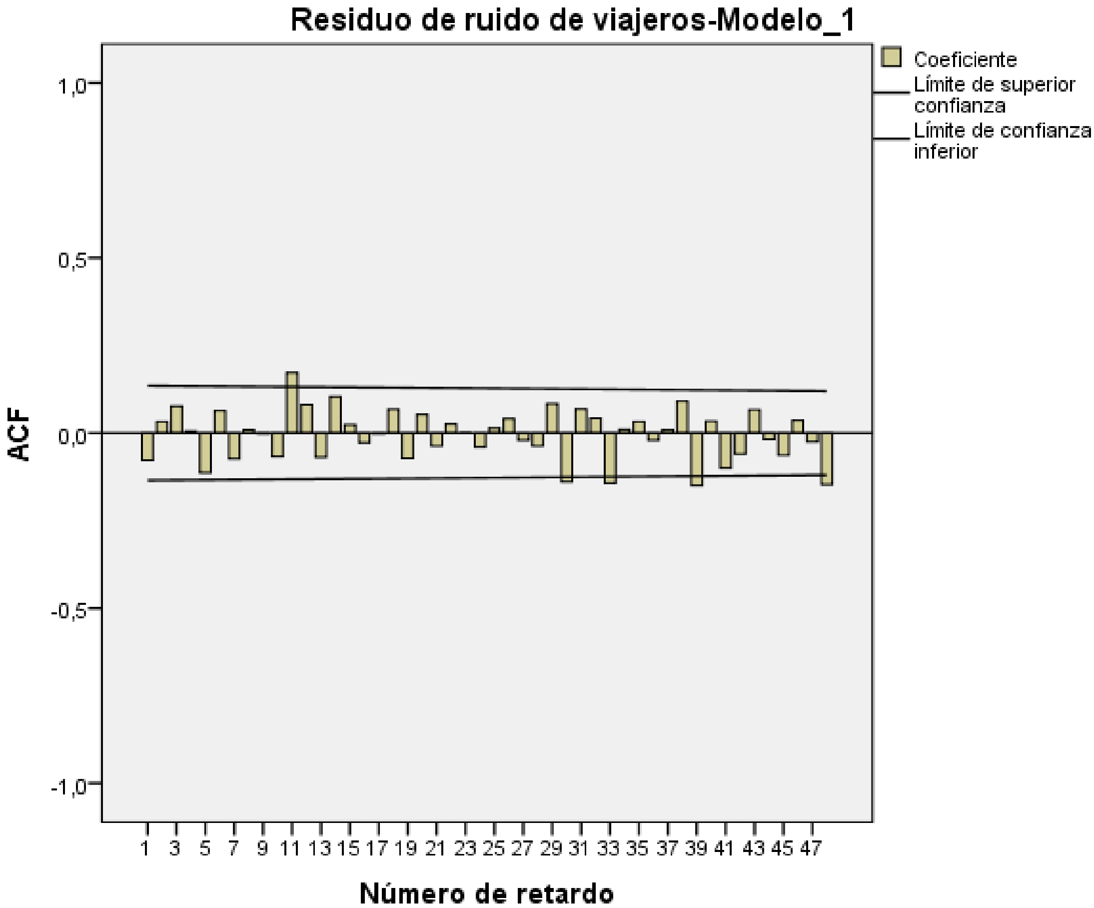
					Estimación	SE	t	Sig.
viajeros de turismo rural en Extremadura-Modelo_1	viajeros de turismo rural en Extremadura	Logaritmo natural	AR	Retardo 1	-,776	,069	-11,277	,000
				Retardo 2	-,480	,080	-6,024	,000
				Retardo 3	-,164	,068	-2,417	,017
			Diferencia		1			
			Diferencia estacional		1			
			MA, estacional	Retardo 1	,719	,060	12,037	,000

Tanto el contraste de Ljung-Box como el de cada uno de los parámetros resultan significativos, lo que quiere decir que el modelo también es factible, pero observamos que los p valores de los contrastes son mayores, es decir, tenemos una menor fiabilidad, además el BIC es mayor que el modelo anterior y hemos utilizado más parámetros que el modelo anterior (criterio de parsimonia). Todos estos indicios nos invitan a ajustar el otro modelo, de todas formas, tendremos este modelo en mente en el caso de que el otro no cumpla las hipótesis de validez.

