

METODOLOGÍA ARIMAX APLICADA A SERIES DE IMPUESTOS (SERIES ESTACIONALES)

Profesor: César Pérez López

Alumno: José Manuel Mendoza Gómez

ÍNDICE

<u>INTRODUCCIÓN</u>	Pág. 3
---------------------------	--------

ARIMA EN SPSS

- VARIABLE IRPF.....Pág. 7
- VARIABLE SOCIEDADES.....Pág. 10
- VARIABLE IVA.....Pág. 13
- VARIABLE ESPECIALES.....Pág. 16

ARIMA EN R

- VARIABLE IRPF.....Pág. 24
- VARIABLE SOCIEDADES.....Pág. 27
- VARIABLE IVA.....Pág. 29
- VARIABLE ESPECIALES.....Pág. 34

• INTRODUCCIÓN

En la anterior práctica vimos como elegir el mejor modelo de predicción teniendo en cuenta que las series eran estacionales y también que se realizaba un análisis de la intervención a través de la predicción automática.

En esta ocasión aplicaremos al modelo de predicción automática, la función de transferencia (siempre y cuando sea necesario). Esta metodología devuelve el mejor modelo ARIMAX según el valor AIC, AICc o BIC.

La predicción automática con función de transferencia realiza una búsqueda sobre el mejor modelo posible dentro de las restricciones de orden proporcionadas.

• BASE DE DATOS

	Nombre	Tipo	Anchura	Decimales	Etiqueta	Valores	Perdidos	Columnas
1	PERIODO	Cadena	4	0		Ninguno	Ninguno	4
2	V2	Cadena	10	0		Ninguno	Ninguno	10
3	INGRESOS...	Numérico	12	0		Ninguno	Ninguno	16
4	IRPF	Numérico	12	0		Ninguno	Ninguno	15
5	SOCIEDAD...	Numérico	12	0		Ninguno	Ninguno	19
6	IVA	Numérico	12	0		Ninguno	Ninguno	30
7	ESPECIALES	Numérico	12	0		Ninguno	Ninguno	35
8	T	Numérico	8	0		Ninguno	Ninguno	10
9	YEAR_	Numérico re...	4	0	YEAR, not peri...	Ninguno	Ninguno	7
10	MONTH_	Numérico	8	0	MONTH, period...	Ninguno	Ninguno	10
11	DATE_	Cadena	8	0	Date. Format: ...	Ninguno	Ninguno	10
12	T2	Numérico	8	2		Ninguno	Ninguno	10
13	T3	Numérico	8	2		Ninguno	Ninguno	10

La BBDD está formada por 13 variables.

- Variable periodo: Indica el año
- Variable V2: Indica el mes
- Ingresos tributarios totales
- IRPF: Impuesto sobre la renta de personas físicas
- Sociedades: Impuesto de sociedades
- IVA: Impuesto sobre el Valor Añadido

Vamos a trabajar con las variables IRPF, SOCIEDADES, IVA y ESPECIALES.

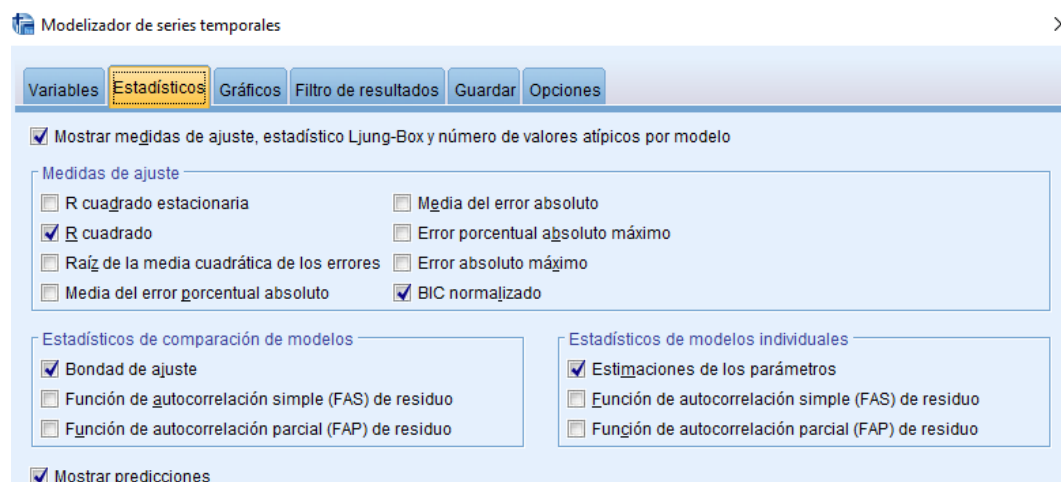
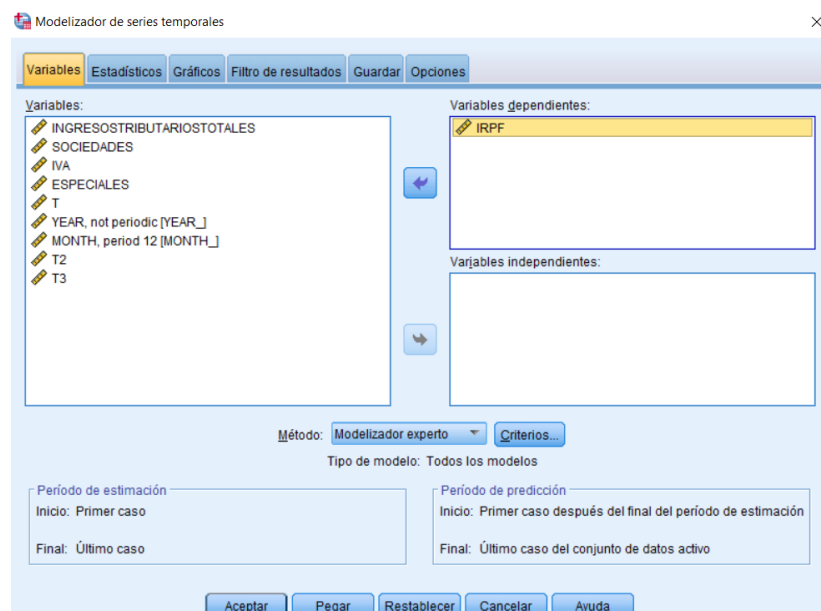
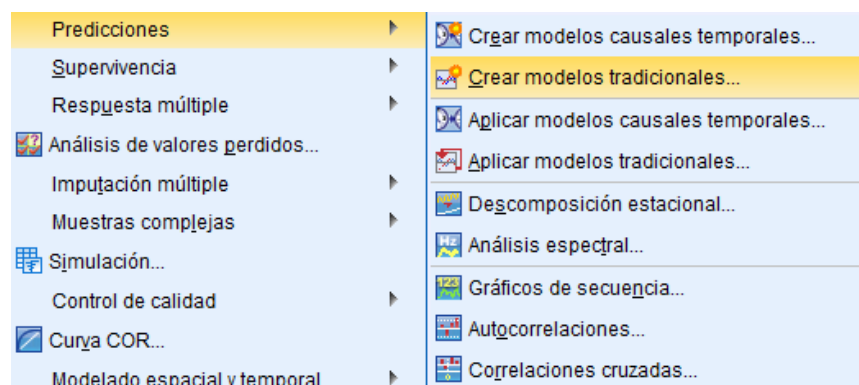
Este proyecto servirá para predecir futuros valores a partir de *la predicción automática con transferencia (incluyendo intervenciones)*.

La BBDD está formada por 224 registros. Cada observación representa un mes, por lo tanto, existe estacionalidad para todas las series que vamos a usar ya que la estacionalidad solo puede darse si el periodo es inferior al año.

Una vez que sabemos que las variables son estacionales, procedemos a ver que modelo nos muestra la PREDICCIÓN AUTOMÁTICA CON FUNCIÓN DE TRANSFERENCIA.

METODOLOGÍA ARIMAX EN SPSS

CÓDIGO



Modelizador de series temporales

Variables Estadísticos **Gráficos** Filtro de resultados Guardar Opciones

Gráficos para comparar modelos

☐ R cuadrado estacionaria
 ☐ Error porcentual absoluto máximo
☐ R cuadrado
 ☐ Error absoluto máximo
☐ Raíz de la media cuadrática de los errores
 ☐ BIC normalizado
☐ Media del error porcentual absoluto
 ☐ Función de autocorrelación simple (FAS) de residuo
☐ Media del error absoluto
 ☐ Función de autocorrelación parcial (FAP) de residuo

Gráficos de modelos individuales

☒ Serie
 ☒ Función de autocorrelación simple (FAS) de residuo
☒ Función de autocorrelación parcial (FAP) de residuo

Cada gráfico muestra

☒ Valores observados
☒ Predicciones
☒ Valores ajustados
☒ Intervalos de confianza de las predicciones
☒ Intervalos de confianza de los valores ajustados

Modelizador de series temporales

Variables Estadísticos Gráficos Filtro de resultados Guardar **Opciones**

Período de predicción

☐ Primer caso después del final del período de estimación hasta el último caso del conjunto de datos activo
☒ Primer caso después del final del período de estimación hasta una fecha especificada

Fecha:

Año	Mes
2020	12

Valores perdidos del usuario

☒ Tratar como no válidos
☐ Tratar como válidos

Ancho del intervalo de confianza (%): 95

Prefijo de los identificadores de modelo de los resultados: Modelo

Número máximo de retardos que se muestran en resultados de las FAS y FAP: 96

Modelizador de series temporales: Criterios de modelizador experto

Modelo Valores atípicos

Tipo de modelo

☐ Todos los modelos
☐ Sólo modelos de alisado exponencial
☒ Sólo modelos ARIMA
☒ El modelizador experto considera modelos estacionales
 Periodicidad actual: Ninguna

Eventos

Variables independientes:

Evento	Tipo	Variable
--------	------	----------

Las variables de evento son variables independientes especiales que se utilizan para modelar efectos de acontecimientos externos como inundaciones, huelgas o la presentación de una nueva línea de productos.

Active todas las variables que quiera tratar como variables de evento. Deberá codificarlas de manera que un 1 indique un punto temporal en el que se piensa que un evento ha producido un efecto.

Continuar Cancelar Ayuda

Modelizador de series temporales: Criterios de modelizador experto

Modelo Valores atípicos

☒ Detectar automáticamente los valores atípicos

Tipos de valores atípicos que se detectarán

☒ Aditivo
☒ Cambio de nivel
☒ Innovador
☒ Transitorio
☐ Aditivo estacional
☐ Tendencia local
☐ Parche aditivo

Continuar Cancelar Ayuda

VARIABLE IRPF

Modelizador de series temporales

Descripción del modelo

			Tipo de modelo
ID de modelo	IRPF	Modelo_1	ARIMA(2,1,3) (1,1,1)

SPSS ha elegido como mejor modelo para la variable IRPF el ARIMA(2,1,3)(1,1,1).

Antes de todo veremos si los estadísticos son significativos y posteriormente lo compararemos con el modelo que estimamos en la anterior práctica.

Estadísticos del modelo

Modelo	Número de predictores	Estadísticos de ajuste del modelo		Ljung-Box Q(18)			Número de valores atípicos
		R cuadrado	BIC normalizado	Estadísticos	DF	Sig.	
IRPF-Modelo_1	0	,987	26,335	15,318	13	,288	11

Parámetros del modelo ARIMA

					Estimación	SE	t	Sig.
IRPF-Modelo_1	IRPF	Ninguna transformación	AR	Retardo 1	-,882	,031	-28,388	,000
				Retardo 2	-,754	,047	-16,112	,000
			Diferencia		1			
			MA	Retardo 3	,550	,061	9,005	,000
			AR, estacional	Retardo 1	-,274	,272	-1,006	,316
			Diferencia estacional		1			
			MA, estacional	Retardo 1	-,397	,252	-1,578	,116

Este modelo tiene un r-cuadrado muy bueno y nos indica que el modelo logra explicar el 98,7% de la variabilidad total.

