

ENTREGA SERIES TEMPORALES



UNIVERSIDAD COMPLUTENSE DE MADRID

FACULTAD DE CIENCIAS MATEMÁTICAS
MÁSTER EN INGENIERÍA MATEMÁTICA

ASIGNATURA: Estadística Aplicada y Minería de Datos

PROFESOR: Carlos Rivero Rodríguez

ALUMNAS: Siria Catherine Íñiguez Brito,
Ana Marta Oliveira dos Santos

CURSO: 2025-2026

Madrid, 11 de Diciembre de 2025

Índice

1. INTRODUCCIÓN	2
2. DATOS	3
3. AJUSTE	4
3.1. Análisis gráfico	4
3.2. Ajuste del modelo	4
4. PREDICCIÓN	16
5. CONCLUSIONES	18

1. INTRODUCCIÓN

Este trabajo se enfoca en el análisis de series temporales, una herramienta clave en la estadística y el análisis de datos, especialmente cuando se trata de estudiar temas económicos. En particular, se centra en el análisis del índice de comercio al por menor, excluyendo la venta de vehículos de motor y motocicletas.

El comercio al por menor es un sector muy importante de la economía, ya que refleja tanto el consumo de bienes y servicios por parte de las personas, como las decisiones de las empresas y las tendencias económicas generales. A través del estudio de las series temporales de estos índices, se intenta identificar patrones repetitivos, ciclos de crecimiento y factores externos que puedan influir en los cambios de la actividad comercial.

El objetivo principal de este estudio es, por tanto, aplicar técnicas avanzadas de modelización y predicción (como los modelos ARIMA/SARIMA) para analizar la trayectoria del índice y generar pronósticos. Más allá del ejercicio predictivo, el trabajo tiene un carácter académico y formativo, cuyo propósito es profundizar en la metodología de análisis de series temporales, evaluar la precisión de diferentes enfoques de modelado y extraer conclusiones robustas sobre el posible comportamiento futuro de un indicador clave para la economía.

2. DATOS

El conjunto de datos seleccionado para el estudio corresponde al **índice de comercio al por menor, excluyendo la venta de vehículos de motor y motocicletas**. Estos datos han sido obtenidos de la página oficial del INE <https://www.ine.es/jaxiT3/Tabla.htm?t=60096> atendiendo a las siguientes opciones:

- **Actividad según jerarquía (CNAE-2009):** Nivel 1
- **Categoría:** 47 Comercio al por menor, excepto de vehículos de motor y motocicletas
- **Tipo de datos:** Índices
- **Periodo:** Desde enero de 2006 (2006M01) hasta septiembre de 2025 (2025M09)

Estos datos fueron exportados a un archivo Excel denominado `ic.xlsx`, al que se le añadió una columna adicional llamada FECHANUEVA. Se decide crear esta nueva fecha tomando el primer día de cada mes para que luego pueda ser interpretada por SAS, aunque en realidad los datos corresponden al final del mes.

Los datos se registran mensualmente, lo que permite analizar la evolución del índice de comercio al por menor a lo largo del tiempo. Este índice, correspondiente al nivel 1 de la clasificación CNAE-2009, constituye una medida clave para entender la dinámica y el comportamiento del sector comercial.

La siguiente figura muestra la gráfica de la serie temporal del índice de comercio al por menor:

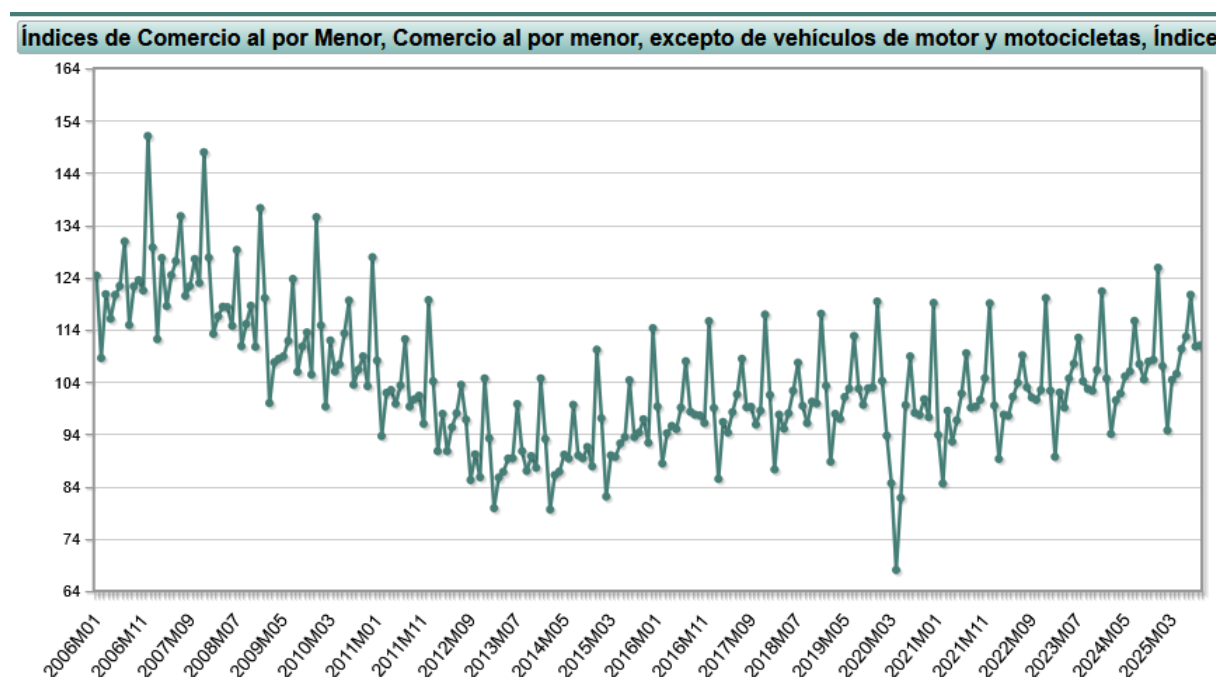


Figura 1: Gráfica de la serie temporal del Índice de Comercio al por menor.

3. AJUSTE

3.1. Análisis gráfico

Primeramente, antes de comenzar con el ajuste del modelo, se realiza un análisis gráfico preliminar con el fin de obtener una primera aproximación al comportamiento de la serie temporal. Este análisis permite identificar características como la estacionariedad, la estacionalidad y la presencia de posibles valores atípicos, entre otros aspectos relevantes.

Atendiendo a la gráfica de la Figura 1, se observa que la serie presenta un patrón repetitivo a lo largo de los años: en febrero alcanza sus valores más bajos, mientras que en diciembre experimenta un incremento notable en comparación con los valores de los demás meses del mismo año.