Notar que este modelo no detecta ningún valor atípico.

Hipótesis de validez:

Incorreción de los residuos:



Estudiamos la incorrelación de forma gráfica a través de la FAS muestral de los residuos, en ella podemos observar que únicamente un retardo es significativo dentro de los primeros 30 retardos, por tanto, podemos permitirnos suponer incorrelación

Linealidad:

Descriptivos

			Estadístico	Error estándar
Residuo de ruido de viajeros-Modelo_1	Media		,00	,011
	95% de intervalo de confianza para la media	Límite inferior	-,02	
		Límite superior	,02	

Para ello basta que el valor medio de los residuos esté próximo a 0, como podemos observar en este caso la media de los residuos es 0 y el intervalo de confianza para la media contiene al 0, por tanto, podemos suponer que la media de los residuos es nula.

Normalidad:

Para ello nos basaremos tanto en el test de Shapiro-Wilk como en el de Kolmogorov:

Pruebas de normalidad

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Estadístico	gl	Sig.	Estadístico	gl	Sig.
Residuo de ruido de viajeros-Modelo_1	,057	215	,087	,989	215	,092

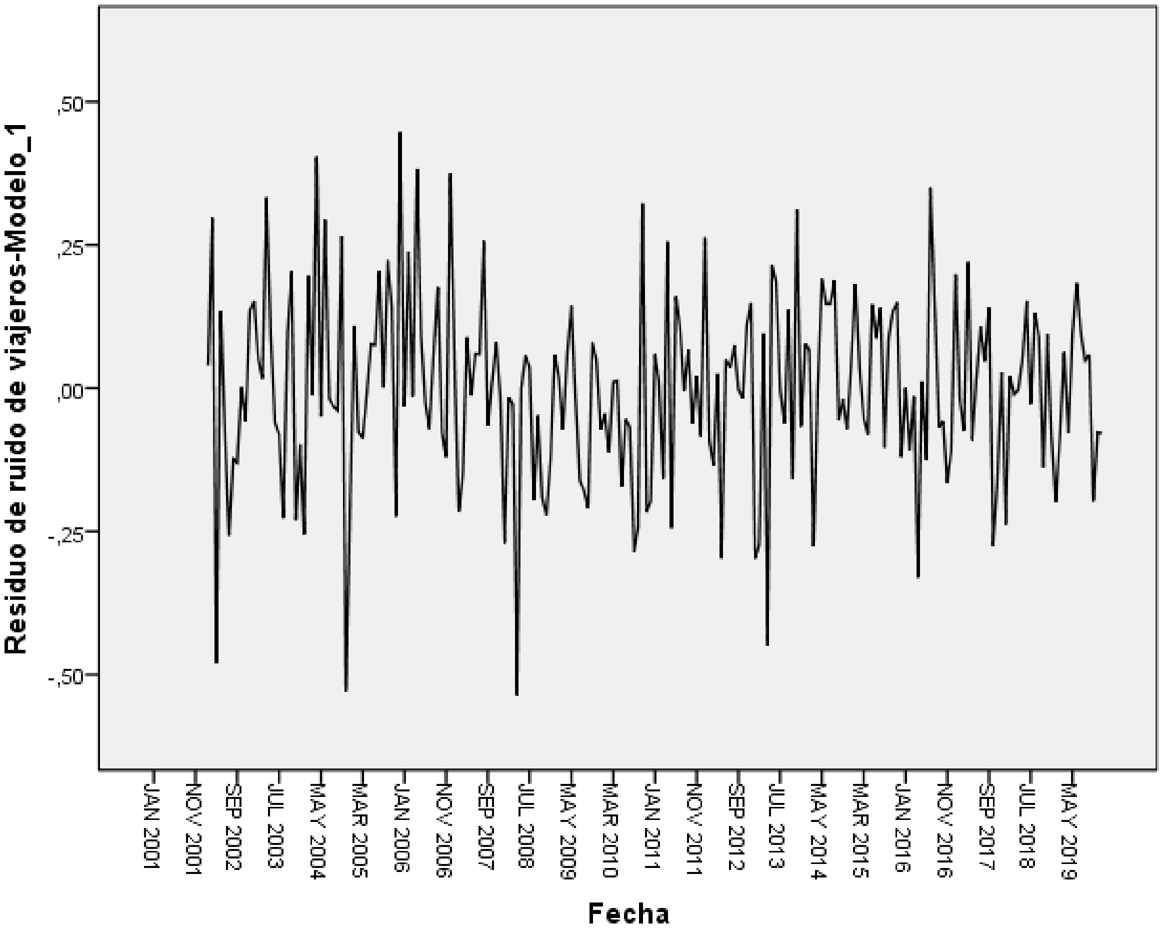
Ambos arrojan resultados no significativos, es decir, podmeo suponer la normalidad,