El estadístico de Ljung-Box, que nos indica la aleatoriedad residual, también es muy bueno ya que supera el nivel de significación alfa igual a 0.05

El valor de la capacidad predictiva del BIC normalizado es de 26,335. Este valor no tiene sentido por si solo y sirve para compararlo con otros modelos. Cuanto más bajo sea, mayor capacidad predictiva tiene el modelo

En cuanto a la significatividad individual, hay problema ya que no todos los parámetros son significativos. La parte AR estacional y la parte MA estacional del primer retardo no son significativos. Luego, hay problemas de significatividad individual.

Valores atípicos

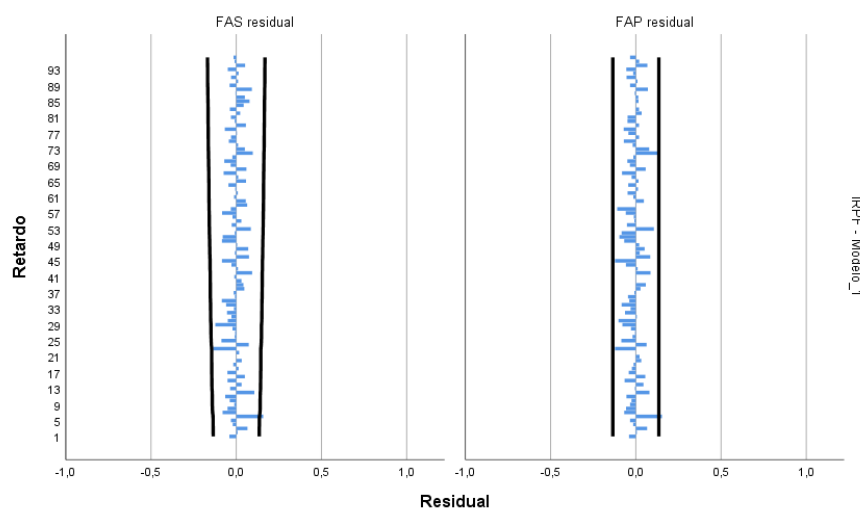
			Estimación	SE	t	Sig.
IRPF-Modelo_1	Ene 2003	Aditivo	1410076,855	273422,928	5,157	,000
	Jul 2007	Innovador	2121969,511	429136,937	4,945	,000
	Jun 2008	Aditivo	1555957,855	278893,528	5,579	,000
	Jul 2008	Innovador	-5966314,297	439169,789	-13,585	,000
	Jul 2009	Innovador	2903597,423	510748,881	5,685	,000
	Abr 2014	Aditivo	-1522103,849	267115,470	-5,698	,000
	Mayo 2014	Aditivo	1893372,489	269083,025	7,036	,000
	Ago 2014	Innovador	4023961,127	431230,866	9,331	,000
	Sep 2014	Innovador	-4701796,142	448228,671	-10,490	,000
	Abr 2019	Innovador	-2627837,478	427680,956	-6,144	,000
	Mayo 2019	Innovador	2737413,086	433573,859	6,314	,000

Predicción

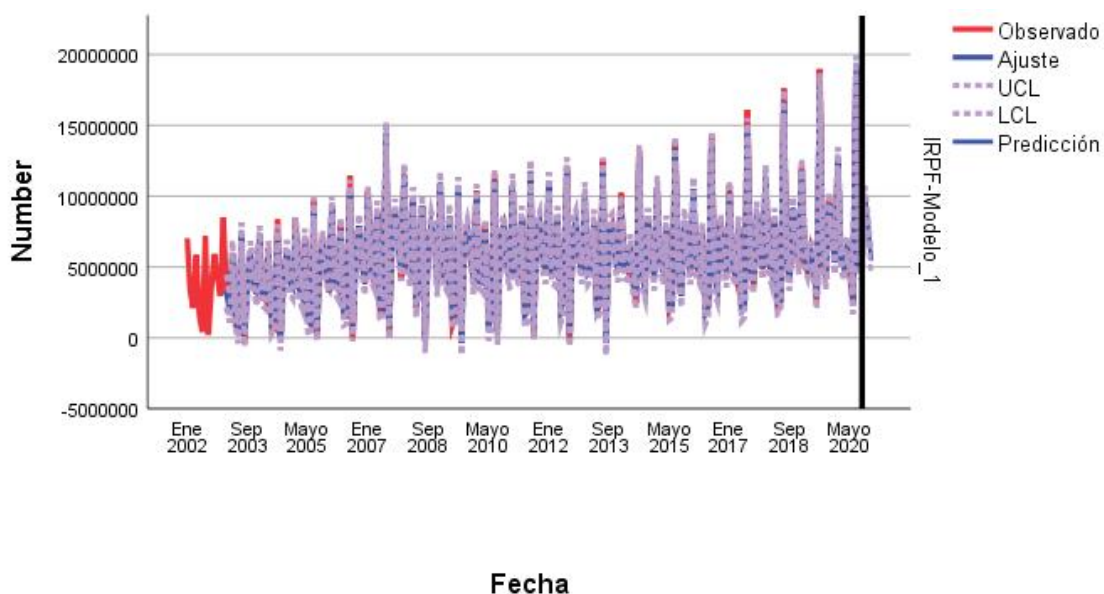
Modelo		Sep 2020	Oct 2020	Nov 2020	Dic 2020
IRPF-Modelo_1	Predicción	4195552	9854220	7997550	5466720
	UCL	5035242	10699724	8851398	6343345
	LCL	3355862	9008716	7143703	4590095

Para cada modelo, las predicciones empiezan después del último valor no perdido en el rango del período de estimación solicitado, y finaliza en el último período para el cual los valores perdidos de todos los predictores están disponibles o al final de la fecha del período de predicción solicitado, lo que suceda antes.

SPSS ha logrado detectar outliers aditivos e innovadores para la variable IRPF en los distintos momentos del tiempo.



Viendo el gráfico de autocorrelación y autocorrelación parcial de los residuos vemos que ningún retardo se sale de las bandas de confianza, luego la aleatoriedad residual es buena.

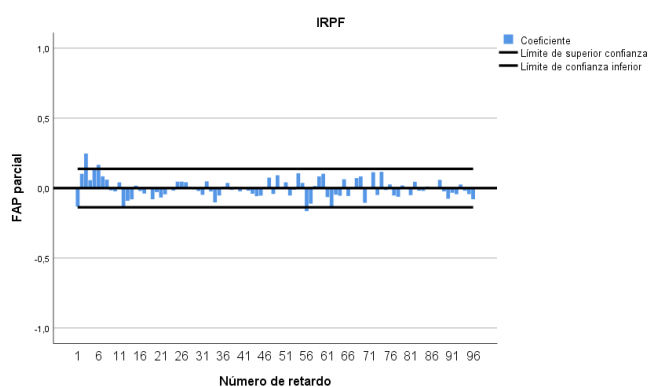
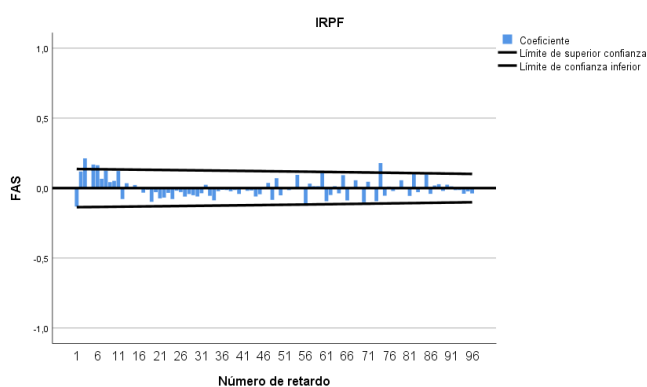


COMPARACIÓN CON NUESTRO MODELO

En la anterior práctica, no logramos encontrar ningún modelo valido ya que todos nos fallaban por la aleatoriedad residual. Es decir, nos mostraban un p-valor de ljung box muy bajo.

Por lo tanto, si tuviésemos que quedarnos con algún modelo nos quedaríamos con el modelo sugerido por la predicción automática. El $ARIMA(2,1,3)(1,1,1)$. *PESE A QUE LA PREDICCIÓN AUTOMÁTICA HA APLICADO UNA DIFERENCIA ESTACIONARIA CUANDO NO DEBERÍA.*

Modelo con $d=0$ y $D=1$ (sin aplicar diferencia estacionaria)



VARIABLE SOCIEDADES

Descripción del modelo

			Tipo de modelo
ID de modelo	SOCIEDADES	Modelo_1	ARIMA(0,0,0) (0,1,0)

SPSS ha elegido como mejor modelo para la variable SOCIEDADES el ARIMA(0,0,0)(0,1,0).

Antes de todo veremos si los estadísticos son significativos y posteriormente lo compararemos con el modelo que estimamos en la anterior práctica.

Estadísticos del modelo

Modelo	Número de predictores	Estadísticos de ajuste del modelo		Ljung-Box Q(18)			Número de valores atípicos
		R cuadrado	BIC normalizado	Estadísticos	DF	Sig.	
SOCIEDADES-Modelo_1	0	,958	27,265	19,245	18	,377	11

Parámetros del modelo ARIMA

				Estimación
SOCIEDADES-Modelo_1	SOCIEDADES	Ninguna transformación	Diferencia estacional	1

Este modelo tiene un r-cuadrado muy bueno y nos indica que el modelo logra explicar el 95,8% de la variabilidad total.

El estadístico de Ljung-Box, que nos indica la aleatoriedad residual, también es muy bueno ya que supera el nivel de significación alfa igual a 0.05

El valor de la capacidad predictiva del BIC normalizado es de 27,265. Este valor no tiene sentido por si solo y sirve para compararlo con otros modelos. Cuanto más bajo sea, mayor capacidad predictiva tiene el modelo

En cuanto a la significatividad individual vemos que, para el ARIMA(0,0,0)(0,1,0), no se puede contrastar ninguna significatividad individual ya que no tenemos parte AR ni parte MA.

Valores atípicos

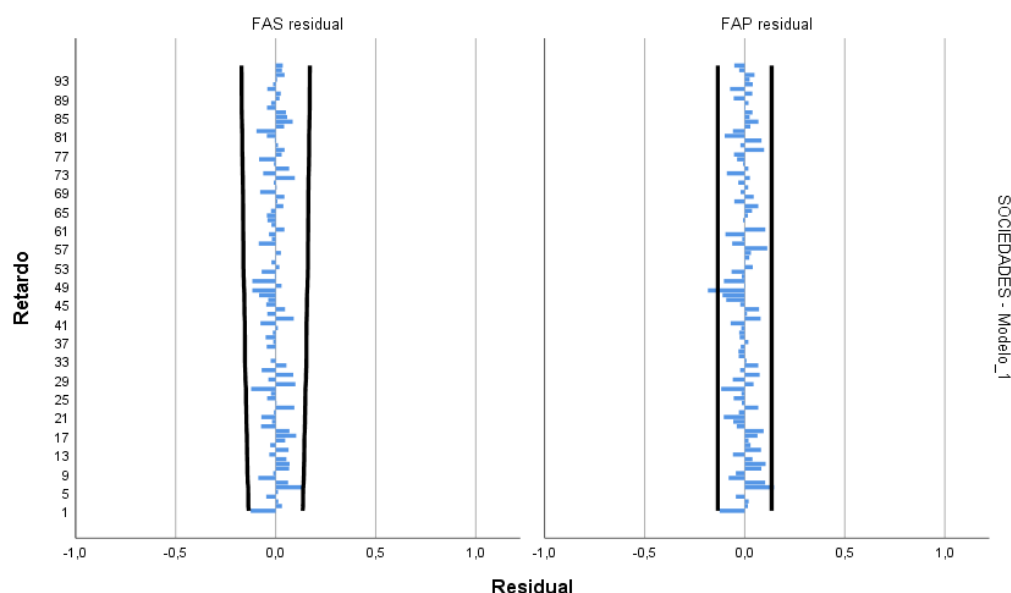
			Estimación	SE	t	Sig.
SOCIEDADES-Modelo_1	Ago 2005	Innovador	2728498,733	724866,952	3,764	,000
	Ago 2007	Aditivo	6031911,909	512558,337	11,768	,000
	Abr 2008	Innovador	-5086471,501	724866,952	-7,017	,000
	Mayo 2008	Aditivo	3245167,682	512558,337	6,331	,000
	Oct 2008	Innovador	-5286988,482	724866,952	-7,294	,000
	Ago 2009	Innovador	-4013657,607	724866,952	-5,537	,000
	Ago 2010	Innovador	-2578247,747	724866,952	-3,557	,000
	Abr 2016	Aditivo	-2872232,718	512558,337	-5,604	,000
	Oct 2016	Innovador	4350318,574	724866,952	6,002	,000
	Oct 2018	Aditivo	2473327,758	512558,337	4,825	,000
	Abr 2020	Innovador	-3573462,650	724866,952	-4,930	,000