Por un lado, en diciembre, debido al periodo navideño, los hogares suelen aumentar significativamente sus gastos en regalos, celebraciones, y otros eventos festivos. Además, en muchos casos, los trabajadores reciben la paga extraordinaria de Navidad, lo que amplía su capacidad de consumo. Por otro lado, en febrero el gasto de los hogares tiende a reducirse: tras las festividades y las compras producidas por las rebajas de enero, los presupuestos se encuentran más ajustados. Este efecto se ve acentuado por el hecho de que febrero es el mes más corto, lo cual se traduce directamente en un menor volumen total de comercio.

Además, también se puede apreciar que el segundo mes con mayor actividad comercial es julio, lo cual es consistente con la realidad del mercado debido al efecto de las rebajas de verano. No obstante, su impacto es menor que el observado en las rebajas de enero y, especialmente, en la campaña navideña.

El análisis exploratorio inicial identifica la presencia de una estacionalidad marcada con patrón anual, evidenciada por un ciclo de 12 meses. Esta característica sugiere que los coeficientes de autocorrelación para los rezagos estacionales (12, 24, 36...) serán estadísticamente significativos y de magnitud considerable.

Respecto a los valores atípicos, se identifica un descenso pronunciado que se inicia en marzo de 2020 y se prolonga durante abril y mayo. Este patrón es claramente atribuible al impacto de la pandemia de COVID-19, ya que las medidas de restricción implementadas provocaron una caída abrupta en la actividad comercial. La serie comenzó a recuperar su trayectoria habitual tras este período excepcional de aproximadamente tres meses.

3.2. Ajuste del modelo

Esta sección tiene como objetivo documentar el proceso iterativo de identificación de un modelo ARIMA adecuado que capture la dinámica de la serie temporal, considerando también intervenciones exógenas, empleando la metodología Box-Jenkins.

Antes de iniciar el ajuste del modelo es necesario comprobar la estacionariedad de la serie. En primer lugar, se analiza la estacionariedad en varianza, evaluando la necesidad de aplicar una transformación. Para ello se emplearon dos procedimientos (boxcar y logtest) cuyos resultados se muestran en la Figura 2.

LAMBDA	LOGLIK	RMSE	AIC	SBC
1.0	-795.779	69.7371	1603.56	1624.05
0.5	-792.139	69.6833	1596.28	1616.77
0.0	-790.640	69.8618	1593.28	1613.78

TRANS	LOGLIK	RMSE	AIC	SBC
NONE	-795.779	69.7371	1603.56	1624.05
LOG	-790.640	69.8618	1593.28	1613.78

Figura 2: Contraste para aplicar una transformación

En vista de los resultados de las pruebas estadísticas, se recomienda aplicar una transformación logarítmica a la serie con el propósito de estabilizar la varianza y cumplir con los supuestos de estacionariedad.

En segundo lugar, se analiza la estacionariedad en media. En esta etapa se evalúa la necesidad de diferenciar la serie mediante la realización del test de Dickey-Fuller, un test de raíces unitarias. Primero, se examina el gráfico asociado al test (Figura 3), y luego se analizan los resultados obtenidos mediante la ejecución del código (Figuras 4 y 5).

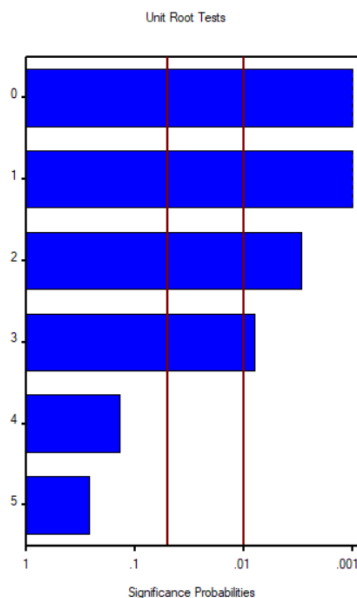


Figura 3: Gráfico de Probabilidades

```

*test de raíces unitarias;
%dfest(TesteTl.ic_entrenamiento,INDICE,dlag=1,ar=4,trend=0);
%put $dfpvalue;

proc arima data=TesteTl.ic_entrenamiento;
identify var=indice stationarity=(adf=(1 2 3 4 5) dlag=1);
run;
quit;

```

Figura 4: Código SAS para el Test de Raíces Unitarias

Pruebas aumentadas de la raíz unidad de Dickey-Fuller							
Tipo	Retardos	Rho	Pr < Rho	Tau	Pr < Tau	F	Pr > F
Media cero	1	-0.6848	0.5307	-0.62	0.4497		
	2	-0.3923	0.5932	-0.52	0.4882		
	3	-0.3100	0.6117	-0.46	0.5148		
	4	-0.2057	0.6354	-0.49	0.5017		
	5	-0.1726	0.6430	-0.51	0.4935		
Media simple	1	-56.6835	0.0014	-5.31	<.0001	14.09	0.0010
	2	-29.6850	0.0014	-3.76	0.0039	7.10	0.0010
	3	-26.2632	0.0022	-3.44	0.0108	5.93	0.0143
	4	-10.6988	0.1102	-2.34	0.1617	2.76	0.3668
	5	-7.2983	0.2526	-1.98	0.2964	2.01	0.5594
Tendencia	1	-74.8580	0.0006	-6.03	<.0001	18.20	0.0010
	2	-39.2401	0.0007	-4.14	0.0065	8.66	0.0010
	3	-36.0630	0.0016	-3.80	0.0181	7.32	0.0236
	4	-12.5821	0.2756	-2.27	0.4479	2.88	0.6020
	5	-7.7809	0.5952	-1.75	0.7242	1.96	0.7855

Figura 5: Resultados del Test Dickey-Fuller

A partir de la Figura 3, se observa que bajo algunas estructuras AR(4,5) pueden existir raíces unitarias. El valor obtenido en el primer bloque de código ha sido: 0.5016877614. Este valor, junto con la información de la Figura 5, sugieren realizar una diferencia regular. No obstante, la diferencia estacional no parece necesaria.

Sin embargo, dada que los tests de raíces unitarias presentan una potencia limitada, se optó por iniciar el proceso de ajuste analizando directamente las funciones de autocorrelación simple (f.a.s) y autocorrelación parcial (f.a.p) de la serie previamente transformada.

