

Análisis de Series Temporales de Venta de Viviendas de Segunda Mano

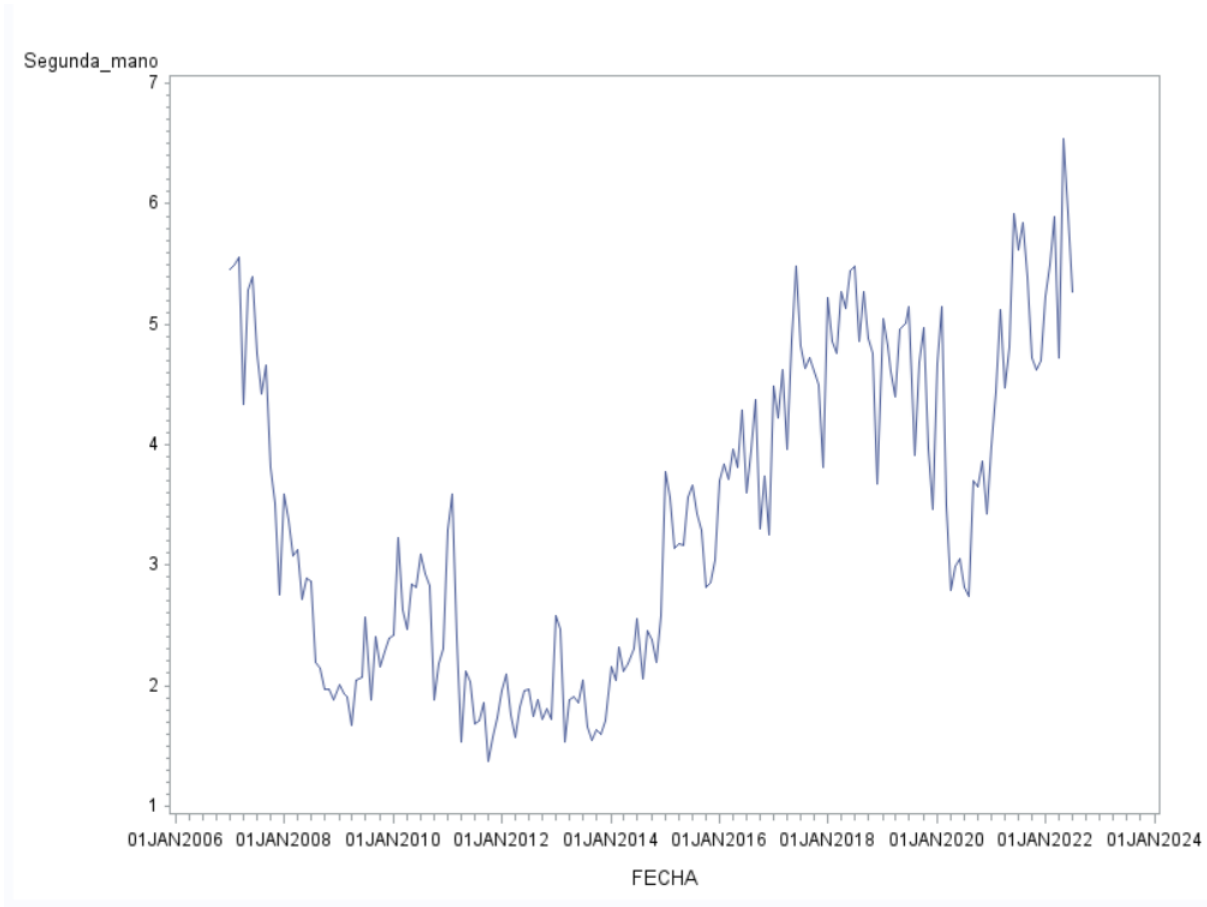
Introducción al Proyecto

Este proyecto se centra en el estudio de la venta de viviendas de segunda mano en España. Se utiliza una base de datos específica para este propósito, realizando un análisis exhaustivo de la serie temporal asociada.