Predicción

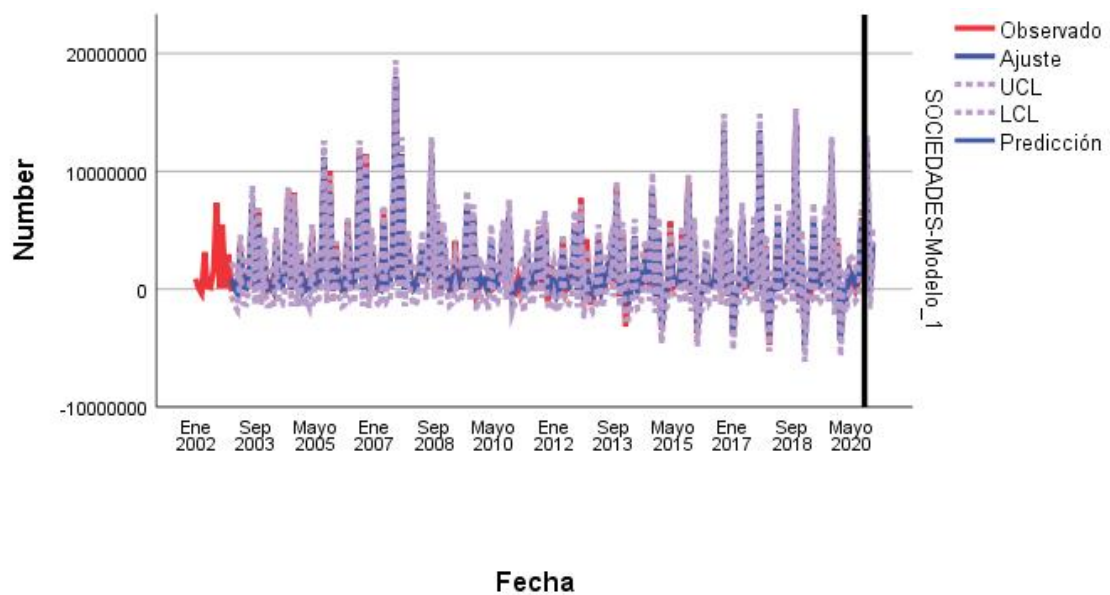
Modelo		Sep 2020	Oct 2020	Nov 2020	Dic 2020
SOCIEDADES-Modelo_1	Predicción	201996	11613426	-223624	3903401
	UCL	1631313	13042743	1205693	5332718
	LCL	-1227321	10184109	-1652941	2474084

Para cada modelo, las predicciones empiezan después del último valor no perdido en el rango del período de estimación solicitado, y finaliza en el último período para el cual los valores perdidos de todos los predictores están disponibles o al final de la fecha del período de predicción solicitado, lo que suceda antes.

SPSS ha logrado detectar outliers innovadores y aditivos en los distintos periodos del tiempo.



Viendo el gráfico de autocorrelación y autocorrelación parcial de los residuos vemos que ningún retardo se sale de las bandas de confianza, luego la aleatoriedad residual es buena.



COMPARACIÓN CON NUESTRO MODELO

ARIMA(0,0,0)(1,1,0)

En la anterior práctica, de todos los modelos, nos habíamos quedado con el ARIMA (1,1,0) ya que era el que mostraba los mejores valores para los distintos contrastes.

Estadísticos del modelo

Modelo	Número de predictores	Estadísticos de ajuste del modelo		Ljung-Box Q(18)			Número de valores atípicos
		R cuadrado	BIC normalizado	Estadísticos	DF	Sig.	
SOCIEDADES-Modelo_1	0	,866	28,130	30,395	17	,024	0

Parámetros del modelo ARIMA

					Estimación	SE	t	Sig.
SOCIEDADES-Modelo_1	SOCIEDADES	Ninguna transformación	AR, estacional	Retardo 1	-,120	,071	-1,694	,092
			Diferencia estacional		1			

Sin embargo, nos fallaba también la aleatoriedad residual.

Ahora conocemos el porqué, y fue porque sobrestimamos el retardo de la parte MA estacional.

Por lo tanto, si tuviéramos que elegir entre alguno de los dos, nos quedaríamos con el modelo sugerido por la predicción automática. El ARIMA(0,0,0)(0,1,0).

VARIABLE IVA

Descripción del modelo

			Tipo de modelo
ID de modelo	IVA	Modelo_1	ARIMA(0,0,0) (0,1,0)

SPSS ha elegido como mejor modelo para la variable IVA el ARIMA(0,0,0)(0,1,0).

Antes de todo veremos si los estadísticos son significativos y posteriormente lo compararemos con el modelo que estimamos en la anterior práctica.

Estadísticos del modelo

Modelo	Número de predictores	Estadísticos de ajuste del modelo		Ljung-Box Q(18)			Número de valores atípicos
		R cuadrado	BIC normalizado	Estadísticos	DF	Sig.	
IVA-Modelo_1	0	,976	27,233	19,939	18	,336	17

Parámetros del modelo ARIMA

				Estimación	SE	t	Sig.
IVA-Modelo_1	IVA	Ninguna transformación	Constante	186598,154	45234,606	4,125	,000
			Diferencia estacional	1			

Este modelo tiene un r-cuadrado muy bueno y nos indica que el modelo logra explicar el 97,6% de la variabilidad total.

El estadístico de Ljung-Box, que nos indica la aleatoriedad residual, también es muy bueno ya que supera el nivel de significación alfa igual a 0.05

El valor de la capacidad predictiva del BIC normalizado es de 27,233. Este valor no tiene sentido por si solo y sirve para compararlo con otros modelos. Cuanto más bajo sea, mayor capacidad predictiva tiene el modelo

En cuanto a la significatividad individual no hay problema ya que todos los parámetros son significativos.

Valores atípicos

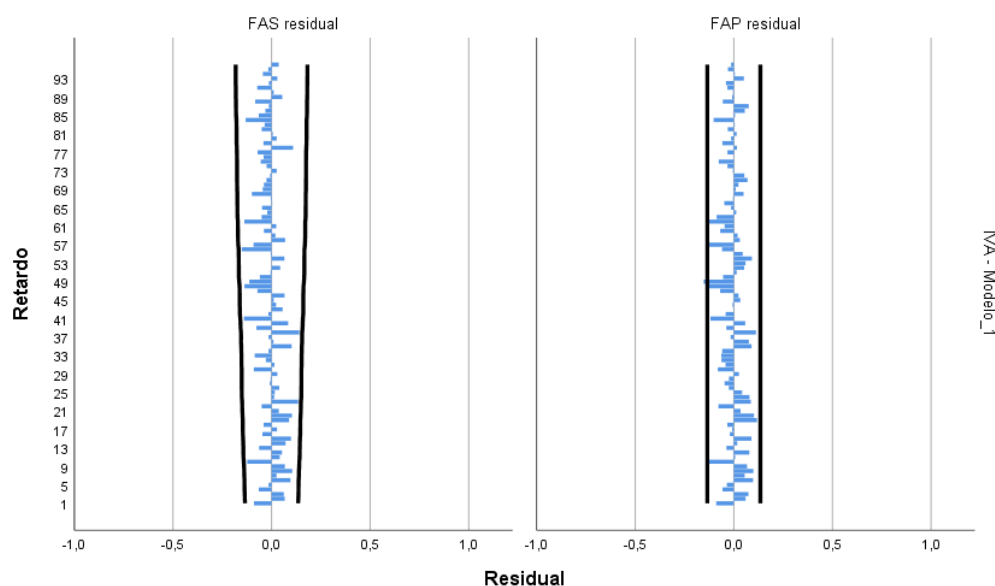
			Estimación	SE	t	Sig.
IVA-Modelo_1	Feb 2007	Aditivo	1768589,000	450078,414	3,930	,000
	Jul 2007	Aditivo	-1960591,882	450078,938	-4,356	,000
	Mayo 2008	Aditivo	-2922259,509	460455,235	-6,346	,000
	Ago 2008	Innovador	-2928104,154	638113,809	-4,589	,000
	Feb 2009	Innovador	-2620107,154	638113,809	-4,106	,000
	Mar 2009	Transitorio	-2816724,107	298740,143	-9,429	,000
		Magnitud				
	Jul 2010	Transitorio	1867107,609	443081,911	4,214	,000
		Magnitud				
		Factor de decrecimiento	,855	,026	33,058	,000
	Mayo 2014	Aditivo	1803027,166	450078,732	4,006	,000
	Ago 2014	Innovador	3579451,753	638114,430	5,609	,000
	Sep 2014	Innovador	-3744016,401	638114,282	-5,867	,000
	Abr 2015	Innovador	2693015,734	638113,865	4,220	,000
	Ago 2017	Aditivo	-3673398,892	450079,462	-8,162	,000
	Ene 2018	Innovador	3310045,883	638113,809	5,187	,000
	Feb 2018	Innovador	-3159092,977	638113,794	-4,951	,000
	Abr 2019	Innovador	-2924579,246	638113,794	-4,583	,000
	Mayo 2019	Innovador	2546266,767	638113,794	3,990	,000
	Abr 2020	Cambio de nivel	-1841021,565	288226,302	-6,387	,000

Predicción

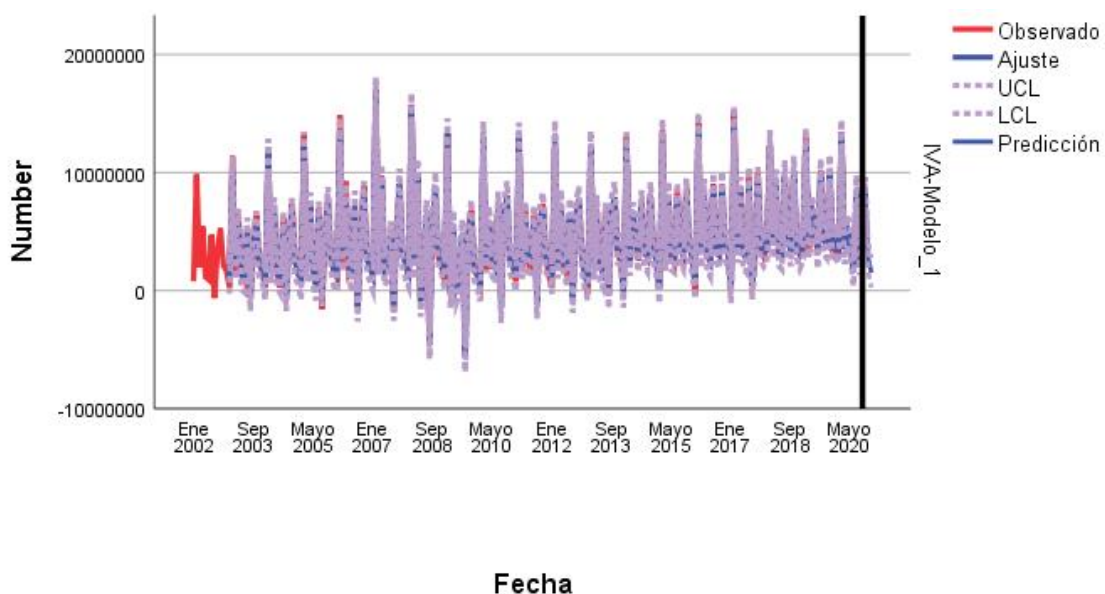
Modelo		Sep 2020	Oct 2020	Nov 2020	Dic 2020
IVA-Modelo_1	Predicción	2515680	8337844	2658772	1522162
	UCL	3771125	9593289	3914217	2777607
	LCL	1260235	7082399	1403327	266717

Para cada modelo, las predicciones empiezan después del último valor no perdido en el rango del período de estimación solicitado, y finaliza en el último período para el cual los valores perdidos de todos los predictores están disponibles o al final de la fecha del período de predicción solicitado, lo que suceda antes.