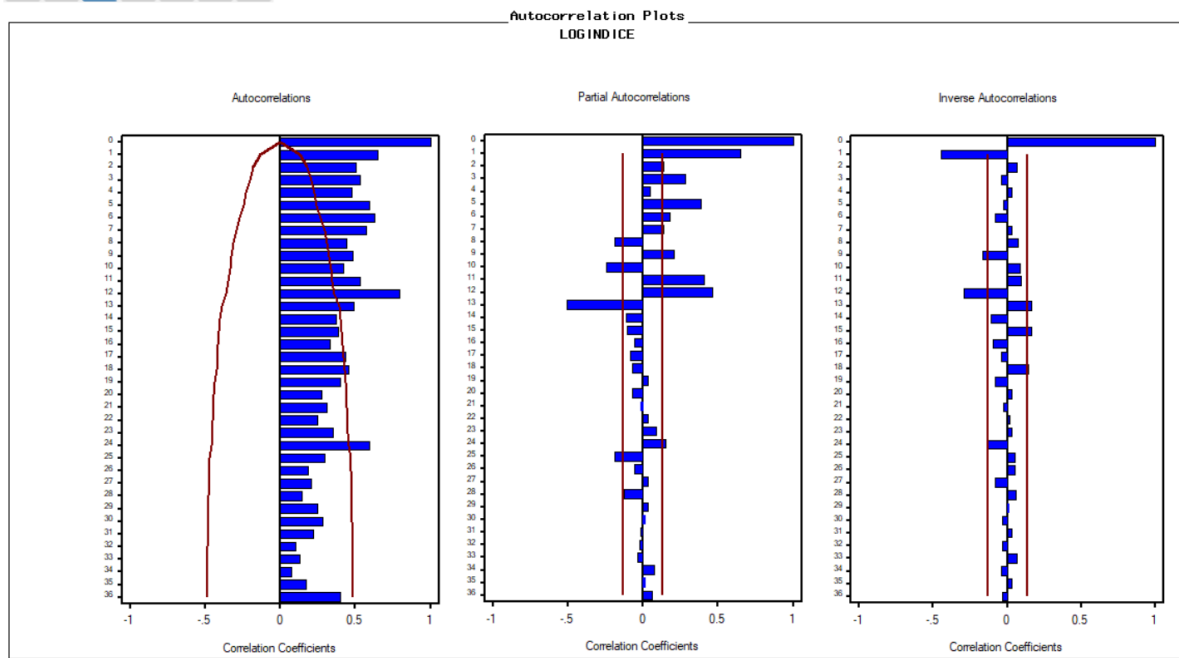


Figura 6: F.A.S y F.S.P serie transformada

El análisis de la función de autocorrelación simple (FAS) revela que los rezagos estacionales (12 y 24) presentan los valores más elevados, un patrón que apunta a la existencia de un componente autorregresivo estacional. Esta evidencia se ve reforzada por la función de autocorrelación parcial (FAP), donde el rezago 13 también muestra significatividad, sugiriendo la posible presencia de una estructura AR estacional.

Cabe señalar que los rezagos adyacentes (11 y 13) también presentan autocorrelaciones considerables. Este fenómeno puede atribuirse a la fuerte dependencia en el rezago estacional de orden 12, cuya influencia se propaga a sus retardos más cercanos.

Además, el descenso lento y persistente observado en la FAS confirma que la serie no es estacionaria en media. Este hallazgo valida la necesidad de aplicar una diferenciación regular, tal y como se había intuido en la fase de exploración gráfica inicial.

No obstante, se propone en primer lugar un modelo con AR(12), equivalente a un ARIMA(1,0,0)s.

ARIMA(1,0,0)s

Model Parameter	Estimate	Std. Error	T	Prob> T
Intercept	4.67826	0.0235	198.7498	<.0001
Seasonal Autoregressive, Lag 12	0.87499	0.0276	31.6770	<.0001
Model Variance (sigma squared)	0.00367	.	.	.

Figura 7: Parámetros estimados - ARIMA(1,0,0)s

Al estimar los parámetros del modelo, todos resultan significativos tal como se observa en la Figura 7. La estimación del parámetro AR estacional es relativamente alta (0.87499), lo que, en principio, podría sugerir la necesidad de una diferenciación estacional. Sin embargo, el intervalo de confianza correspondiente

$$(0,87499 - 2 * 0,0276; 0,87499 + 2 * 0,0276) = (0,81979; 0,93019)$$

no contiene el 1, por lo que, de momento, se descarta aplicar la diferenciación.

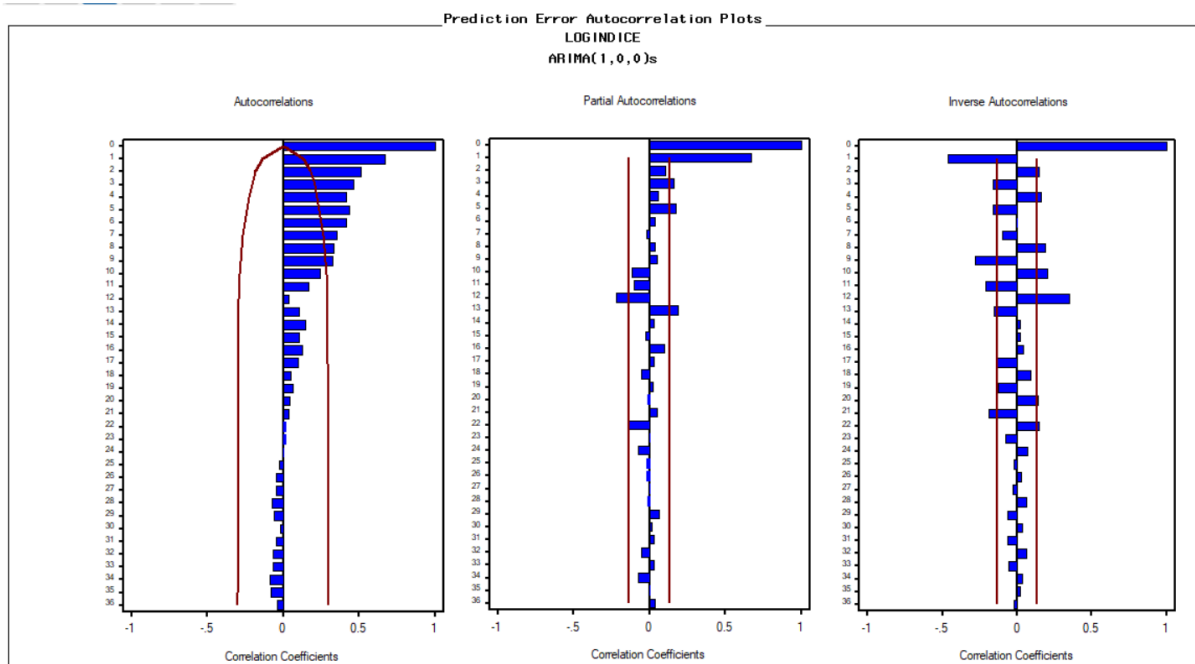


Figura 8: F.A.S y F.A.P - ARIMA(1,0,0)s

Tras examinar las funciones obtenidas en la Figura 8, se aprecia nuevamente que la F.A.S decae exponencialmente, siendo este un síntoma característico de procesos autoregresivos. Por su parte, la F.A.P muestra una autocorrelación parcial de orden 1 significativamente alta. En consecuencia, se decide ajustar un modelo AR(1).

Nota: Cada vez que se ajusta un nuevo modelo, se siguen los mismos pasos: análisis de significancia de los parámetros y, posteriormente, evaluación de las funciones de autocorrelación simple y parcial.

ARIMA(1,0,0)x(1,0,0)s

Model Parameter	Estimate	Std. Error	T	Prob> T
Intercept	4.67834	0.0556	84.1702	<.0001
Autoregressive, Lag 1	0.69547	0.0459	15.1461	<.0001
Seasonal Autoregressive, Lag 12	0.87687	0.0278	31.5062	<.0001
Model Variance (sigma squared)	0.00189	.	.	.