Preparación y Exploración de Datos

Inicialmente, los datos se preparan adecuadamente, creando una nueva variable de fecha y seleccionando las variables relevantes para el análisis. Se realiza una revisión preliminar de los datos mediante la impresión de las primeras observaciones, proporcionando una visión inicial de la estructura de los datos.

Obs	Segunda_mano	FECHA
1	5.464	01JAN2007
2	5.498	01FEB2007
3	5.552	01MAR2007
4	4.340	01APR2007
5	5.278	01MAY2007
6	5.394	01JUN2007
7	4.754	01JUL2007
8	4.424	01AUG2007
9	4.659	01SEP2007
10	3.813	01OCT2007



- **Análisis Descriptivo**

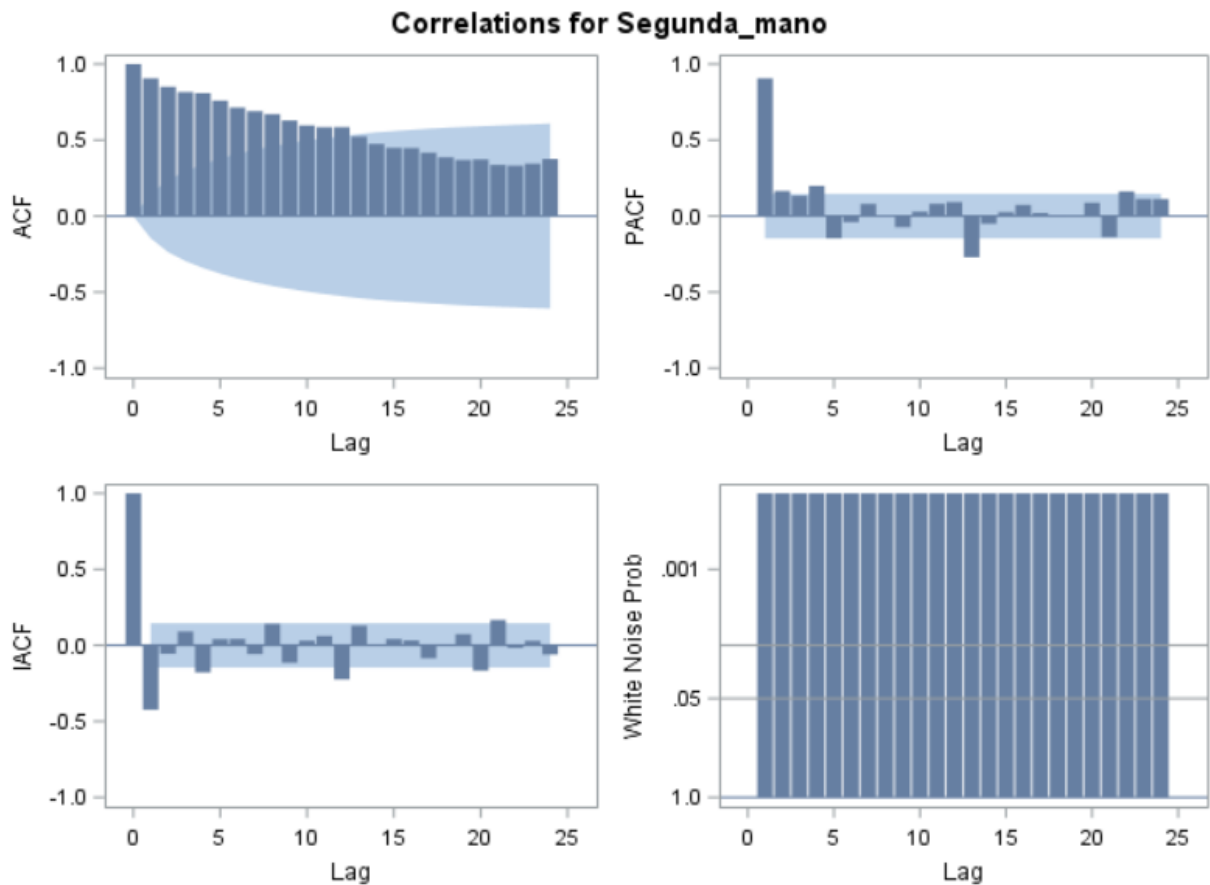
El análisis descriptivo de la variable “segunda mano” revela estadísticas clave que proporcionan una comprensión fundamental del comportamiento de las ventas de viviendas de segunda mano. Se observa que, a lo largo de 187 observaciones, el promedio de ventas es de aproximadamente 3.4, con una desviación estándar de 1.3. Esto indica una variabilidad moderada en los datos. El valor mínimo registrado es de 1.376, mientras que el máximo alcanza 6.541. Estos extremos destacan la variabilidad en las ventas de viviendas durante el período de estudio.