También podemos estudiar la normalidad a través del gráfico Q-Q plot:

En el cual observamos que los cuantiles muestrales de los residuos se ajustan muy bien a los teóricos, observamos 4 valores que no se ajustan muy bien al modelo normal, podría tratarse de atípicos aditivos, estudiaremos dichos valores una vez representemos el modelo.

Estabilidad en la variabilidad de los residuos:

Contrastaremos esta hipótesis de forma gráfica, representando el gráfico de los residuos:



Pese a observar una mayor variabilidad al principio de la serie que al final de la misma la diferencia no es muy grande, por lo tanto, podemos suponer que la dispersión de los residuos se mantiene constate.

Notar que resulta bastante fácil detectar los 4 valores atípicos que hemos observado en el gráfico Q-Q plot, los estudiaremos más adelante.

Tras estudiar las hipótesis de validez del modelo ajustado concluimos que todas ellas se cumplen y por tanto podemos utilizar el modelo para realizar predicciones.

Estudiemos el sobreajuste:

Para ello consideramos los 4 modelos

ARIMA(1,1,1)xARIMA(0,1,1):

Estadísticos del modelo

Modelo	Número de predictores	Estadísticos de ajuste del modelo: R cuadrado estacionaria	Estadísticos de ajuste del modelo: BIC normalizado	Ljung-Box Q(18): Estadísticos	Ljung-Box Q(18): DF	Ljung-Box Q(18): Sig.	Número de valores atípicos
viajeros de turismo rural en Extremadura-Modelo_1	0	,651	14,854	19,326	15	,199	3

Parámetros del modelo ARIMA

					Estimación	SE	t	Sig.
viajeros de turismo rural en Extremadura-Modelo_1	viajeros de turismo rural en Extremadura	Logaritmo natural	AR	Retardo 1	-,117	,081	-1,446	,150
			Diferencia		1			
			MA	Retardo 1	,791	,051	15,650	,000
			Diferencia estacional		1			
			MA, estacional	Retardo 1	,687	,061	11,263	,000

Observamos que el coeficiente AR(1) que hemos añadido resulta no significativo, es decir, sobra en el modelo, descartamos este modelo.

ARIMA(0,1,2)xARIMA(0,1,1):

Parámetros del modelo ARIMA

					Estimación	SE	t	Sig.
viajeros de turismo rural en Extremadura-Modelo_1	viajeros de turismo rural en Extremadura	Logaritmo natural	Diferencia		1			
			MA	Retardo 1	,909	,068	13,350	,000
				Retardo 2	-,097	,068	-1,432	,154
			Diferencia estacional		1			
			MA, estacional	Retardo 1	,685	,061	11,233	,000