Figura 9: Parámetros - ARIMA(1,0,0)x(1,0,0)s

Las nuevas estimaciones tal como se aparece en la Figura 9 muestran que todos los parámetros son significativos y que ninguno de sus intervalos de confianza contiene el valor 1.

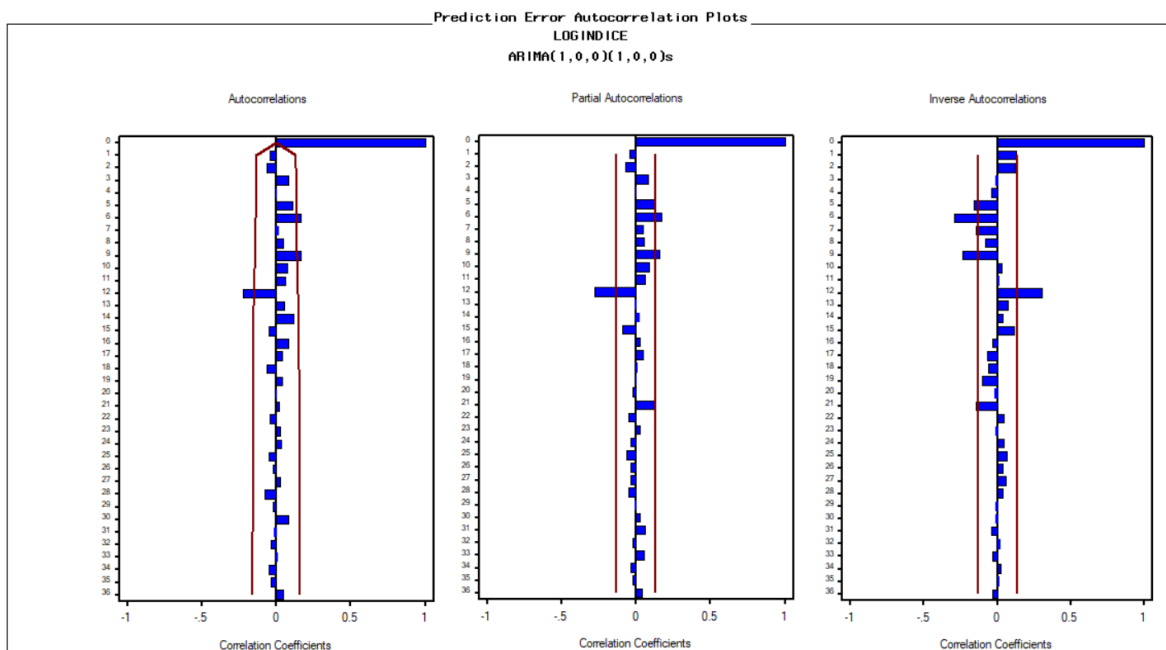


Figura 10: F.A.S y F.A.P. - ARIMA(1,0,0)x(1,0,0)s

A partir de las funciones F.A.S y F.A.P. mostradas en la Figura 10 se aprecia que la autocorrelación de orden 12 vuelve a ser dominante. Esto podría indicar un componente MA(1) estacional, es decir, un MA(12).

ARIMA(1,0,0)x(1,0,1)s

Al intentar ajustar este modelo en la interfaz de SAS se produce un problema de convergencia durante el proceso de estimación de parámetros. Para investigar esta situación, se ejecuta el correspondiente código en SAS:

```
*error en ARIMA(1,0,0)(1,0,1)s;
proc arima data=TesteT1.ic_entrenamiento;
  identify var=logIndice;
  estimate p =(1) (12) q=(12) plot;
run;
```

Figura 11: Código SAS

Estimación por mínimos cuadrados condicional					
Parámetro	Estimación	Error estándar	t valor	Aprox. Pr > t	Retardo
MU	4.85459	0.03352	144.83	<.0001	0
MA1,1	0.36202	0.07165	5.05	<.0001	12
AR1,1	0.75899	0.04599	16.50	<.0001	1
AR2,1	0.97823	0.02168	45.13	<.0001	12

Figura 12: Parámetros

El análisis de los resultados revela que el intervalo de confianza del parámetro AR(2) (identificado como AR2,1 en la salida) incluye el valor 1:

$$(0,97823 - 2 * 0,0216; 0,97823 + 2 * 0,0216) = (0,93503; 1,02143)$$

Se concluye que la serie requiere una diferencia estacional. En consecuencia, en el modelo se sustituye el término autoregresivo estacional AR(12) por una diferencia estacional.

ARIMA(1,0,0)x(0,1,1)s

Model Parameter	Estimate	Std. Error	T	Prob> T
Intercept	-0.00885	0.0070	-1.2676	0.2064
Seasonal Moving Average, Lag 12	0.73881	0.0565	13.0803	<.0001
Autoregressive, Lag 1	0.88971	0.0322	27.6446	<.0001
Model Variance (sigma squared)	0.00145	.	.	.

Figura 13: Parámetros ARIMA(1,0,0)x(0,1,1)s con Intercept

Inicialmente se estimó este modelo incluyendo el parámetro *Intercept*. Sin embargo, dado que dicho parámetro no resultó significativo (Veáse Figura 13), se eliminó del modelo de manera definitiva. Una vez ajustado el modelo sin intercept, todos los parámetros restantes se mantienen significativos 14).

Model Parameter	Estimate	Std. Error	T	Prob> T
Seasonal Moving Average, Lag 12	0.73494	0.0560	13.1355	<.0001
Autoregressive, Lag 1	0.90175	0.0306	29.4942	<.0001
Model Variance (sigma squared)	0.00145	.	.	.