SPSS ha logrado detectar outliers innovadores, aditivos, transitorios en los distintos periodos del tiempo y un cambio de nivel en abril de 2020, coincidente cuando se tomaron medidas frente al COVID.



Viendo el gráfico de autocorrelación y autocorrelación parcial de los residuos vemos que ningún retardo se sale de las bandas de confianza, luego la aleatoriedad residual es buena.



COMPARACIÓN CON NUESTRO MODELO

ARIMA(0,0,0)(0,1,1)

Estadísticos del modelo

Modelo	Número de predictores	Estadísticos de ajuste del modelo			Ljung-Box Q(18)			Número de valores atípicos
		R cuadrado estacionaria	R cuadrado	BIC normalizado	Estadísticos	DF	Sig.	
IVA-Modelo_1	0	,042	,912	27,987	41,101	17	,001	0

Parámetros del modelo ARIMA

				Estimación	SE	t	Sig.
IVA-Modelo_1	IVA	Ninguna transformación	Constante	134035,871	61768,965	2,170	,031
			Diferencia estacional	1			
			MA, estacional Retardo 1	,243	,070	3,444	,001

Nos ha pasado lo mismo que en el anterior caso, **hemos sobrestimado el retardo de la parte MA estacional. Por ese motivo fue que nos fallaba la aleatoriedad residual.**

Por lo tanto, si tuviéramos que elegir entre alguno de los dos, nos quedaríamos con el modelo sugerido por la predicción automática. El ARIMA(0,0,0)(0,1,0).

VARIABLE ESPECIALES

Descripción del modelo

			Tipo de modelo
ID de modelo	ESPECIALES	Modelo_1	ARIMA(1,0,1) (0,1,1)

SPSS ha elegido como mejor modelo para la variable ESPECIALES el ARIMA(1,0,1)(0,1,1).

Antes de todo veremos si los estadísticos son significativos y posteriormente lo compararemos con el modelo que estimamos en la anterior práctica.

Estadísticos del modelo

Modelo	Número de predictores	Estadísticos de ajuste del modelo		Ljung-Box Q(18)			Número de valores atípicos
		R cuadrado	BIC normalizado	Estadísticos	DF	Sig.	
ESPECIALES-Modelo_1	0	,606	23,304	30,581	15	,010	5

Parámetros del modelo ARIMA

					Estimación	SE	t	Sig.
ESPECIALES-Modelo_1	ESPECIALES	Ninguna transformación	AR	Retardo 1	,982	,017	58,774	,000
			MA	Retardo 1	,863	,042	20,349	,000
			Diferencia estacional		1			
			MA, estacional	Retardo 1	,629	,053	11,893	,000

Este modelo tiene un r-cuadrado muy bueno y nos indica que el modelo logra explicar el 60,6% de la variabilidad total.

El estadístico de Ljung-Box, que nos indica la aleatoriedad residual, ES MALO ya que no logra superar el nivel de significación alfa igual a 0.05

El valor de la capacidad predictiva del BIC normalizado es de 23,304. Este valor no tiene sentido por si solo y sirve para compararlo con otros modelos. Cuanto más bajo sea, mayor capacidad predictiva tiene el modelo

En cuanto a la significatividad individual no hay problema ya que todos los parámetros son significativos.

Por lo tanto, el modelo de predicción automática falla en la aleatoriedad residual.

Valores atípicos

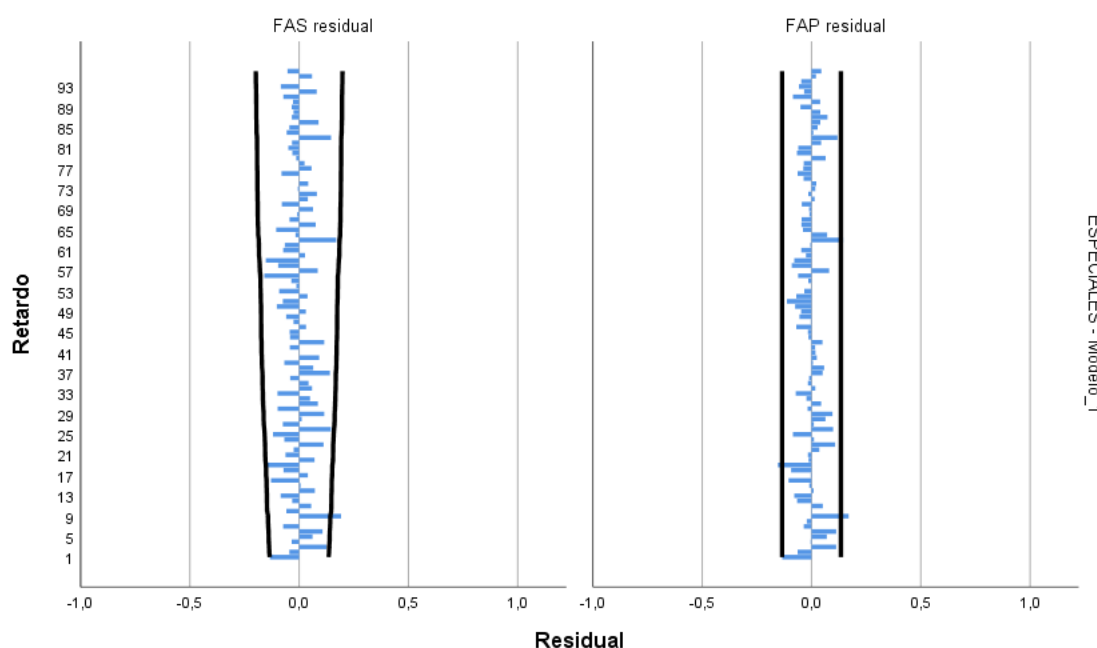
			Estimación	SE	t	Sig.
ESPECIALES-Modelo_1	Jul 2003	Innovador	735975,254	115723,969	6,360	,000
	Jul 2005	Innovador	397510,999	105909,679	3,753	,000
	Jun 2013	Aditivo	378549,436	88241,256	4,290	,000
	Dic 2014	Innovador	-374273,281	101017,567	-3,705	,000
	Mayo 2020	Transitorio	-803192,735	96274,879	-8,343	,000
		Magnitud				
		Factor de decrecimiento	,628	,077	8,105	,000

Predicción

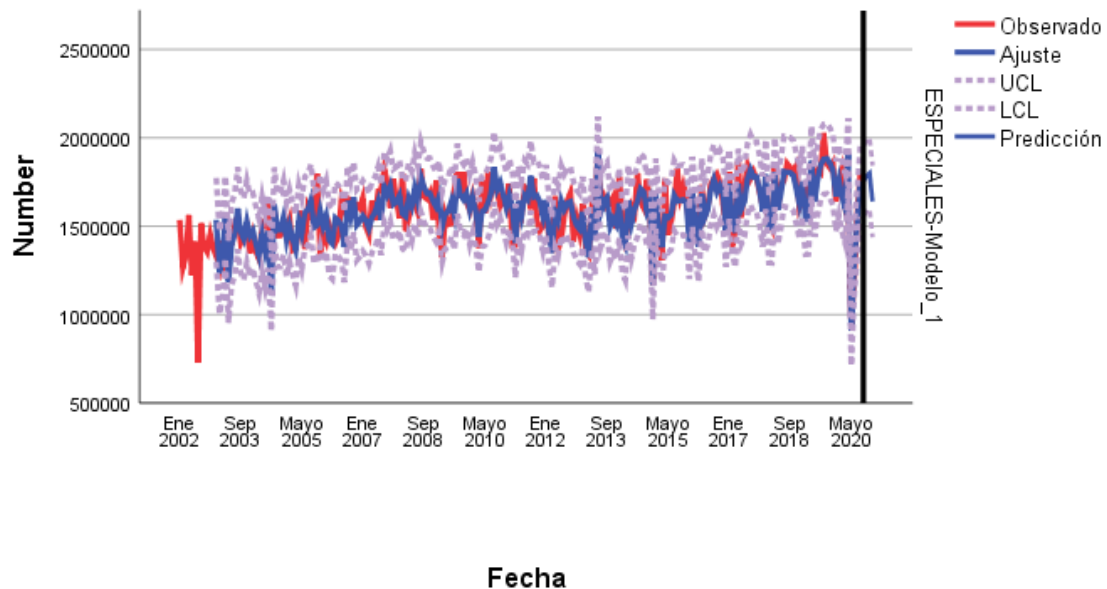
Modelo		Sep 2020	Oct 2020	Nov 2020	Dic 2020
ESPECIALES-Modelo_1	Predicción	1758013	1782284	1798954	1639756
	UCL	1956467	1982150	2000175	1842275
	LCL	1559558	1582417	1597734	1437237

Para cada modelo, las predicciones empiezan después del último valor no perdido en el rango del período de estimación solicitado, y finaliza en el último período para el cual los valores perdidos de todos los predictores están disponibles o al final de la fecha del período de predicción solicitado, lo que suceda antes.

SPSS ha logrado detectar outliers innovadores y aditivos en los distintos periodos del tiempo y un cambio transitorio en mayo de 2020, coincidente cuando se tomaron medidas frente al COVID.



Viendo el gráfico de autocorrelación y autocorrelación parcial de los residuos vemos que hay retardos que se sale de las bandas de confianza, luego la aleatoriedad residual no es del todo buena.



COMPARACIÓN CON NUESTRO MODELO

ARIMA(2,1,0)(0,1,1)

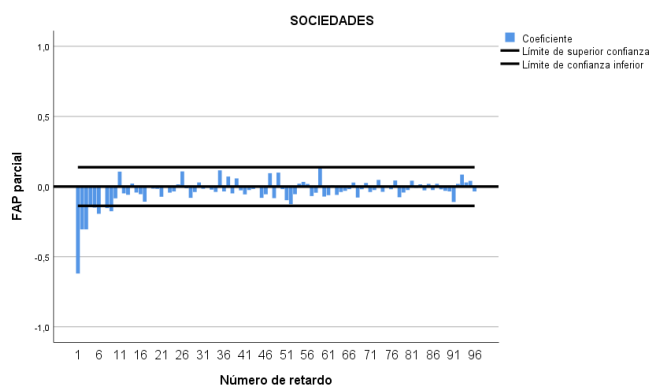
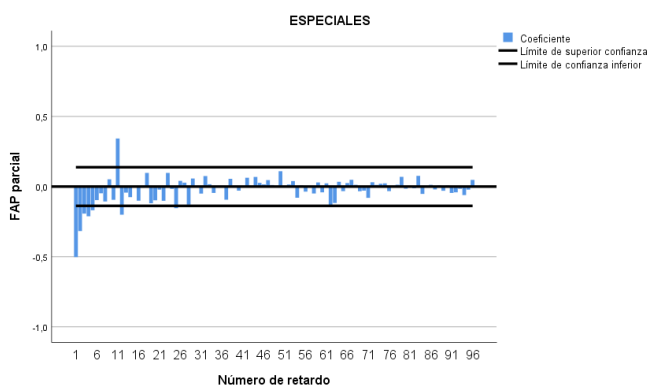
Nosotros decidimos aplicar una diferencia regular porque tras aplicar la diferencia estacionaria nos aparecía que los retardos en la FAP parcial decrecían lentamente.

Es decir, decrecían tan lentamente que había 5 términos significativos consecutivos.

$d=0$ y $D=1$

→

$d=1$ y $D=1$



Estadísticos del modelo

Modelo	Número de predictores	Estadísticos de ajuste del modelo		Ljung-Box Q(18)			Número de valores atípicos
		R cuadrado	BIC normalizado	Estadísticos	DF	Sig.	
ESPECIALES-Modelo_1	0	,140	23,896	24,945	15	,051	0

Parámetros del modelo ARIMA

					Estimación	SE	t	Sig.
ESPECIALES-Modelo_1	ESPECIALES	Ninguna transformación	AR	Retardo 1	-,622	,066	-9,358	,000
				Retardo 2	-,330	,066	-4,984	,000
			Diferencia		1			
			Diferencia estacional		1			
			MA, estacional	Retardo 1	,799	,069	11,579	,000

Hemos conseguido superar la aleatoriedad residual del modelo de predicción automática y además hemos obtenemos parecidas significatividades individuales.