Analysis Variable : Segunda_mano				
N	Mean	Std Dev	Minimum	Maximum
187	3.4026578	1.3081317	1.3760000	6.5410000

- **Autocorrelaciones**

Al analizar las correlaciones parciales (PACF) y las autocorrelaciones (ACF) de la serie temporal, se identifican patrones significativos. En PACF, los retardos 1, 2, 3 y 13 muestran significancia, lo cual es indicativo de influencias de corto plazo y una posible estacionalidad anual en los datos. Específicamente, el retardo significativo en 13 sugiere un patrón recurrente que ocurre aproximadamente cada año.

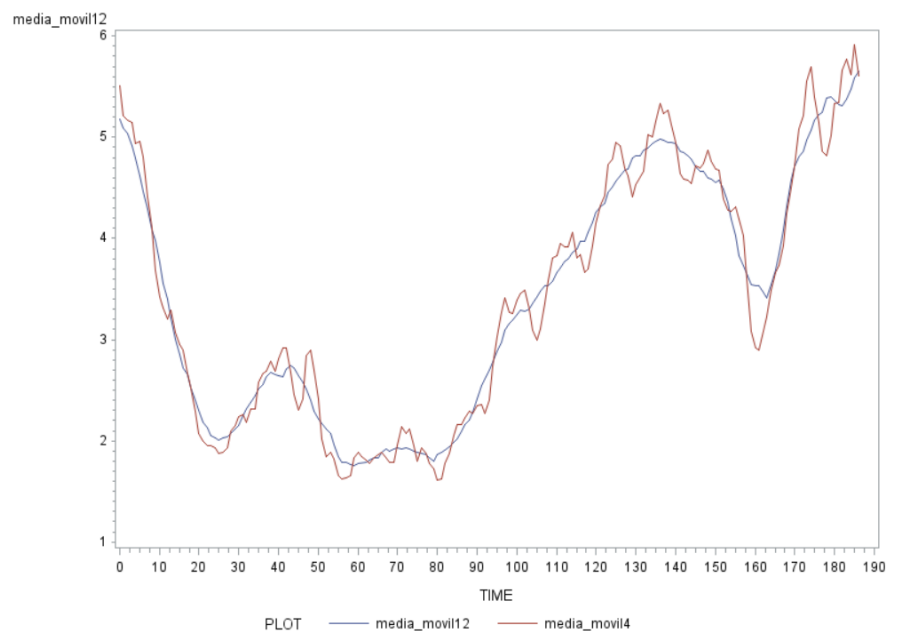
En cuanto a ACF, todos los retardos hasta el 13 son significativos, con una tendencia a disminuir gradualmente desde el primer retardo hasta el último. Esto implica una fuerte autocorrelación en el corto plazo que va disminuyendo lentamente, lo cual es típico en series temporales donde las observaciones cercanas en el tiempo tienden a ser más correlacionadas entre sí. La disminución gradual en la autocorrelación sugiere que, aunque las ventas de viviendas están influidas por sus valores pasados, este efecto se reduce con el tiempo.



- Medias Móviles

Media Móvil de 12 Meses (Anual): Esta media móvil es especialmente útil para identificar cambios significativos en los datos a lo largo del tiempo, como un aumento o disminución sostenida en las ventas de viviendas.

Media Móvil de 4 Meses (Trimestral): La media móvil trimestral proporciona una visión más detallada de las fluctuaciones a corto plazo y puede ayudar a identificar patrones estacionales o cíclicos que ocurren a lo largo del año.