Figura 14: Parámetros - ARIMA(1,0,0)x(0,1,1)s

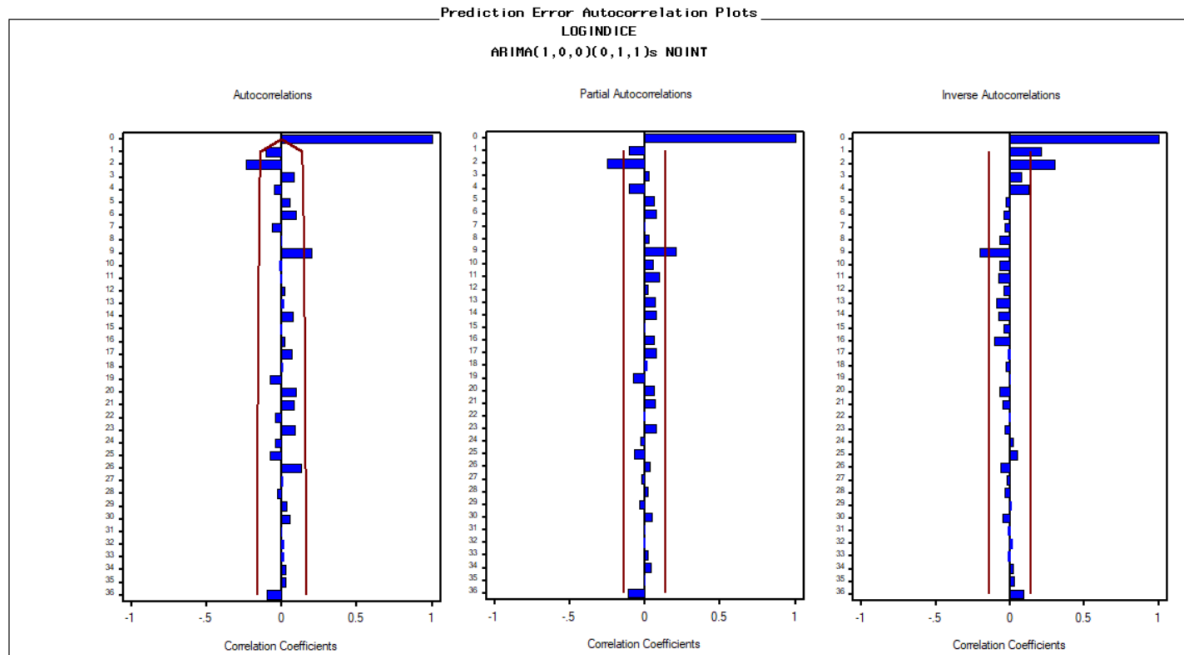


Figura 15: F.A.S y F.A.P - ARIMA(1,0,0)x(0,1,1)s

A partir de la F.A.S y F.A.P, se observa que las autocorrelaciones de orden 2 se situán fuera de las bandas de confianza, lo cual sugiere la presencia de un componente MA(2). Asimismo, se detectan autocorrelaciones significativas de orden 9, aunque estas podrían estar relacionadas con la presencia de datos atípicos previamente identificados. En consecuencia, se decide introducir un MA(2).

$$\text{ARIMA}(1,0,2)\times(0,1,1)s + \text{NOINT}$$

Model Parameter	Estimate	Std. Error	T	Prob> T
Moving Average, Lag 1	0.23574	0.0643	3.6681	0.0003
Moving Average, Lag 2	0.37370	0.0643	5.8134	<.0001
Seasonal Moving Average, Lag 12	0.77928	0.0558	13.9683	<.0001
Autoregressive, Lag 1	0.98850	0.0110	89.6125	<.0001
Model Variance (sigma squared)	0.00126	.	.	.

Figura 16: Parámetros ARIMA(1,0,2)x(0,1,1)s + NOINT

En este ajuste, todos los parámetros estimados han resultado significativos tal como se observa en la Figura 16. No obstante, el intervalo de confianza centrado en la estimación del parámetro AR(1) regular, con amplitud el doble de su desviación típica, incluye el 1. Este hecho indica la necesidad de reemplazar el término AR(1) regular por una diferencia regular.

ARIMA(0,1,2)x(0,1,1)s + NOINT

Model Parameter	Estimate	Std. Error	T	Prob> T
Moving Average, Lag 1	0.24048	0.0626	3.8436	0.0002
Moving Average, Lag 2	0.37886	0.0628	6.0367	<.0001
Seasonal Moving Average, Lag 12	0.78168	0.0532	14.7030	<.0001
Model Variance (sigma squared)	0.00126	.	.	.

Figura 17: Parámetros - ARIMA(0,1,2)x(0,1,1)s

Tras la estimación, todos los parámetros del modelo son estadísticamente significativos, (Veáse Figura 17) y ninguno de sus intervalos de confianza incluye al 1, confirmando la estabilidad de la especificación. Asimismo, los residuos exhiben las propiedades de un ruido blanco (Veáse Figura 19).