Valores atípicos

				Estimación	SE	t	Sig.
viajeros de turismo rural en Extremadura-Modelo_1	Feb 2003	Cambio de nivel		,359	,083	4,336	,000
	Nov 2004	Transitorio	Magnitud	-,356	,097	-3,664	,000
			Factor de decrecimiento	,869	,078	11,112	,000
	Abr 2008	Innovador		-,555	,155	-3,588	,000

Volvemos a observar que el segundo coeficiente del MA(2) que hemos añadido resulta no significativo, es decir, sobra en el modelo. Descartamos el modelo

ARIMA(0,1,1)xARIMA(1,1,1):

Estadísticos del modelo

Modelo	Número de predictores	Estadísticos de ajuste del modelo		Ljung-Box Q(18)			Número de valores atípicos
		R cuadrado estacionaria	BIC normalizado	Estadísticos	DF	Sig.	
viajeros de turismo rural en Extremadura-Modelo_1	0	,607	14,903	19,435	15	,195	1

Parámetros del modelo ARIMA

					Estimación	SE	t	Sig.
viajeros de turismo rural en Extremadura-Modelo_1	viajeros de turismo rural en Extremadura	Logaritmo natural	Diferencia		1			
			MA	Retardo 1	,804	,043	18,737	,000
			AR, estacional	Retardo 1	,163	,097	1,678	,095
			Diferencia estacional		1			
			MA, estacional	Retardo 1	,846	,082	10,365	,000

El coeficiente añadido al modelo resulta no significativo, sobra. Tambien descartamos este modelo.

ARIMA(0,1,1)xARIMA(0,1,2):

Estadísticos del modelo

Modelo	Número de predictores	Estadísticos de ajuste del modelo		Ljung-Box Q(18)			Número de valores atípicos
		R cuadrado estacionaria	BIC normalizado	Estadísticos	DF	Sig.	
viajeros de turismo rural en Extremadura-Modelo_1	0	,607	14,902	19,363	15	,198	1

Parámetros del modelo ARIMA

					Estimación	SE	t	Sig.
viajeros de turismo rural en Extremadura-Modelo_1	viajeros de turismo rural en Extremadura	Logaritmo natural	Diferencia		1			
			MA	Retardo 1	,803	,043	18,733	,000
			Diferencia estacional		1			
			MA, estacional	Retardo 1	,680	,076	8,921	,000
				Retardo 2	,129	,075	1,718	,087