En el análisis de la serie temporal de la venta de viviendas, se han identificado dos descensos significativos en las ventas. El primero ocurrió durante la crisis financiera de 2008, y el segundo coincidió con la pandemia de COVID-19, donde se observaron reducciones drásticas en las ventas de viviendas. Sin embargo, aparte de estos periodos, los datos revelan una tendencia creciente en las ventas de viviendas. Especialmente notoria es esta tendencia creciente si se consideran los datos desde 2008 hasta 2020 y, posteriormente, desde el bajón de 2020 hasta la actualidad.

En cuanto a las autocorrelaciones, se observa una alta correlación en los primeros retrasos que disminuye gradualmente para los retrasos más largos. Este patrón sugiere la posible existencia de estacionalidad en la serie, un aspecto que también se puede apreciar en el gráfico inicial. El análisis del gráfico de PACF revela que las autocorrelaciones parciales para los tres primeros retrasos son significativas, así como la del retraso 13. Esto podría indicar la presencia de un ciclo estacional, posiblemente anual.

Finalmente, se han creado dos filtros de medias móviles centradas para 12 meses (1 año) y 4 meses (1 trimestre) para poder suavizar la serie del ruido y observar más claramente sus características. Con el primer filtro podemos observar la tendencia que sospechamos que será un factor predominante y la estacionalidad de la serie.

Descomposición de la Serie Temporal

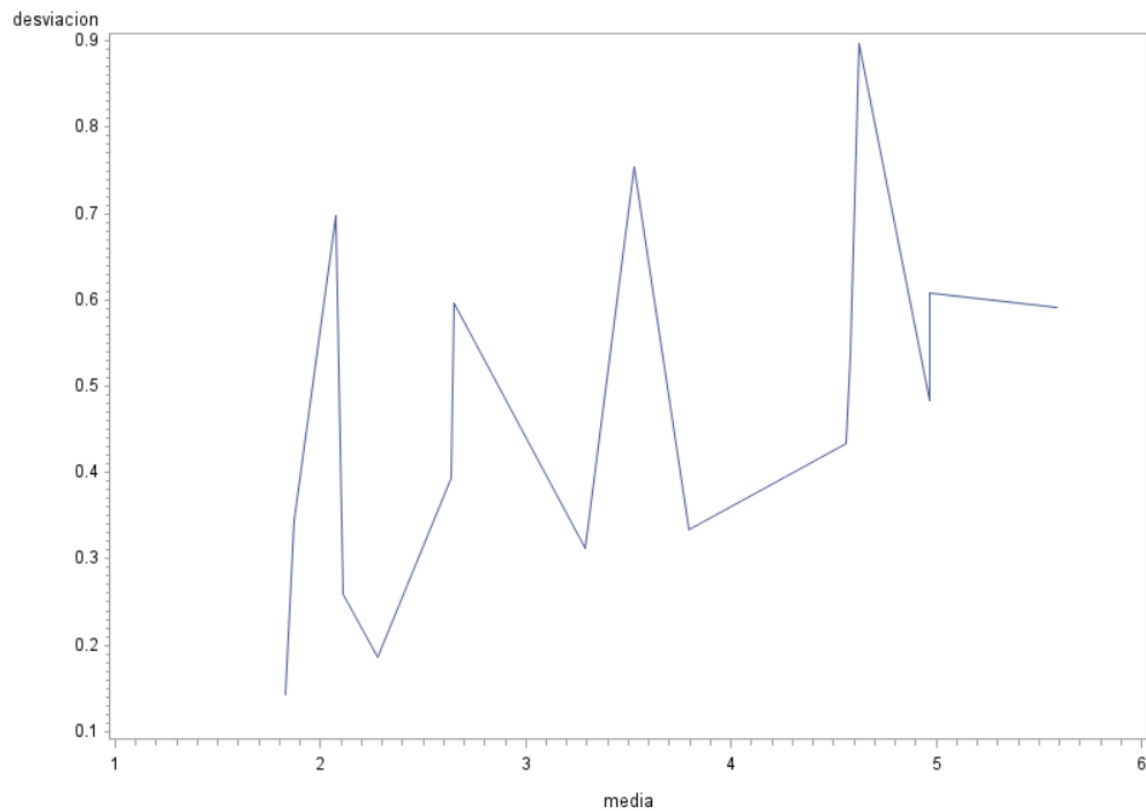
Al analizar gráficamente la serie temporal de las ventas de viviendas, resulta difícil determinar si las fluctuaciones en los datos se vuelven más pronunciadas con el tiempo. Para clarificar esto, se optó por usar un método matemático, el coeficiente de variación. Este análisis llevó a la decisión de aplicar una descomposición multiplicativa a la serie, un enfoque que sugiere que los cambios en las ventas no son constantes, sino que varían en función del nivel de ventas en cada momento.