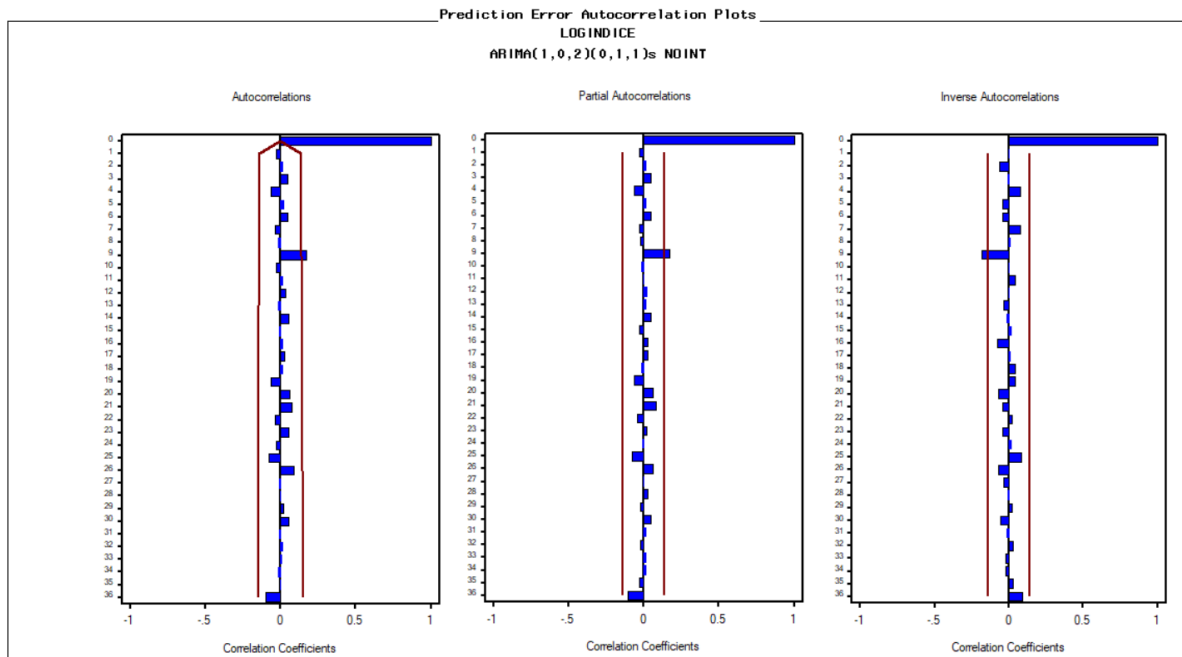


Figura 18: F.A.S y F.A.P - ARIMA(0,1,2)x(0,1,1)s

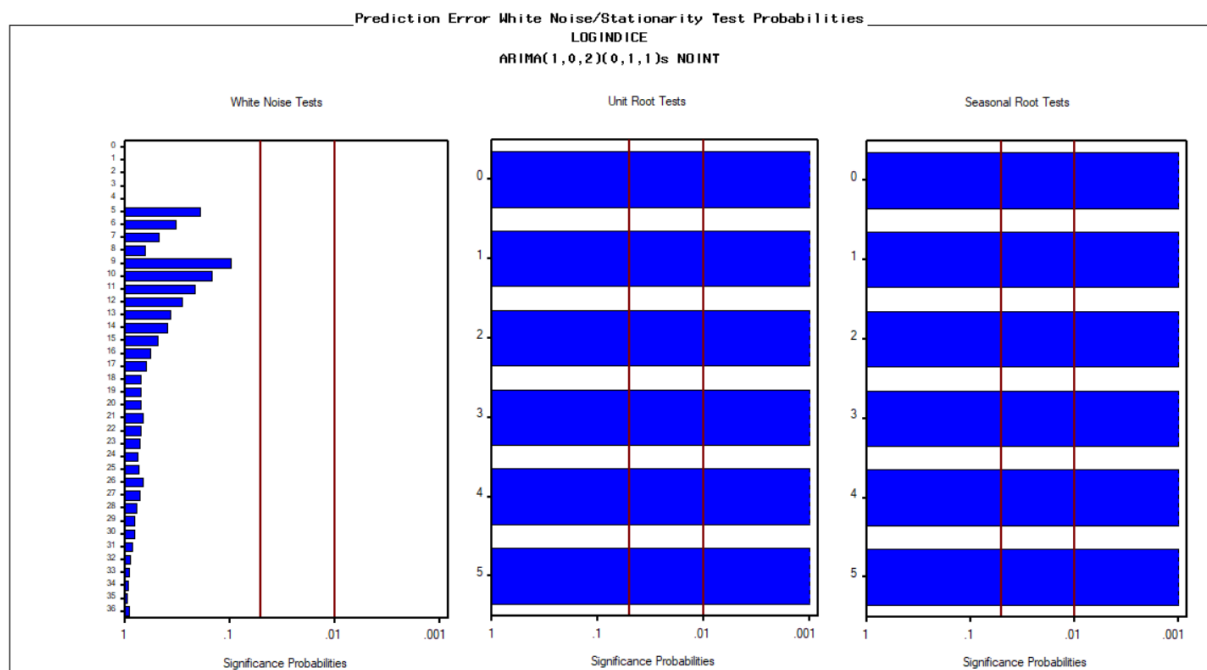


Figura 19: Test Ruido Blanco y Raíces Unitarias

No obstante, se detecta una autocorrelación residual significativa en el rezago 9, tanto en la función de autocorrelación simple (FAS) como en la parcial (FAP). Se plantea la hipótesis de que pueda deberse a un efecto atípico o de intervención. Para verificar esta hipótesis y modelar adecuadamente el impacto, se procede a realizar un análisis de intervención y para ello se introducen las intervenciones correspondientes al periodo afectado por la COVID-19:

```
*regresores;
Data Trabajo1.ic_entrenamiento_regresores;
SET Trabajo1.ic_entrenamiento;
COVID = (FECHANUEVA >= '01MAR2020'd and FECHANUEVA <= '01MAY2020'd);
run;
```

Figura 20: Código para Regresores

A continuación, se presenta un resumen de los valores de RMSE obtenidos para los distintos modelos estimados hasta este punto. (Figura 21).

Forecast Model	Model Title	Root Mean Square Error
<input type="checkbox"/>	ARIMA(1,0,0)s	0.06770
<input type="checkbox"/>	ARIMA(1,0,0)(1,0,0)s	0.04746
<input type="checkbox"/>	ARIMA(1,0,0)(0,1,1)s	0.03836
<input type="checkbox"/>	ARIMA(1,0,0)(0,1,1)s NOINT	0.03845
<input checked="" type="checkbox"/>	ARIMA(1,0,2)(0,1,1)s NOINT	0.03575
<input type="checkbox"/>	ARIMA(0,1,2)(0,1,1)s NOINT	0.03589

Figura 21: Comparación de modelos mediante RMSE

ARIMA(0,1,2)x(0,1,1)s+COVID

Model Parameter	Estimate	Std. Error	T	Prob> T
Intercept	0.0001802	0.000316	0.5702	0.5692
Moving Average, Lag 1	0.83694	0.0662	12.6391	<.0001
Moving Average, Lag 2	-0.30562	0.0659	-4.6345	<.0001
Seasonal Moving Average, Lag 12	0.66870	0.0568	11.7659	<.0001
COVID	-0.25128	0.0131	-19.1217	<.0001
Model Variance (sigma squared)	0.0006987	.	.	.

Figura 22: Parámetros estimados ARIMA(0,1,2)x(0,1,1)s+COVID

Al estimar este modelo, el parámetro Intercept vuelve a no ser significativo como se muestra en la Figura 22, por lo que es eliminado. El modelo final considerado es:

ARIMA(0,1,2)x(0,1,1)s+COVID NOINT

Model Parameter	Estimate	Std. Error	T	Prob> T
Moving Average, Lag 1	0.83621	0.0661	12.6539	<.0001
Moving Average, Lag 2	-0.30703	0.0658	-4.6681	<.0001
Seasonal Moving Average, Lag 12	0.66207	0.0569	11.6434	<.0001
covid	-0.25137	0.0131	-19.1810	<.0001
Model Variance (sigma squared)	0.0006970	.	.	.

Figura 23: Parámetros Estimados ARIMA(0,1,2)x(0,1,1)s+COVID NOINT

Con esta nueva especificación, todos los parámetros estimados resultan estadísticamente significativos pudiéndose comprobar en los resultados obtenido en la Figura 23. Además, se resuelve el problema de autocorrelación residual previamente identificado: para el rezago 9, que antes se situaban fuera de las bandas de confianza, ahora se encuentran dentro de los límites establecidos tocándolos ligeramente. Por último, los diagnósticos de los residuos, representados en la Figura 25, confirman que estos se comportan como un ruido blanco, validando así la idoneidad del modelo final.