Sin embargo, nuestro modelo presenta un R-cuadrado de 0,140 en contraste al modelo de predicción automática que llega a un R-cuadrado de 0,606.

Por lo tanto, si tuviéramos que elegir entre alguno de los dos, nos quedaríamos con el modelo sugerido por la predicción automática. El ARIMA(1,0,1)(0,1,1). *PESE A QUE LOS RESIDUOS TAMPOCO SON DEL TODO ALEATORIOS Y TAMBIÉN QUE LA PREDICCIÓN AUTOMÁTICA DEBERÍA HABER APLICADO UNA DIFERENCIA REGULAR TRAS LA DIFERENCIA ESTACIONARIA.*

Función de transferencia

Vamos a realizar el modelo con función de transferencia a ver si se puede aumentar la aleatoriedad residual ya que un valor de 0,01 del estadístico de Ljung-Box es demasiado bajo.

Antes de nada, se analizan las correlaciones existentes entre las variables. Lo vemos con la matriz de correlaciones. A p-valores bajos, mayor dependencia existente.

Correlacion

Correlaciones bivariadas

Variables:

- ESPECIALES
- INGRESOSTRIBUT...
- IRPF
- SOCIEDADES
- IVA

Coeficientes de correlación:

☒ Pearson ☐ Tau-b de Kendall ☐ Spearman

Prueba de significación:

☒ Bilateral ☐ Unilateral

☒ Señalar las correlaciones significativas

Aceptar Pegar Restablecer Cancelar Ayuda

ESPECIALES

Variable	Correlación de Pearson	Sig. (bilateral)	N
ESPECIALES	1		
INGRESOSTRIBUTARIO STOTALES	,291**	,000	224
IRPF	,260**	,000	224
SOCIEDADES	,221**	,001	224
IVA	,042	,535	224

** La correlación es significativa en el nivel 0,01 (bilateral).

Las variables que están más correladas con ESPECIALES son INGRESOS TOTALES, IRPF y SOCIEDADES. Vamos a realizar un análisis factorial para reducir su dimensión y además, que se hagan incorreladas.

Análisis factorial

Variables:

- INGRESOSTRIBUT...
- IRPF
- SOCIEDADES

Variable de selección:

Valor...

Aceptar Pegar Restablecer Cancelar Ayuda

Análisis factorial: Descriptivos

Estadísticos:

☐ Descriptivos univariados

☒ Solución inicial

Matriz de correlaciones:

☐ Coeficientes ☐ Inverso

☐ Niveles de significación ☐ Reproducible

☒ Determinante ☐ Anti-imagen

☒ KMO y prueba de esfericidad de Bartlett

Continuar Cancelar Ayuda

Análisis factorial: Rotación

Método:

☐ Ninguno ☐ Quartimax

☒ Varimax ☐ Equamax

☐ Oblimin directo ☐ Promax

Delta: 0 Kappa: 4

Mostrar:

☒ Solución rotada ☐ Gráficos de cargas

N.º máximo de iteraciones para convergencia: 25

Continuar Cancelar Ayuda

Análisis factorial: Puntuaciones factoriales

☒ Guardar como variables

Método

☒ Regresión

☐ Bartlett

☐ Anderson-Rubin

☐ Mostrar matriz de coeficientes de las puntuaciones factoriales

Continuar Cancelar Ayuda

Prueba de KMO y Bartlett		
Medida Kaiser-Meyer-Olkin de adecuación de muestreo		,313
Prueba de esfericidad de Bartlett	Aprox. Chi-cuadrado	306,944
	gl	3
	Sig.	,000

Matriz de componente^a

	Componente	
	1	2
INGRESOSTRIBUTARIO STOTALES	,968	,022
IRPF	,783	-,585
SOCIEDADES	,520	,841

Método de extracción: análisis de componentes principales.

a. 2 componentes extraídos.

KMO MUY BAJO, Se desaconseja meter SOCIEDADES en el análisis factorial. Lo volvemos a hacer, pero sin SOCIEDADES.

Prueba de KMO y Bartlett		
Medida Kaiser-Meyer-Olkin de adecuación de muestreo		,500
Prueba de esfericidad de Bartlett	Aprox. Chi-cuadrado	144,273
	gl	1
	Sig.	,000

Matriz de componente^a

	Componente
	1
INGRESOSTRIBUTARIO STOTALES	,920
IRPF	,920

Método de extracción: análisis de componentes principales.

a. 1 componentes extraídos.

FAC1_1
-,03622
-,32605
-1,23826
,20661
-1,33742
-1,62469
,26508
-1,25937
-,72686
,36025

Vamos a predecir la variable regresora FAC1_1, que está compuesta por el Ingreso Tributario Total y el IRPF.

Modelizador de series temporales

Variables: INGRESOSTRIBUTARIOTOTALES, IRPF, SOCIEDADES, IVA, ESPECIALES, T, YEAR, not periodic [YEAR_], MONTH, period 12 [MONTH_], T2, T3

Variables dependientes: REGR factor score_1 for analysis 1 [FAC1_1]

Variables independientes:

Método: Modelizador experto Criterios...

Tipo de modelo: Sólo modelos ARIMA

Período de estimación: Inicio: Primer caso, Final: Último caso

Período de predicción: Inicio: Primer caso después del final del período de estimación, Final: Último caso del conjunto de datos activo

Guardar variables

Descripción	Guardar	Prefijo del nombre de variable
Valores pronosticados	<input checked="" type="checkbox"/>	Pronosticado
Límites de confianza inferiores	<input type="checkbox"/>	LCL
Límites de confianza superiores	<input type="checkbox"/>	UCL
Residuos de ruido	<input type="checkbox"/>	ResiduoN

Por cada elemento seleccionado, se guardará una variable para cada variable dependiente.

-,49112	Cortar
2,09878	Copiar
,47304	Copiar con nombres de variable
,06870	

Hacemos lo mismo para la variable SOCIEDADES y volvemos a plantear el MODELO.

Modelizador de series temporales

Variables: INGRESOSTRIBUTARIOTOTALES, IRPF, IVA, T, YEAR, not periodic [YEAR_], MONTH, period 12 [MONTH_], T2, T3, Valor pronosticado de FAC1_1-Modelo_1 [Pronosticado...], Valor pronosticado de SOCIEDADES-Modelo_1 [Pronosti...]

Variables dependientes: ESPECIALES

Variables independientes: SOCIEDADES, REGR factor score_1 for analysis 1 [FAC1_1]

Método: Modelizador experto Criterios...

Tipo de modelo: Sólo modelos ARIMA

Período de estimación: Inicio: Primer caso, Final: Último caso

Período de predicción: Inicio: Primer caso después del final del período de estimación, Final: Último caso del conjunto de datos activo

Aceptar Pegar Restablecer Cancelar Ayuda

Estadísticos del modelo							
Modelo	Número de predictores	Estadísticos de ajuste del modelo		Ljung-Box Q(18)			Número de valores atípicos
		R cuadrado	BIC normalizado	Estadísticos	DF	Sig.	
ESPECIALES-Modelo_1	2	,634	23,283	21,406	14	,092	4

Este modelo tiene un r-cuadrado muy bueno y nos indica que el modelo logra explicar el 63,6% de la variabilidad total. Hemos aumentado en un 3% la explicación de la varianza

El estadístico de Ljung-Box, que nos indica la aleatoriedad residual sigue siendo malo pero no tanto como el 0,01 del modelo anterior

El valor de la capacidad predictiva del BIC normalizado es de 23,283. Este valor no tiene sentido por si solo y sirve para compararlo con otros modelos. Cuanto más bajo sea, mayor capacidad predictiva tiene el modelo. Por lo tanto, hemos conseguido disminuirla con la función de transferencia.

En cuanto a la significatividad individual no hay problema ya que todos los parámetros son significativos.

Parámetros del modelo ARIMA								
					Estimación	SE	t	Sig.
ESPECIALES-Modelo_1	ESPECIALES	Ninguna transformación	AR	Retardo 1	,979	,018	54,027	,000
				Retardo 2	-,192	,066	-2,933	,004
			MA	Retardo 1	1,048	,066	15,769	,000
				Retardo 2	-,192	,066	-2,933	,004
			Diferencia estacional		1			
			MA, estacional	Retardo 1	,560	,061	9,202	,000
	SOCIEDADES	Ninguna transformación	Retardo		6			
			Numerador	Retardo 0	,011	,006	2,010	,046
			Diferencia estacional		1			
	REGR factor score 1 for analysis 1	Ninguna transformación	Numerador	Retardo 0	86152,250	23703,436	3,635	,000
				Retardo 6	61279,607	26750,679	2,291	,023
			Diferencia estacional		1			

					Estimación	SE	t	Sig.
ESPECIALES-Modelo_1	Jul 2003	Innovador			733878,980	111603,010	6,576	,000
	Jun 2013	Aditivo			383072,286	82657,706	4,634	,000
	Dic 2014	Aditivo			-347755,282	81636,340	-4,260	,000
	Mayo 2020	Transitorio	Magnitud		-833627,425	87482,114	-9,529	,000
				Factor de decrecimiento	,616	,066	9,312	,000

Conclusión:

Nos quedamos con el modelo propuesto por la función de transferencia ya que tiene mejor R-cuadrado, mejor capacidad predictiva y mayor aleatoriedad residual.

METODOLOGÍA ARIMAX EN R

En esta ocasión analizaremos la predicción automática con el software R para series estacionales. A diferencia de SPSS, **no hemos encontrado una función de predicción automática que tome en cuenta las intervenciones en la serie.**

VARIABLE IRPF

```
#IRPF
```

```
seriex=ts(IRPF,start=c(2002,1),frequency =12)
```

Transformamos la variable IRPF a formato series temporales antes de realizar la predicción automática. Ponemos de argumento frecuencia 12 porque la serie ya la hemos analizado anteriormente y concluimos que era estacional.

```
auto.arima(seriex)
coeftest(auto.arima(seriex))
checkresiduals(auto.arima(seriex))
```

Modelo propuesto

```
> auto.arima(seriex)
Series: seriex
ARIMA(2,0,2)(0,1,1)[12]
```

Coefficients:

	ar1	ar2	ma1	ma2	sma1
	1.2633	-0.3156	-1.4630	0.6174	-0.1202
s.e.	0.2244	0.2157	0.1926	0.1654	0.0725

```
sigma^2 estimated as 6.64e+11: log likelihood=-3184.23
AIC=6380.46 AICc=6380.87 BIC=6400.6
```

```
> coeftest(auto.arima(seriex))
```

z test of coefficients:

	Estimate	Std. Error	z value	Pr(> z)	
ar1	1.263276	0.224389	5.6298	1.804e-08	***
ar2	-0.315626	0.215724	-1.4631	0.1434397	
ma1	-1.462972	0.192638	-7.5944	3.092e-14	***
ma2	0.617413	0.165378	3.7334	0.0001889	***
sma1	-0.120168	0.072463	-1.6583	0.0972497	.

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

El modelo que nos aconseja la predicción automática es el $ARIMA(2,0,2)(0,1,1)$.

El valor de la capacidad predictiva del BIC normalizado es de 6400,6 y el AIC es de 6380,46. Estos valores no tienen sentido por sí solos y sirven para compararlo con otros modelos. Cuanto más bajo sean estos valores, mayor capacidad predictiva tiene el modelo.

En cuanto a la significatividad individual no hay problema ya que todos los parámetros son significativos. Tal vez tendríamos algún problema para la parte AR de la parte regular del segundo retardo ($ar2$). Ya que su p-valor es de 0,1434.