Variable	Coeff of Variation
d	79.1842183
h	21.1542165

$h < d \rightarrow$ se opta por descomposición multiplicativa.

Al explorar la descomposición de la serie, se elige un enfoque multiplicativo basado en el gráfico de dispersión de los datos en relación con su media. Este enfoque se justifica observando el crecimiento en la relación entre la media y la desviación

estándar a lo largo del tiempo, lo que sugiere una tendencia creciente en la volatilidad.



- Descomposición con PROC TIMESERIES

Seasonal Decomposition for Variable Segunda_mano								
Obs	Time	Season	Trend-Cycle	Seasonal	Irregular	Seasonally Adjusted	Trend	Cycle
1	JAN2007	1	.	1.105916	.	4.940699	.	.
2	FEB2007	2	.	1.137606	.	4.832955	.	.
3	MAR2007	3	.	1.006052	.	5.518601	.	.
4	APR2007	4	.	0.940602	.	4.614067	.	.
5	MAY2007	5	.	1.011038	.	5.220378	.	.
6	JUN2007	6	.	1.068730	.	5.047114	.	.
7	JUL2007	7	4.542875	1.064234	0.983312	4.467062	4.224825	0.3180501
8	AUG2007	8	4.376292	0.954217	1.059404	4.636261	4.087074	0.2892177
9	SEP2007	9	4.184750	1.021927	1.089440	4.559036	3.949522	0.2352283
10	OCT2007	10	4.031292	0.908668	1.040919	4.196250	3.812548	0.2187438
11	NOV2007	11	3.874458	0.909744	0.998365	3.868123	3.676679	0.1977795
12	DEC2007	12	3.663417	0.871266	0.864087	3.165509	3.542578	0.1208386
13	JAN2008	1	3.480042	1.105916	0.932019	3.243464	3.411032	0.0690094
14	FEB2008	2	3.308167	1.137606	0.897329	2.968514	3.282904	0.0252628
15	MAR2008	3	3.110542	1.006052	0.982946	3.057496	3.159099	-0.0485569
16	APR2008	4	2.928875	0.940602	1.137244	3.330846	3.040538	-0.1116626
17	MAY2008	5	2.787167	1.011038	0.965602	2.691294	2.928112	-0.1409452
18	JUN2008	6	2.686125	1.068730	1.005315	2.700403	2.822643	-0.1365177
19	JUL2008	7	2.584042	1.064234	1.040354	2.688318	2.724863	-0.1408215
20	AUG2008	8	2.457958	0.954217	0.934585	2.297171	2.635421	-0.1774627
21	SEP2008	9	2.348708	1.021927	0.894923	2.101912	2.554876	-0.2061676

- Descomposición con PROC X11

Summary of Results and Combined Test for the Presence of Identifiable Seasonality	
Seasonality Tests:	Probability Level
Stable Seasonality F-test	0.000
Moving Seasonality F-test	0.289
Kruskal-Wallis Chi-square Test	0.000
Combined Measures:	Value
$T1 = 7/F_Stable$	0.52
$T2 = 3 \cdot F_Moving / F_Stable$	0.27
$T = (T1 + T2)/2$	0.39
Combined Test of Identifiable Seasonality:	Present