Valores atípicos

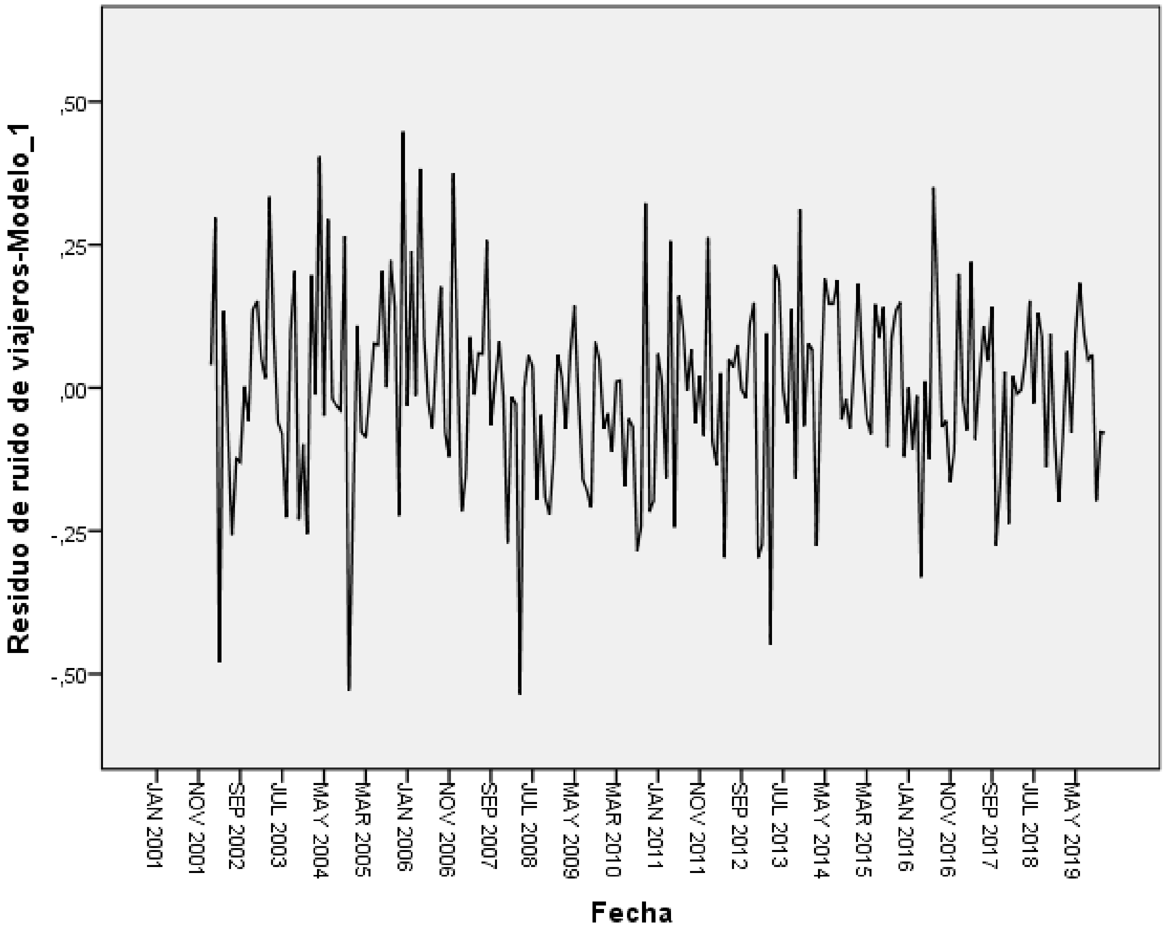
			Estimación	SE	t	Sig.
viajeros de turismo rural en Extremadura-Modelo_1	Feb 2003	Cambio de nivel	,375	,093	4,044	,000

El coeficiente añadido resulta significativo, descartamos el modelo.

Por tanto, concluimos que el modelo que mejor se ajusta a los datos es el:

ARIMA(0,1,1)xARIMA(0,1,1) con un cambio de nivel en el dato de la serie nº 26 (febrero 2003)

Como comentamos anteriormente en la gráfica de los residuos parecen existir más valores atípicos, de naturaleza aditiva.



El primer valor atípico se produce en Abril de 2002, el dato nº 16 de la serie, el segundo descenso abrupto y puntual se produce en Noviembre de 2004 el dato nº47 de la serie, el 3º valor atípico se encuentra en Abril de 2008, el dato nº 88 de la serie.

Todos estos valores atípicos parecen ser aditivos pues se trata de una caída brusca de los residuos que recupera rápidamente su valor normal y no tiene ningún efecto a la larga en la serie, añadiendo estos valores atípicos al modelo obtenemos:

Estadísticos del modelo

Modelo	Número de predictores	Estadísticos de ajuste del modelo		Ljung-Box Q(18)			Número de valores atípicos
		R cuadrado estacionaria	BIC normalizado	Estadísticos	DF	Sig.	
viajeros de turismo rural en Extremadura-Modelo_1	0	,591	15,006	19,644	16	,237	4

Parámetros del modelo ARIMA

					Estimación	SE	t	Sig.
viajeros de turismo rural en Extremadura-Modelo_1	viajeros de turismo rural en Extremadura	Sin transformación	Diferencia		1			
			MA	Retardo 1	,851	,038	22,493	,000
			Diferencia estacional		1			
			MA, estacional	Retardo 1	,434	,068	6,354	,000