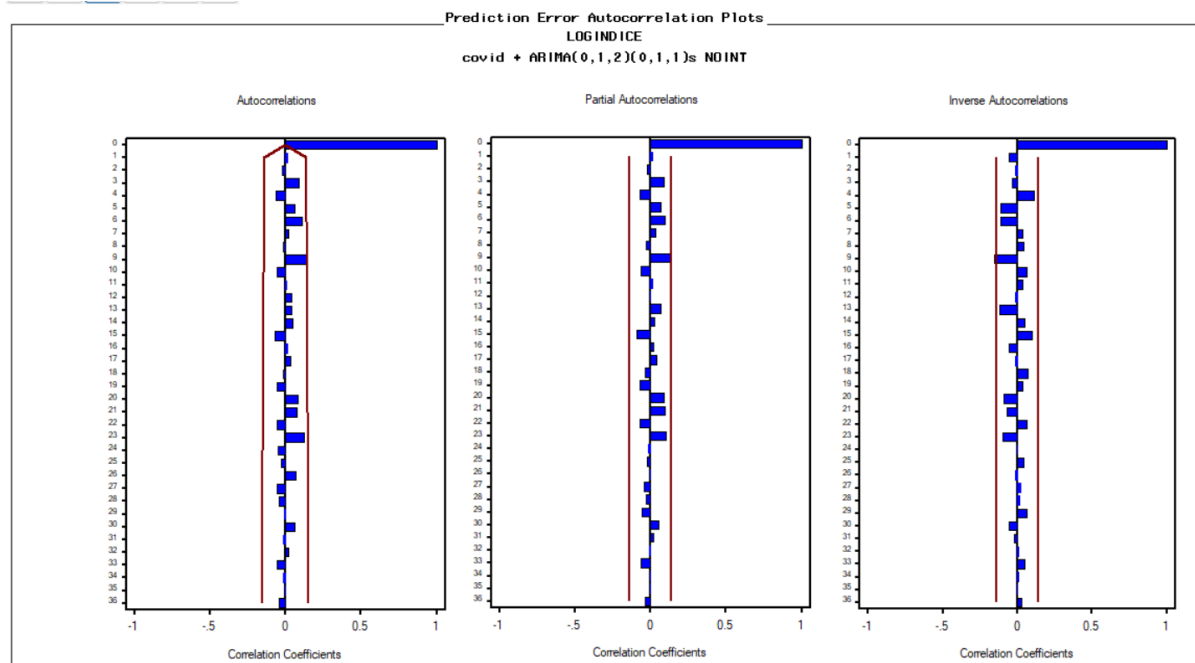


Figura 24: F.A.S y F.A.P - ARIMA(0,1,2)x(0,1,1)s+COVID NOINT

Como se señaló anteriormente, en la Figura 25 se aprecia la obtención de ruido blanco, tras añadir la variable exógenas COVID:

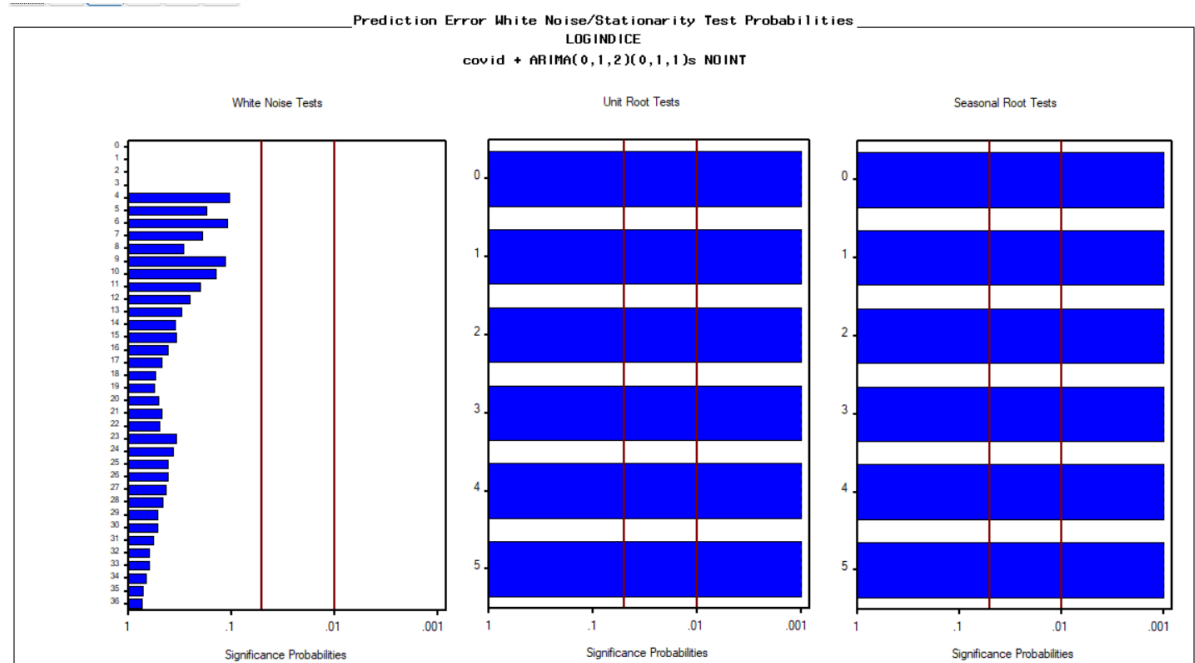


Figura 25: Test Ruido Blanco y Raices Unitarias - final

Para concluir el proceso de validación, se aplican pruebas formales de diagnóstico a los residuos del modelo. Los resultados no rechazan la hipótesis nula de ruido blanco, confirmando que los residuos no están correlacionados, tienen media cero y varianza constante. Esto permite concluir que el modelo satisface los supuestos esenciales y captura de manera apropiada la estructura de dependencia de la serie.

```
/*Comprobar final con regresores*/
proc arima data=TesteI1.ic_entrenamiento_regresores;
identify var=logINDICE(1 12)
crosscorr=(COVID(1 12));
estimate plot q = (1 2)(12) noint
input=(COVID)
method=ML;
run;
```

Figura 26: Código SAS

Comprobación de autocorrelación de residuales									
Para retardo	Chi-cuadrado	DF	Pr > ChiSq	Autocorrelaciones					
6	5.38	3	0.1458	0.009	-0.028	0.075	-0.065	0.058	0.102
12	10.69	9	0.2973	0.019	-0.025	0.132	-0.059	-0.001	0.042
18	13.04	15	0.5989	0.038	0.043	-0.073	0.015	0.031	-0.018
24	22.08	21	0.3951	-0.055	0.087	0.081	-0.061	0.121	-0.048
30	25.56	27	0.5432	-0.034	0.066	-0.051	-0.048	-0.006	0.061
36	27.37	33	0.7434	-0.019	0.018	-0.062	-0.017	-0.003	-0.048
42	29.74	39	0.8573	0.037	-0.018	-0.024	0.052	-0.054	0.033

Figura 27: Resultados de autocorrelación residuales

En consecuencia, se selecciona este modelo frente al ARIMA(0,1,2)x(0,1,1)s sin intercepto, ya que presenta un comportamiento más adecuado en varios aspectos clave. Aunque los valores-p asociados a algunas autocorrelaciones son ligeramente más elevados, el modelo mantiene residuos con comportamiento de ruido blanco, las autocorrelaciones en el rezago 9 muestran una mejora notable, y, adicionalmente, el valor RMSE es inferior. Estos elementos combinados justifican la elección del modelo final.

Forecast		Root Mean Square Error	
Model	Model Title		
<input checked="" type="checkbox"/>	COVID + ARIMA(0,1,2)(0,1,1)s covid + ARIMA(0,1,2)(0,1,1)s NOINT		0.02646 0.02648

Figura 28: Comparación de modelos mediante RMSE

Utilizando las estimaciones obtenidas de los parámetros, el modelo final obtenido es entonces el siguiente:

$$(1 - B)(1 - B^{12})\text{LOGINDICE}_t = (1 - 0,83621B + 0,30703B^2)(1 - 0,66207B^{12})\epsilon_t - 0,25137 \cdot (1 - B)(1 - B^{12})\text{COVID}_t$$

4. PREDICCIÓN

En esta sección, se analizan las predicciones generadas por el modelo ajustado, así como sus errores y el comportamiento de las bandas de confianza.