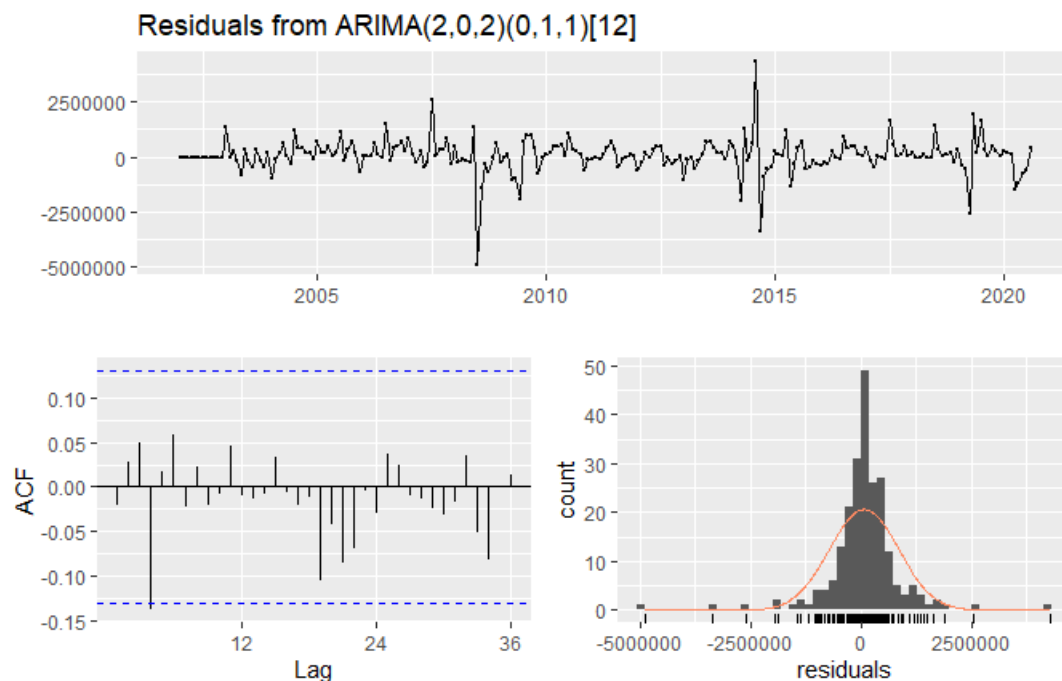
Análisis de los residuos

```
> checkresiduals(auto.arima(serieX))
```

Ljung-Box test

```
data: Residuals from ARIMA(2,0,2)(0,1,1)[12]  
Q* = 13.937, df = 19, p-value = 0.7874
```

```
Model df: 5. Total lags used: 24
```



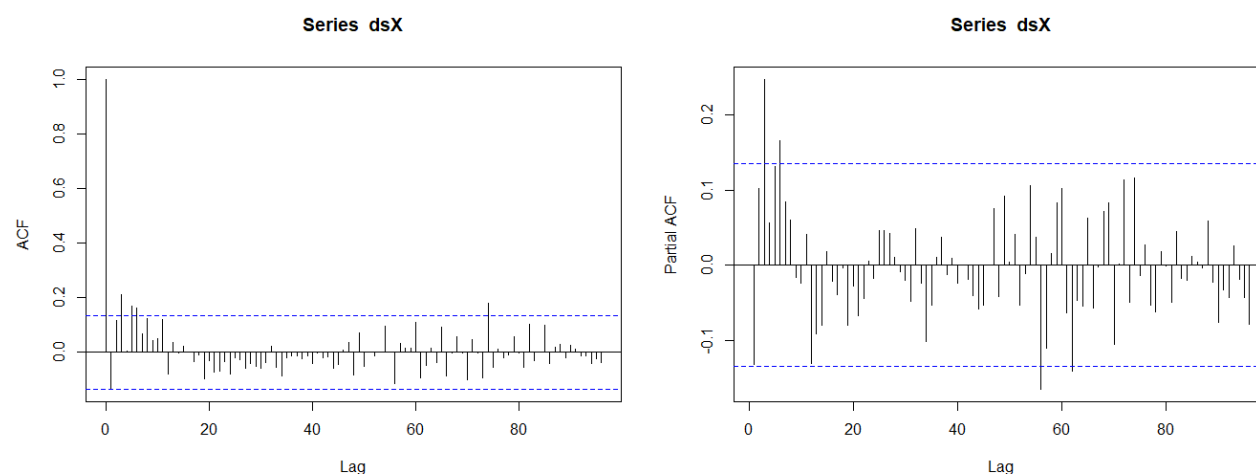
El estadístico de Ljung-Box, lo hemos visto con la función `checkresiduals()` de la librería `FORECAST`. Este estadístico nos indica la aleatoriedad residual. Su valor es muy bueno ya que supera el nivel de significación alfa igual a 0.05.

Además, viendo la gráfica de autocorrelaciones residuales, solamente un retardo se sale de las bandas de confianza, luego los residuos son muy aleatorios.

En este modelo, los residuos consiguen distribuirse normalmente.

COMPARACIÓN CON NUESTRO MODELO

Nosotros no logramos estimar este modelo porque claramente no hay 2 retardos ni en la parte MA regular, ni en la parte AR regular.



Nosotros logramos estimar los modelos $ARIMA(0,0,0)(0,1,0)$ y $ARIMA(0,0,0)(0,1,0)$. Sin embargo, ninguno de estos modelos nos mostraba una aleatoriedad residual buena (ambos nos arrojaban p-valores del estadístico de Ljung-Box inferiores a 0.05)

Por lo tanto, nos quedaríamos con el modelo de predicción automática sugerida por R. **A PESAR DE QUE NO VEMOS ESOS RETARDOS EN LAS GRÁFICAS DE AUTOCORRELACIÓN Y AUTOCORRELACIÓN PARCIAL.**

VARIABLE SOCIEDADES

```
#SOCIEDADES
```

```
serie=ts(SOCIEDADES,start=c(2002,1),frequency = 12)
```

Transformamos la variable SOCIEDADES a formato series temporales antes de realizar la predicción automática. Ponemos de argumento frecuencia 12 porque la serie ya la hemos analizado anteriormente y concluimos que era estacional.

```
auto.arima(serie)
coeftest(auto.arima(serie))
checkresiduals(auto.arima(serie))
```

Modelo propuesto

```
> auto.arima(serie)
Series: serie
ARIMA(0,0,0)(0,1,0)[12]

sigma^2 estimated as 1.622e+12: log likelihood=-3280.98
AIC=6563.96 AICc=6563.98 BIC=6567.32
> coeftest(auto.arima(serie))
Error in dimnames(x) <- dn :
  la longitud de 'dimnames' [2] no es igual a la extensión del arreglo
```

El modelo que nos aconseja la predicción automática es el ARIMA(0,0,0)(0,1,0).

El valor de la capacidad predictiva del BIC normalizado es de 6567,32 y el AIC es de 6563,96. Estos valores no tienen sentido por sí solos y sirven para compararlo con otros modelos. Cuanto más bajo sean estos valores, mayor capacidad predictiva tiene el modelo.

En cuanto a la significatividad individual no hemos podido contrastarlo.

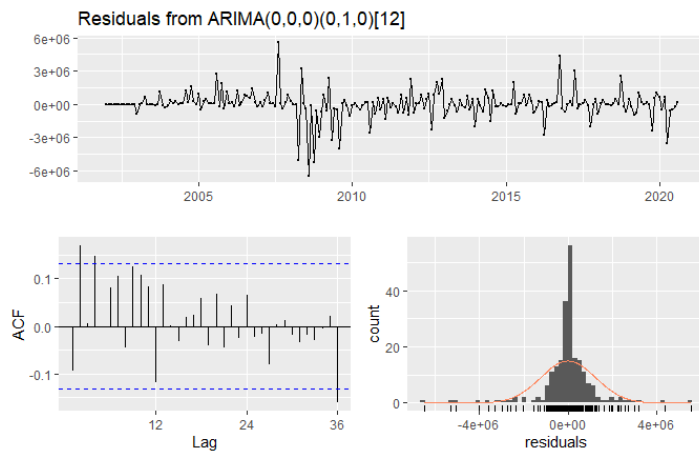
Análisis de los residuos

```
> checkresiduals(auto.arima(serie))

Ljung-Box test

data: Residuals from ARIMA(0,0,0)(0,1,0)[12]
Q* = 36.399, df = 24, p-value = 0.05018

Model df: 0. Total lags used: 24
```



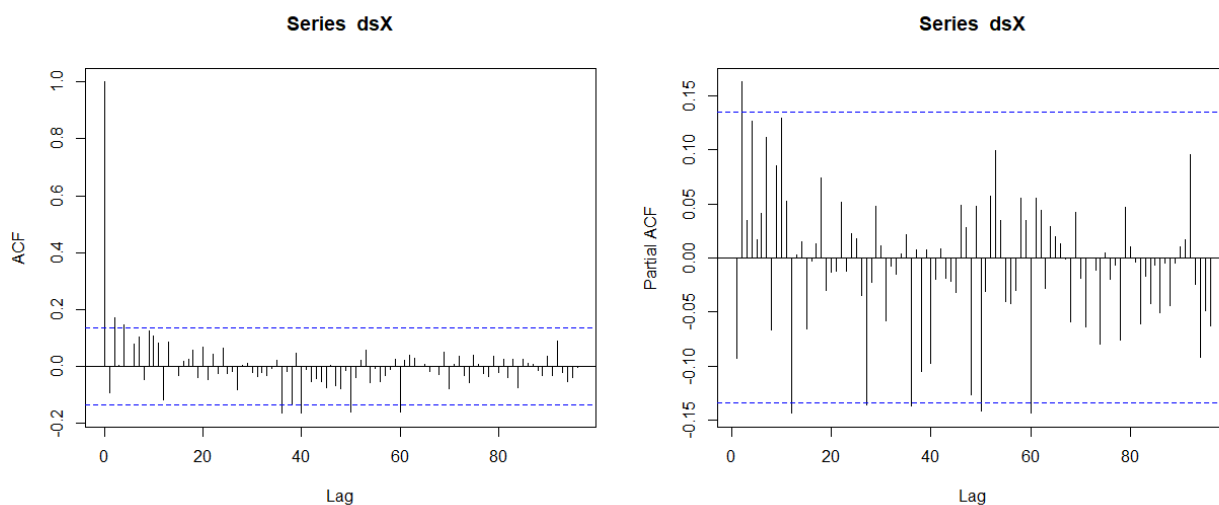
El estadístico de Ljung-Box, lo hemos visto con la función `checkresiduals()` de la librería `FORECAST`. Este estadístico nos indica la aleatoriedad residual. Su valor no es del todo bueno ya que apenas supera el nivel de significación alfa igual a 0.05.

Además, viendo la gráfica de autocorrelaciones residuales, varios retardos se salen de las bandas de confianza, luego los residuos no son muy aleatorios. En este modelo, los residuos consiguen distribuirse normalmente.

COMPARACIÓN CON NUESTRO MODELO

Nosotros logramos estimar el modelo el $ARIMA(0,0,0)(0,1,0)$ y el modelo $ARIMA(0,0,0)(1,1,0)$.