Valores atípicos

			Estimación	SE	t	Sig.
viajeros de turismo rural en Extremadura-Modelo_1	Abr 2002	Aditivo	-2251,229	1421,369	-1,584	,115
	Feb 2003	Cambio de nivel	675,603	822,982	,821	,413
	Nov 2004	Aditivo	-519,331	1367,386	-,380	,704
	Abr 2008	Aditivo	-4948,724	1372,772	-3,605	,000

Es decir, nuestra intuición nos ha fallado, dichos valores atípicos no resultan significativos, de hecho, hacen que el cambio de nivel detectado anteriormente deje de ser significativo por tanto deberíamos ir poniendo 1 a 1 los valores atípicoa aditivos junto al cambio de nivel para ver si alguno resulta significativo, el SPSS me está dando error:

Advertencias

MODEL bloque 1. El subcomando OUTLIER 2: LOCATION tiene un valor perdido o no válido. La ejecución de este comando se detiene.

No encuentro la solución y voy justo de tiempo. La representación del modelo es:

Log(X_t) = 0.368*S_t,26 + ((0.794*B + 0.758*B^12)/((1-B)*(1-B^12)))*Z_t

Ó

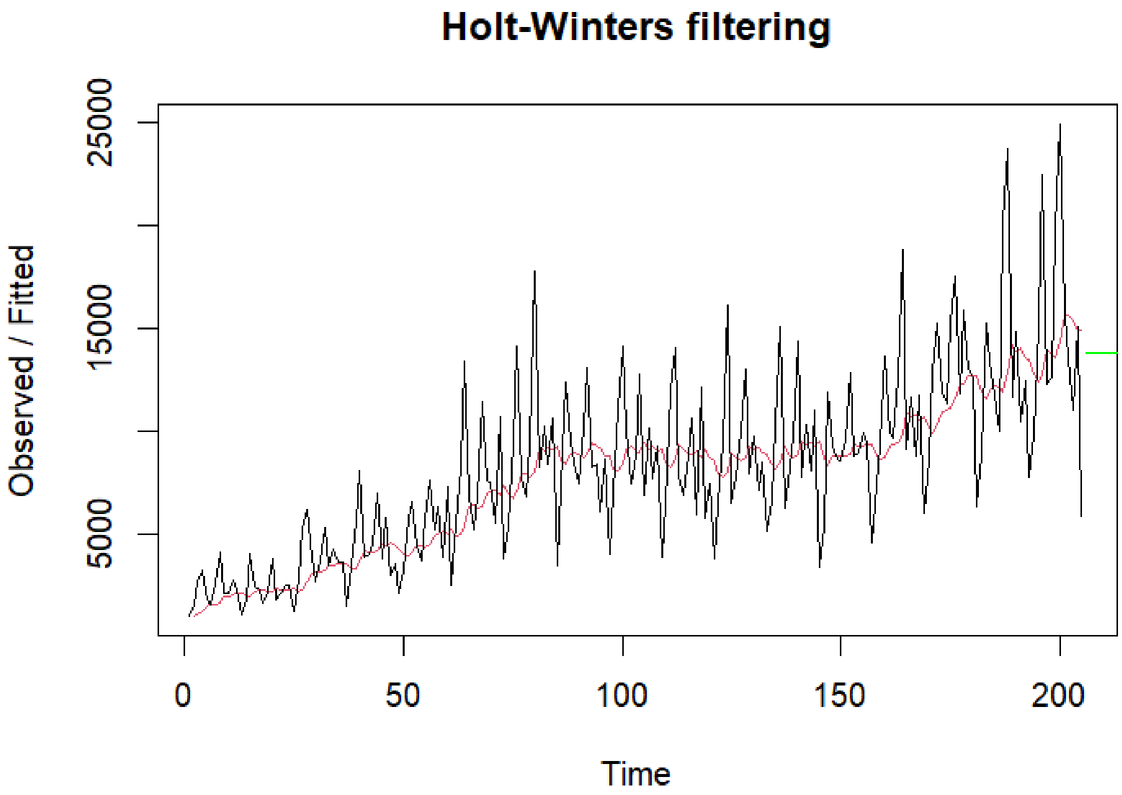
X_t = exp{0.368*S_t,26 + ((0.794*B + 0.758*B^12)/((1-B)*(1-B^12)))*Z_t}