Obs	FECHA	ORIGINAL	C_ESTACIONAL	SA	C_TENDENCIA	C_IRREGULAR
1	01JAN2007	5.464	104.160	5.24578	5.32525	98.508
2	01FEB2007	5.498	106.639	5.15571	5.25837	98.048
3	01MAR2007	5.552	102.231	5.43084	5.18352	104.771
4	01APR2007	4.340	88.590	4.89896	5.08626	96.318
5	01MAY2007	5.278	103.257	5.11151	4.96579	102.934
6	01JUN2007	5.394	107.378	5.02338	4.83937	103.802
7	01JUL2007	4.754	113.257	4.19755	4.69374	89.429
8	01AUG2007	4.424	94.182	4.69729	4.52007	103.921
9	01SEP2007	4.659	103.731	4.49141	4.31858	104.002
10	01OCT2007	3.813	91.744	4.15613	4.08507	101.740
11	01NOV2007	3.519	92.241	3.81502	3.83201	99.557
12	01DEC2007	2.758	91.452	3.01580	3.58219	84.189
13	01JAN2008	3.587	104.634	3.42815	3.34821	102.388
14	01FEB2008	3.377	108.293	3.11840	3.15449	98.856
15	01MAR2008	3.076	102.055	3.01407	3.01442	99.989
16	01APR2008	3.133	88.973	3.52131	2.88987	121.850
17	01MAY2008	2.721	103.112	2.63888	2.75734	95.704
18	01JUN2008	2.886	106.761	2.70324	2.61365	103.428
19	01JUL2008	2.861	112.645	2.53983	2.47195	102.746

La implementación de la descomposición X11, y en particular su variante multiplicativa, refuerza la idea de que la serie temporal de las ventas de viviendas se

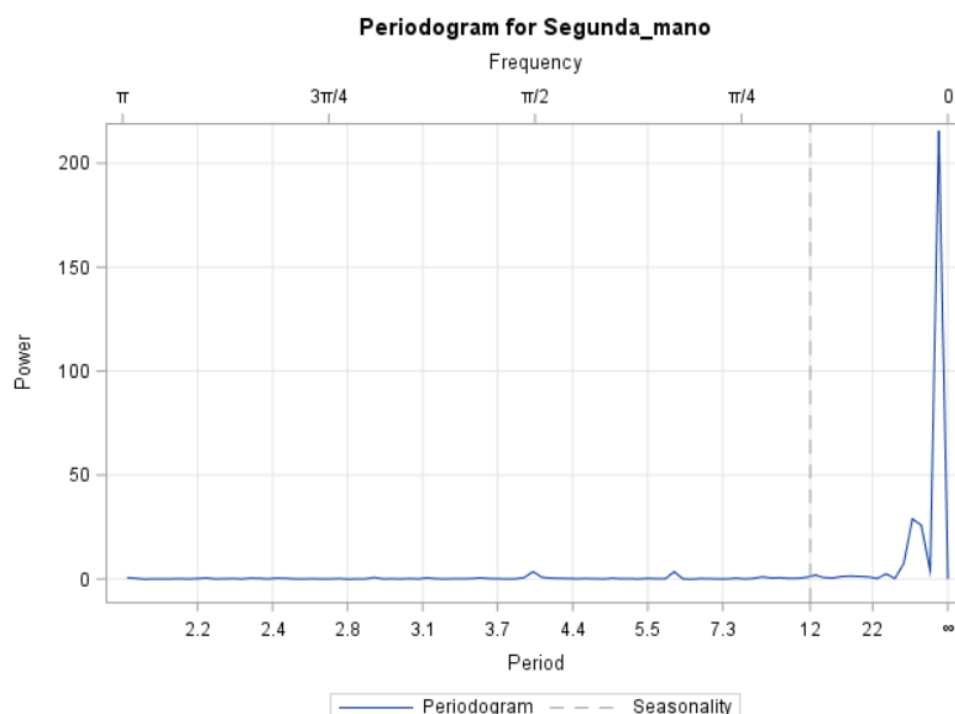
ajusta mejor a un modelo multiplicativo. La descomposición X11 es particularmente efectiva en series temporales donde la intervención o influencia humana es un factor significativo, como es el caso en los mercados inmobiliarios.