El $ARIMA(0,0,0)(0,1,0)$ nos arrojaba mayor aleatoriedad residual, sin embargo, tampoco pudimos realizar el contraste de significatividad individual



Por lo tanto, hemos coincidido con el modelo de la predicción automática. **NOS FALTARÍA CONTRASTAR LA SIGNIFICATIVIDAD INDIVIDUAL.**

VARIABLE IVA

#IVA

```
seriex=ts(IVA,start=c(2002,1),frequency = 12)
```

Transformamos la variable IVA a formato series temporales antes de realizar la predicción automática. Ponemos de argumento frecuencia 12 porque la serie ya la hemos analizado anteriormente y concluimos que era estacional.

```
auto.arima(seriex)
coeftest(auto.arima(seriex))
checkresiduals(auto.arima(seriex))
```

Modelo propuesto

```
> auto.arima(seriex)
Series: seriex
ARIMA(3,0,0)(0,1,2)[12] with drift

Coefficients:
      ar1      ar2      ar3      sma1      sma2      drift
    0.0120  0.1446  0.2273 -0.2791 -0.1524 10622.527
s.e.  0.0672  0.0682  0.0685  0.0768  0.0840  6128.927

sigma^2 estimated as 1.244e+12:  log likelihood=-3250.91
AIC=6515.82  AICc=6516.37  BIC=6539.32
> coeftest(auto.arima(seriex))

z test of coefficients:

      Estimate Std. Error z value Pr(>|z|)
ar1  1.2036e-02  6.7187e-02  0.1791 0.8578242
ar2  1.4464e-01  6.8223e-02  2.1201 0.0339955 *
ar3  2.2735e-01  6.8514e-02  3.3183 0.0009056 ***
sma1 -2.7911e-01  7.6823e-02 -3.6332 0.0002799 ***
sma2 -1.5244e-01  8.4001e-02 -1.8148 0.0695575 .
drift 1.0623e+04  6.1289e+03  1.7332 0.0830639 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

El modelo que nos aconseja la predicción automática es el ARIMA(3,0,0)(0,1,2).

El valor de la capacidad predictiva del BIC normalizado es de 6539,32 y el AIC es de 6515,82. Estos valores no tienen sentido por sí solos y sirven para compararlo con otros modelos. Cuanto más bajo sean estos valores, mayor capacidad predictiva tiene el modelo.

En cuanto a la significatividad individual tenemos problemas ya que la parte regular AR para el primer retardo es muy no significativo. Esto lo sabemos ya que

su p-valor es de 0,8578. Luego, el modelo de predicción automática falla en cuanto a significatividad individual.

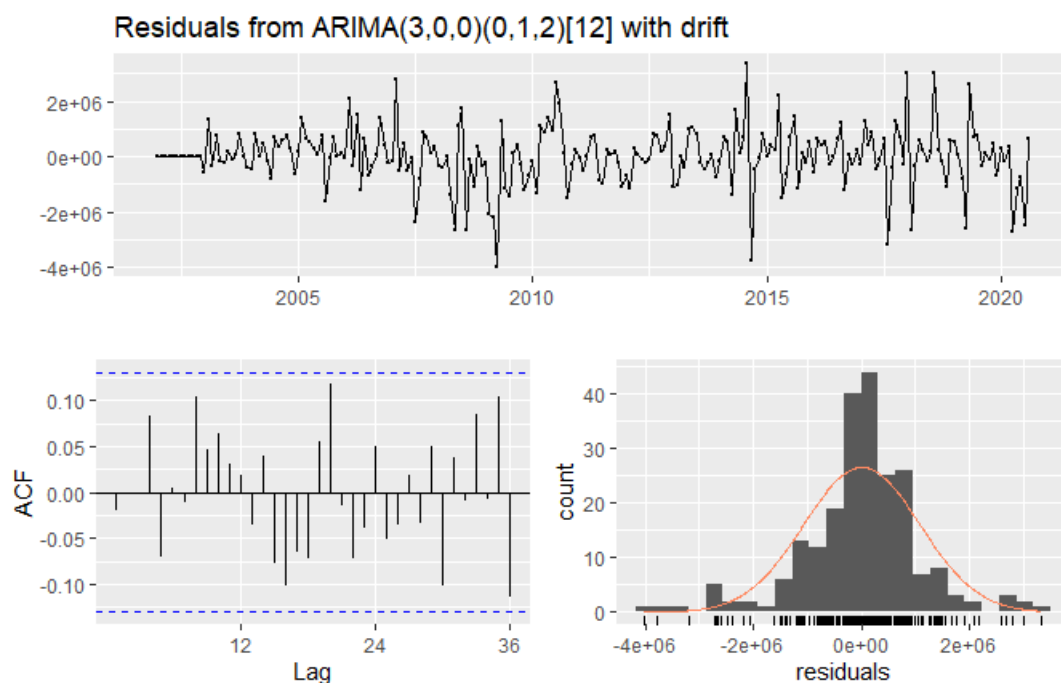
Análisis de los residuos

```
> checkresiduals(auto.arima(serieX))
```

Ljung-Box test

data: Residuals from ARIMA(3,0,0)(0,1,2)[12] with drift
Q* = 20.891, df = 18, p-value = 0.2849

Model df: 6. Total lags used: 24



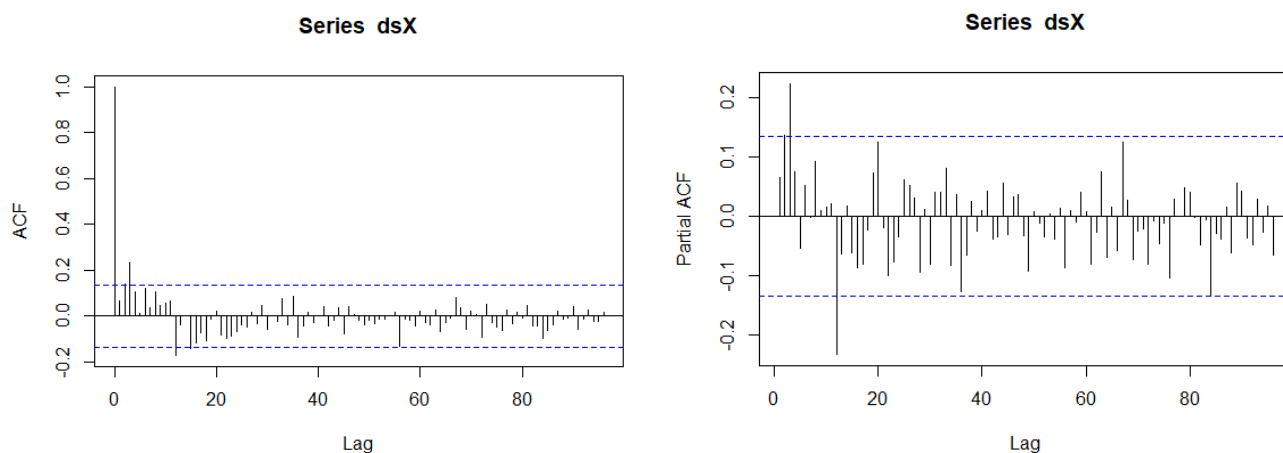
El estadístico de Ljung-Box, lo hemos visto con la función `checkresiduals()` de la librería `FORECAST`. Este estadístico nos indica la aleatoriedad residual. Su valor es muy bueno ya que supera el nivel de significación alfa igual a 0.05.

Además, viendo la gráfica de autocorrelaciones residuales, solamente un retardo se sale de las bandas de confianza, luego los residuos son muy aleatorios.

En este modelo, los residuos consiguen distribuirse normalmente.

COMPARACIÓN CON NUESTRO MODELO

Nosotros no logramos estimar este modelo porque claramente no hay 3 retardos significativos en la parte AR regular, ni tampoco hay 2 retardos significativos en la parte MA estacional.



Nosotros logramos estimar los modelos ARIMA(0,0,0)(1,1,0) y ARIMA(0,0,0)(0,1,1)

Sin embargo, ninguno de estos modelos nos mostraba una aleatoriedad residual buena (ambos nos arrojaban p-valores del estadístico de Ljung-Box inferiores a 0.05).

Por lo tanto, nos quedaríamos con el modelo de predicción automática sugerida por R. **A PESAR DE QUE NO VEMOS ESOS RETARDOS EN LAS GRÁFICAS DE AUTOCORRELACIÓN Y AUTOCORRELACIÓN PARCIAL Y QUE NOS FALLA LA SIGNIFICATIVIDAD INDIVIDUAL.**

Función de transferencia

		Correlaciones				
		INGRESOS RIBUTARIOS TOTALES	IRPF	SOCIEDADE S	IVA	ESPECIALES
INGRESOS RIBUTARIOS TOTALES	Correlación de Pearson	1	,692**	,483**	,757**	,291**
	Sig. (bilateral)		,000	,000	,000	,000
	N	224	224	224	224	224
IRPF	Correlación de Pearson	,692**	1	-,053	,366**	,260**
	Sig. (bilateral)	,000		,429	,000	,000
	N	224	224	224	224	224
SOCIEDADES	Correlación de Pearson	,483**	-,053	1	,062	,221**
	Sig. (bilateral)	,000	,429		,358	,001
	N	224	224	224	224	224
IVA	Correlación de Pearson	,757**	,366**	,062	1	,042
	Sig. (bilateral)	,000	,000	,358		,535
	N	224	224	224	224	224
ESPECIALES	Correlación de Pearson	,291**	,260**	,221**	,042	1
	Sig. (bilateral)	,000	,000	,001	,535	
	N	224	224	224	224	224

** . La correlación es significativa en el nivel 0,01 (bilateral).

Recogemos la tabla de correlaciones de SPSS para ver que variables están correladas con IVA.

Nuestras variables regresoras serán SOCIEDADES e INGRESOS TRIBUTARIOS TOTALES.

Modelo propuesto

#autor.arima con funcion de transferencia

```
predauto=auto.arima(serie,xreg=cbind(INGRESOSTRIBUTARIOSTOTALES,SOCIEDADES))
predauto
coeftest(predauto)
checkresiduals(predauto)
```

```
> predauto
Series: serie
Regression with ARIMA(2,0,1)(0,1,1)[12] errors

Coefficients:
      ar1      ar2      ma1      sma1      drift  INGRESOSTRIBUTARIOSTOTALES  SOCIEDADES
      0.7183  0.2273 -0.8510 -0.3756 -6724.915              0.5706       -0.6324
s.e.    0.0830  0.0723  0.0507  0.0698  5213.004              0.0226       0.0425

sigma^2 estimated as 3.119e+11: log likelihood=-3103.58
AIC=6223.16  AICc=6223.87  BIC=6250.01
```

La predicción automática con transferencia ha detectado el modelo ARIMA(2,0,1)(0,1,1).

El valor de la capacidad predictiva del BIC normalizado es de 6250,01 y el AIC es de 6223,16, por lo tanto, ha disminuido. Estos valores no tienen sentido por sí solos y sirven para compararlo con otros modelos. Cuanto más bajo sean estos valores, mayor capacidad predictiva tiene el modelo.

```
> coeftest(predauto)
```

z test of coefficients:

	Estimate	Std. Error	z value	Pr(> z)	
ar1	7.1826e-01	8.3038e-02	8.6498	< 2.2e-16	***
ar2	2.2734e-01	7.2308e-02	3.1441	0.001666	**
ma1	-8.5097e-01	5.0701e-02	-16.7841	< 2.2e-16	***
sma1	-3.7562e-01	6.9827e-02	-5.3793	7.477e-08	***
drift	-6.7249e+03	5.2130e+03	-1.2900	0.197041	
INGRESOSTRIBUTARIOSTOTALES	5.7055e-01	2.2556e-02	25.2955	< 2.2e-16	***
SOCIEDADES	-6.3240e-01	4.2482e-02	-14.8864	< 2.2e-16	***

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

En cuanto a la significatividad individual no hay problema ya que todos los parámetros son significativos.

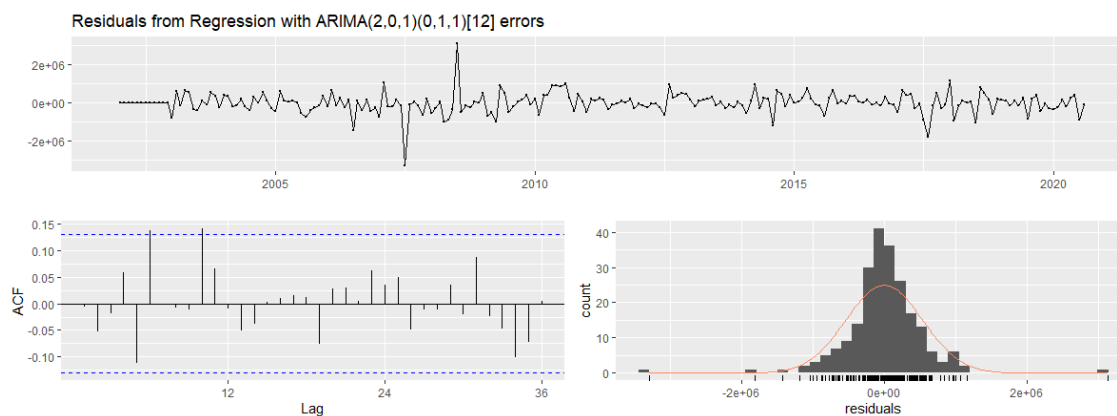
Análisis de los residuos

```
> checkresiduals(predauto)
```

Ljung-Box test

data: Residuals from Regression with ARIMA(2,0,1)(0,1,1)[12] errors
 $Q^* = 18.894$, $df = 17$, $p\text{-value} = 0.3347$

Model df: 7. Total lags used: 24



Aplicando la función de transferencia hemos conseguido un valor de aleatoriedad residual más alto, luego óptimo.

Conclusión

Nos quedamos con el modelo ARIMA con función de transferencia (ARIMAX) porque tiene mayor aleatoriedad residual y mayor capacidad predictiva que el modelo de predicción automática.

VARIABLE ESPECIALES

```
#ESPECIALES
```

```
seriex=ts(ESPECIALES,start=c(2002,1),frequency = 12)
```

Transformamos la variable ESPECIALES a formato series temporales antes de realizar la predicción automática. Ponemos de argumento frecuencia 12 porque la serie ya la hemos analizado anteriormente y concluimos que era estacional.