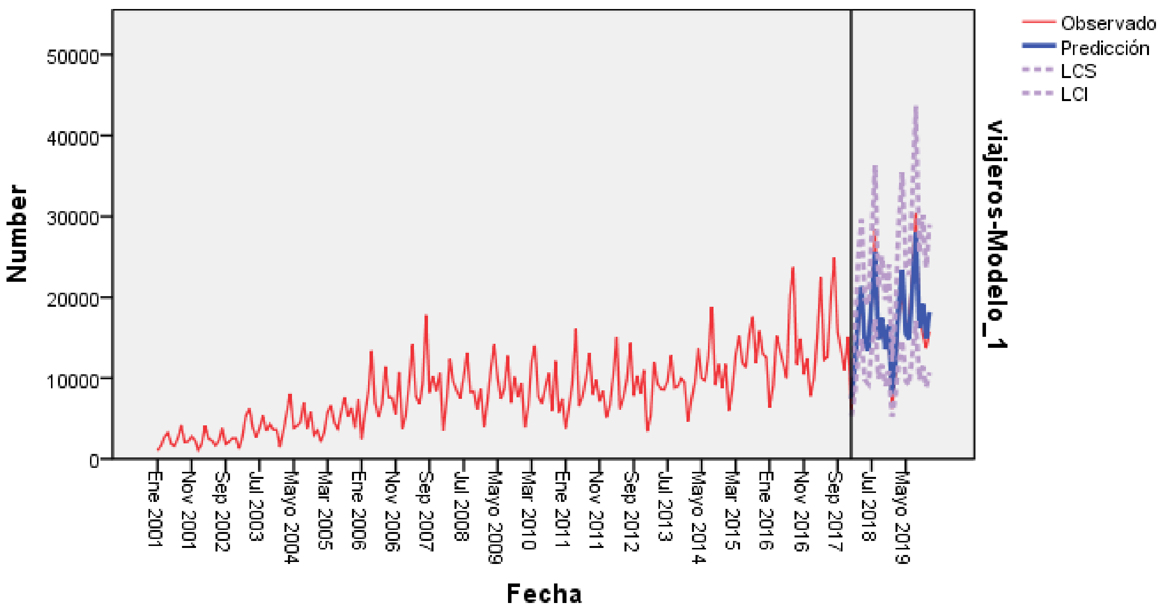
d)

Para estudiar que método aproxima mejor los datos calcularemos el ECM de la estimación de los datos con respecto a los datos reales, en el caso del modelo ARIMA que hemos creado el SPSS no da un ECM de 2854536 para el modelo ARIMA.

Para calcular el ECM he aplicado un filtro a los datos, luego he calculado la variable diff = (1/24)*(predicción - datos_reales)**2 y luego he sumado todas sus componentes.

Por otra parte, tras aplicar el método de suavizado exponencial a los datos y predecir los valores se obtiene un ECM de 37644450, mayor que el ECM del ARIMA, por tanto, concluimos que el método ARIMA arroja mejores predicciones para los resultados, se muestran las predicciones de ambos modelos:





El código de RStudio:

```
library(foreign)

dat<-read.spss("./datos/turismo_rural.sav",to.data.frame=T)

nac<-ts(dat$viajeros[1:205],freq=12)

plot(nac)

des<- decompose(nac,type="multiplicative")

des$seasonal

plot(des)

ses.cantidad<-HoltWinters(ts(dat$viajeros[1:204]),beta=FALSE,gamma=FALSE)

plot(ts(dat$viajeros),ylab="cantidad")

plot(ses.cantidad)

dat$viajeros[205]

p.ses.cantidad<-predict(ses.cantidad, n.ahead= 24)

p.ses.cantidad

dat$viajeros[205:228]

mean((dat$viajeros[205:228]-p.ses.cantidad)^2)

lines(p.ses.cantidad , col="green")
```