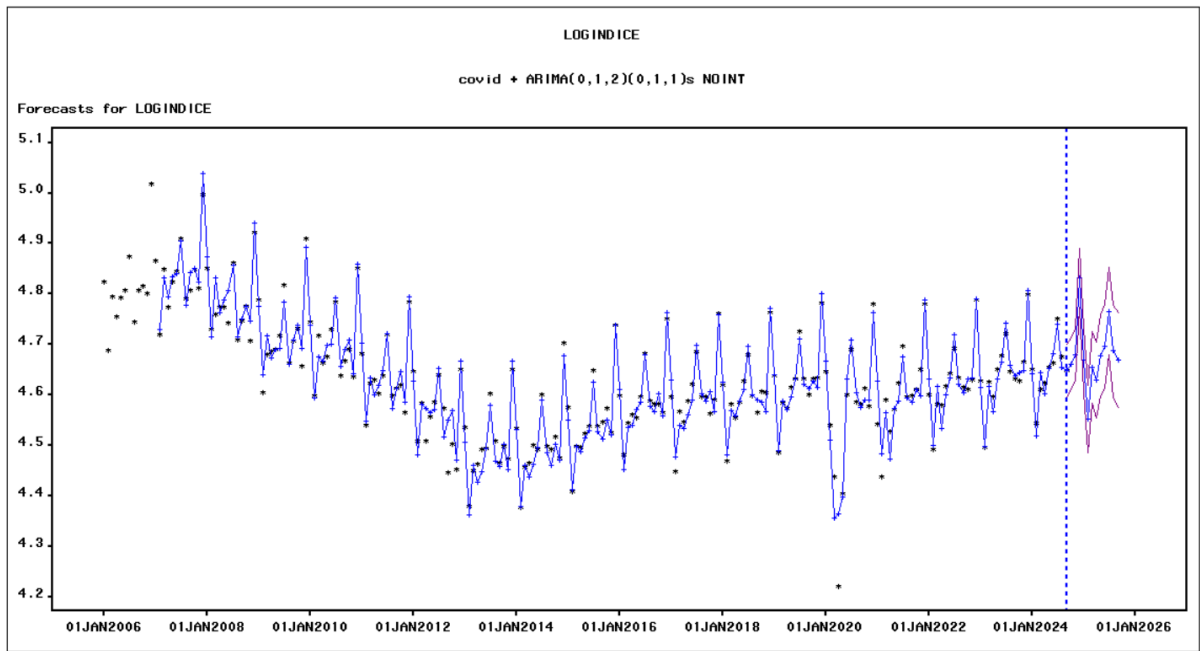


Figura 29: Enter Caption

FECHA	Indice	Predicción	Error
2024M10	107,85	105,8067536	1,897255
2024M11	108,212	107,808621	0,372767
2024M12	125,823	125,4579326	0,290144
2025M1	106,982	106,7134894	0,250987
2025M2	94,735	94,86243731	0,13452
2025M3	104,376	105,0485754	0,644377
2025M4	105,535	102,5077201	2,868508
2025M5	110,327	107,6242159	2,449794
2025M6	112,652	109,5471811	2,756115
2025M7	120,682	117,4908638	2,644252
2025M8	110,784	108,5513019	2,015362
2025M9	110,916	106,6164496	3,876402

Figura 30: Tabla de Predicciones

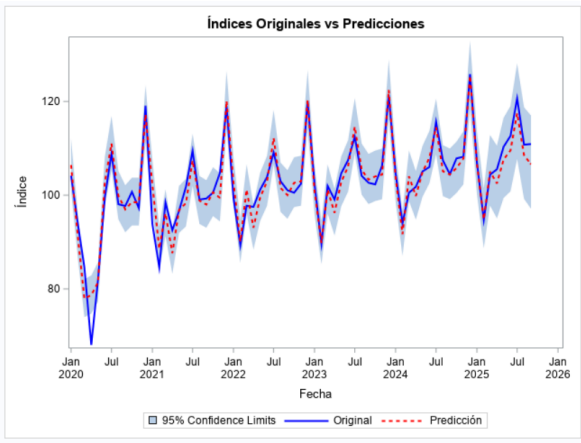


Figura 31: Gráfico predicciones

A partir de la tabla de predicciones (Figura 30), se confirma lo observado en el análisis exploratorio: el mes de diciembre presenta sistemáticamente los valores más altos del año, mientras que febrero corresponde al punto mínimo anual. Este patrón coincide exactamente con la estacionalidad identificada en fases previas del estudio, lo que indica que el modelo ha captado adecuadamente la estructura estacional.

En la misma tabla se aprecia además un comportamiento esperado en términos de precisión: los errores asociados a las predicciones aumentan conforme crece el horizonte temporal. La primera estimación presenta un valor cercano al 1,9 %, y los siguientes valores mantienen errores inferiores al 0,7 %. Sin embargo, las predicciones más alejadas en el tiempo muestran errores próximos al 2 %, lo que es consistente con la teoría de series temporales: la incertidumbre crece a medida que se avanza en el horizonte de predicción. Por tanto, el modelo ofrece una capacidad predictiva aceptable y refleja de manera adecuada la dinámica de la serie temporal analizada.

Por otro lado, las bandas de confianza mostradas en la Figura 31 también evidencian este crecimiento de la incertidumbre. Su expansión progresiva indica que las predicciones son considerablemente más fiables en el corto plazo. También cabe destacar que los valores previstos mantienen niveles similares a los del año anterior.

Las predicciones se ajustan adecuadamente a los valores reales observados. Sin embargo, el modelo no es capaz de anticipar eventos extraordinarios o no recurrentes. Un ejemplo ilustrativo de ello es la pandemia del COVID-19: si el modelo hubiera sido estimado únicamente con datos hasta enero de 2020, las predicciones para marzo y los meses siguientes se habrían desviado significativamente de la realidad, dado que el modelo no puede prever cambios abruptos que no formen parte del patrón histórico.

5. CONCLUSIONES

El presente trabajo ha permitido desarrollar un modelo ARIMA con la inclusión de una variable exógena que representa el impacto de la pandemia del COVID-19: $ARIMA(0,1,2) \times (0,1,1)_s + \text{COVID}$. Tras un proceso iterativo de identificación, estimación y validación, se ha logrado un modelo que satisface los supuestos estadísticos esenciales: sus residuos se comportan como ruido blanco, confirmando que la estructura de dependencia temporal ha sido adecuadamente capturada. La incorporación de un regresor que aísla el impacto de marzo a mayo de 2020 fue esencial para corregir la autocorrelación residual en el rezago 9. Además, los resultados confirman la fuerte estacionalidad anual de la serie, con pico en diciembre debido a la Navidad, y mínimos en febrero, validando así la necesidad de aplicar una diferenciación estacional.

En cuanto a su capacidad predictiva, el modelo ofrece pronósticos verosímiles y coherentes con la dinámica histórica, aunque con incertidumbre creciente en horizontes temporales más largos. En definitiva, el modelo constituye una herramienta robusta para comprender la evolución estructural del índice de comercio minorista y realizar proyecciones a corto y medio plazo en condiciones normales, pero su utilidad operativa debe complementarse con el análisis de factores externos.