```
auto.arima(seriex)
coeftest(auto.arima(seriex))
checkresiduals(auto.arima(seriex))
```

Modelo propuesto

```
> auto.arima(seriex)
Series: seriex
ARIMA(1,1,1)(0,0,2)[12]

Coefficients:
      ar1      ma1      sma1      sma2
      0.1441 -0.9105  0.212  0.3329
s.e.  0.0838  0.0439  0.073  0.0749

sigma^2 estimated as 1.895e+10: log likelihood=-2955.25
AIC=5920.51 AICc=5920.78 BIC=5937.54
> coeftest(auto.arima(seriex))

z test of coefficients:

      Estimate Std. Error z value Pr(>|z|)
ar1  0.144100   0.083786   1.7199  0.085457 .
ma1 -0.910469   0.043926 -20.7271 < 2.2e-16 ***
sma1 0.212021   0.073017   2.9037  0.003688 **
sma2 0.332909   0.074918   4.4436  8.845e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

El modelo que nos aconseja la predicción automática es el ARIMA(1,1,1)(0,0,2).

El valor de la capacidad predictiva del BIC normalizado es de 5937,54 y el AIC es de 5920,51. Estos valores no tienen sentido por sí solos y sirven para compararlo con otros modelos. Cuanto más bajo sean estos valores, mayor capacidad predictiva tiene el modelo.

En cuanto a la significatividad individual no hay problema ya que todos los parámetros son significativos.

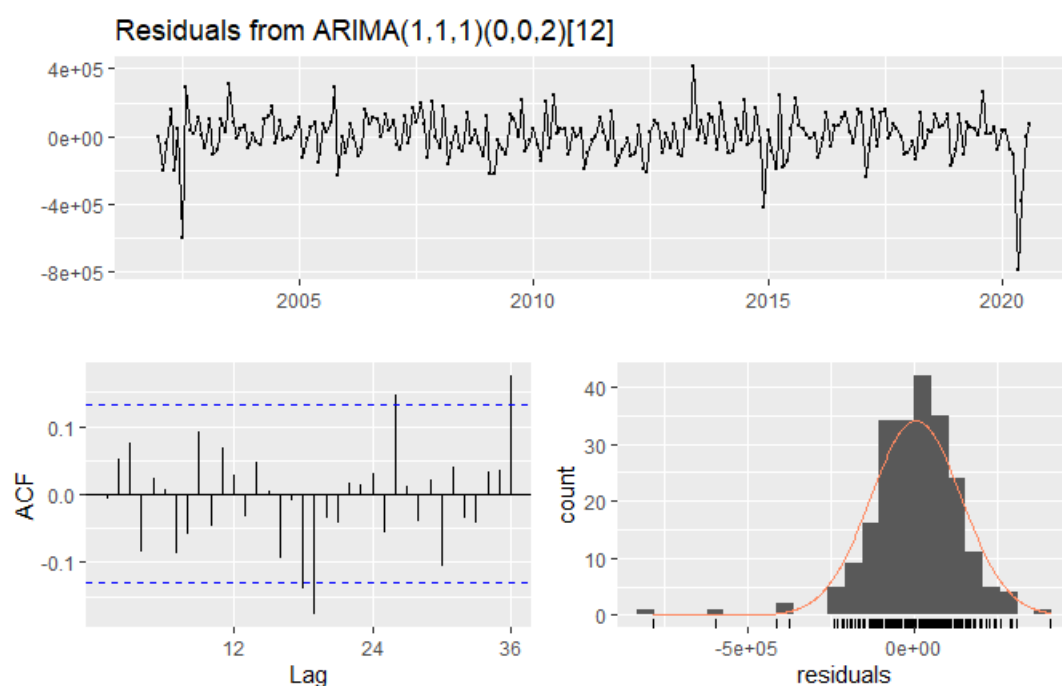
Análisis de los residuos

```
> checkresiduals(auto.arima(serieix))
```

Ljung-Box test

data: Residuals from ARIMA(1,1,1)(0,0,2)[12]
Q* = 26.811, df = 20, p-value = 0.1407

Model df: 4. Total lags used: 24



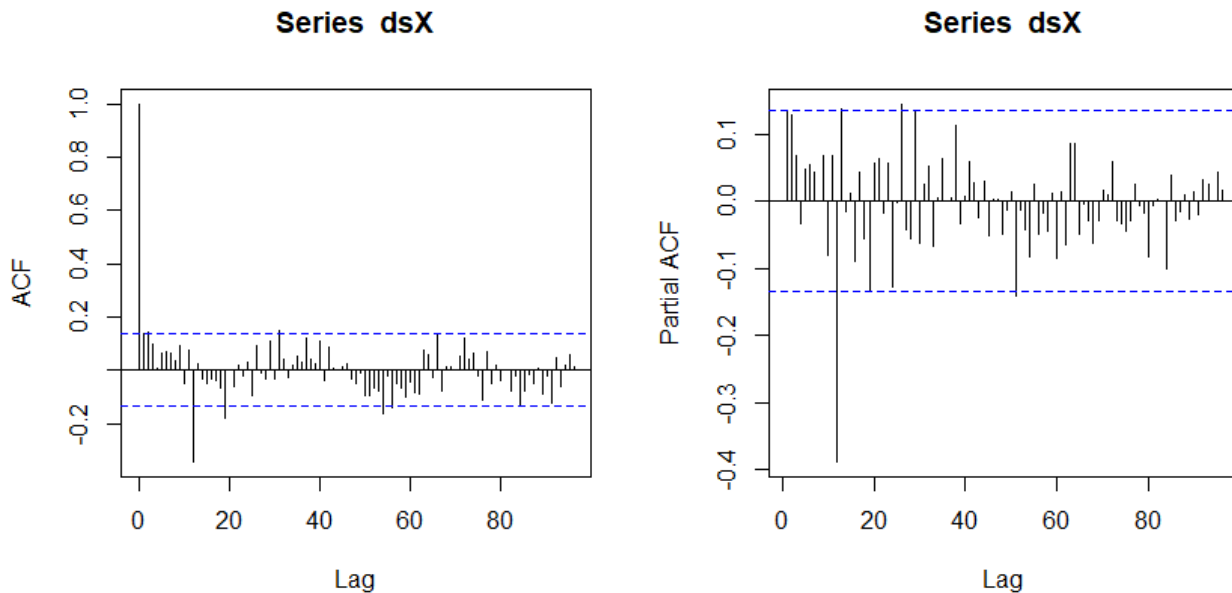
El estadístico de Ljung-Box, lo hemos visto con la función `checkresiduals()` de la librería `FORECAST`. Este estadístico nos indica la aleatoriedad residual. Su valor no es del todo bueno ya que apenas supera el nivel de significación alfa igual a 0.05.

Además, viendo la gráfica de autocorrelaciones residuales, varios retardos se salen de las bandas de confianza, luego los residuos no son muy aleatorios.

En este modelo, los residuos consiguen distribuirse normalmente.

COMPARACIÓN CON NUESTRO MODELO

Nosotros no logramos estimar este modelo porque claramente hay que aplicar diferencia estacional. Así que desde allí, ya esta todo mal.



Nosotros logramos estimar los modelos $ARIMA(1,0,0)(1,1,0)$, $ARIMA(1,0,0)(0,1,1)$, $ARIMA(0,0,2)(1,1,0)$ y $ARIMA(0,0,2)(0,1,1)$, una vez aplicada la diferencia estacional.

Sin embargo, ninguno de estos modelos nos mostraba una aleatoriedad residual buena (ambos nos arrojaban p-valores del estadístico de Ljung-Box inferiores a 0.05).

Por lo tanto, no nos quedaríamos ni con el modelo de predicción automática sugerida por R ni con nuestro modelo.

Vamos a aplicar la función de transferencia para ver si así detecta la estacionalidad de la serie.

Función de transferencia

		Correlaciones				
		INGRESOS RIBUTARIOS TOTALES	IRPF	SOCIEDADE S	IVA	ESPECIALES
INGRESOS RIBUTARIO STOTALES	Correlación de Pearson	1	,692**	,483**	,757**	,291**
	Sig. (bilateral)		,000	,000	,000	,000
	N	224	224	224	224	224
IRPF	Correlación de Pearson	,692**	1	-,053	,366**	,260**
	Sig. (bilateral)	,000		,429	,000	,000
	N	224	224	224	224	224
SOCIEDADES	Correlación de Pearson	,483**	-,053	1	,062	,221**
	Sig. (bilateral)	,000	,429		,358	,001
	N	224	224	224	224	224
IVA	Correlación de Pearson	,757**	,366**	,062	1	,042
	Sig. (bilateral)	,000	,000	,358		,535
	N	224	224	224	224	224
ESPECIALES	Correlación de Pearson	,291**	,260**	,221**	,042	1
	Sig. (bilateral)	,000	,000	,001	,535	
	N	224	224	224	224	224

** . La correlación es significativa en el nivel 0,01 (bilateral).

Recogemos la tabla de correlaciones de SPSS para ver que variables están correladas con ESPECIALES.

Realizamos un análisis factorial entre las variables INGRESOS TOTALES e IVA y nos quedamos con el factor creado.

Nuestras variables regresoras serán SOCIEDADES y FAC1_1.

Modelo propuesto

#autor.arima con funcion de transferencia

```
predauto=auto.arima(serie,xreg=cbind(FAC1_1,SOCIEDADES))
predauto
coeftest(predauto)
checkresiduals(predauto)
```

```
> predauto
Series: seriex
Regression with ARIMA(0,1,1)(2,0,0)[12] errors

Coefficients:
      ma1      sar1      sar2      FAC1_1      SOCIEDADES
      -0.8616    0.2186    0.3222   27686.18         0.0073
s.e.      0.0495    0.0762    0.0785   16650.08         0.0045

sigma^2 estimated as 1.74e+10:  log likelihood=-2945.66
AIC=5903.32   AICc=5903.71   BIC=5923.76
```

R sigue sin detectar la estacionalidad. Esto lo sabemos porque la parte MA estacional tiene un 0 cuando debería tener un 1 o más.

La predicción automática con transferencia ha detectado el modelo ARIMA(0,1,1)(2,0,0).

El valor de la capacidad predictiva del BIC normalizado es de 5923,76 y el AIC es de 5903,71, por lo tanto, ha disminuido. Estos valores no tienen sentido por sí solos y sirven

para compararlo con otros modelos. Cuanto más bajo sean estos valores, mayor capacidad predictiva tiene el modelo.

```
> coeftest(predauto)
```

```
z test of coefficients:
```

	Estimate	Std. Error	z value	Pr(> z)	
ma1	-8.6164e-01	4.9514e-02	-17.4018	< 2.2e-16	***
sar1	2.1858e-01	7.6195e-02	2.8686	0.004122	**
sar2	3.2216e-01	7.8545e-02	4.1016	4.104e-05	***
FAC1_1	2.7686e+04	1.6650e+04	1.6628	0.096347	.
SOCIEDADES	7.3243e-03	4.5031e-03	1.6265	0.103840	

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

En cuanto a la significatividad individual no hay problema ya que todos los parámetros son significativos.

Análisis de los residuos

```
> checkresiduals(predauto)
```

```
Ljung-Box test
```

```
data: Residuals from Regression with ARIMA(0,1,1)(2,0,0)[12] errors  
Q* = 26.022, df = 19, p-value = 0.1296
```

```
Model df: 5. Total lags used: 24
```

Aplicando la función de transferencia hemos conseguido un valor de aleatoriedad residual más alto, luego óptimo.

Conclusión

Nos quedamos con el modelo ARIMA con función de transferencia (ARIMAX) porque tiene mayor aleatoriedad residual y mayor capacidad predictiva que el modelo de predicción automática.