A partir de los resultados estadísticos obtenidos, se confirma que la serie temporal de las ventas de viviendas presenta una estacionalidad móvil. Esto significa que la estacionalidad no es constante a lo largo del tiempo, sino que varía, reflejando posiblemente cómo cambian los patrones de compra de viviendas en diferentes periodos del año y cómo estos patrones pueden evolucionar a lo largo del tiempo.

Un aspecto a destacar es que, en ninguna de las descomposiciones realizadas, se logró obtener ruido blanco en la componente irregular. La ausencia de ruido blanco indica que aún hay información o patrones en los residuos que el modelo no ha capturado.

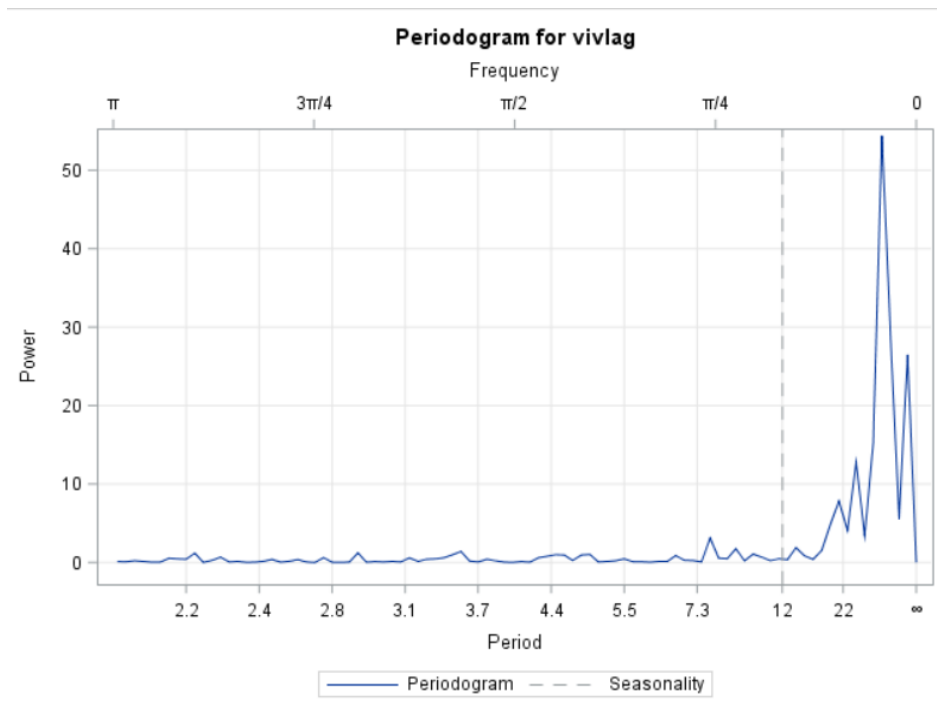
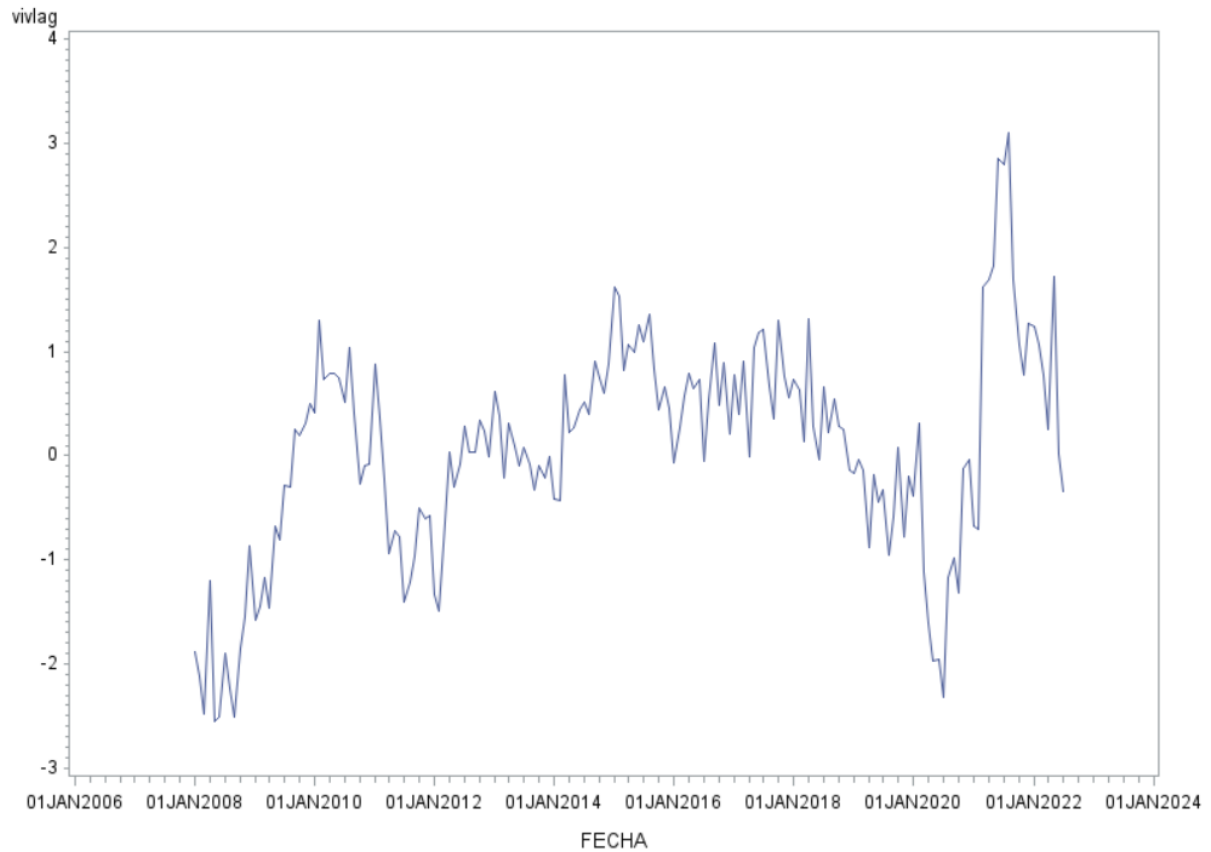
Dado este contexto, se está considerando si es más adecuado utilizar la serie ajustada sin la estacionalidad o trabajar directamente con la componente irregular para seguir modelando hasta que se consiga ruido blanco.

Análisis de Tendencias y Ciclos



En el análisis, se observa que la tendencia en las ventas de viviendas es tan fuerte que dificulta ver otros patrones como los ciclos. Para solucionar esto y ver mejor los

posibles ciclos, se decide realizar una diferenciación de los datos con el mismo mes del año anterior. Este método ayuda a reducir el efecto de la tendencia y hace más fáciles de ver los ciclos en los datos.



Una vez que se eliminó la tendencia predominante de la serie temporal, se hicieron evidentes ciclos de aproximadamente 44 meses, o alrededor de 4 años. Estos ciclos podrían estar influenciados por factores como ciclos económicos, variaciones en las tasas de interés, o incluso políticas gubernamentales que suelen cambiar o revisarse cada cuatro años.

Obs	FREQ	PERIOD	P_01	S_01
1	0.14362	43.750	45.2173	3.59828
2	0.17952	35.000	43.6006	3.46963
3	0.03590	175.000	24.8801	1.97990
4	0.21542	29.167	14.4533	1.15016
5	0.10771	58.333	9.9134	0.78888
6	0.25133	25.000	6.3874	0.50829
7	0.28723	21.875	3.4534	0.27481
8	0.07181	87.500	3.2576	0.25923
9	0.43085	14.583	2.1016	0.16724
10	0.46675	13.462	2.0927	0.16653

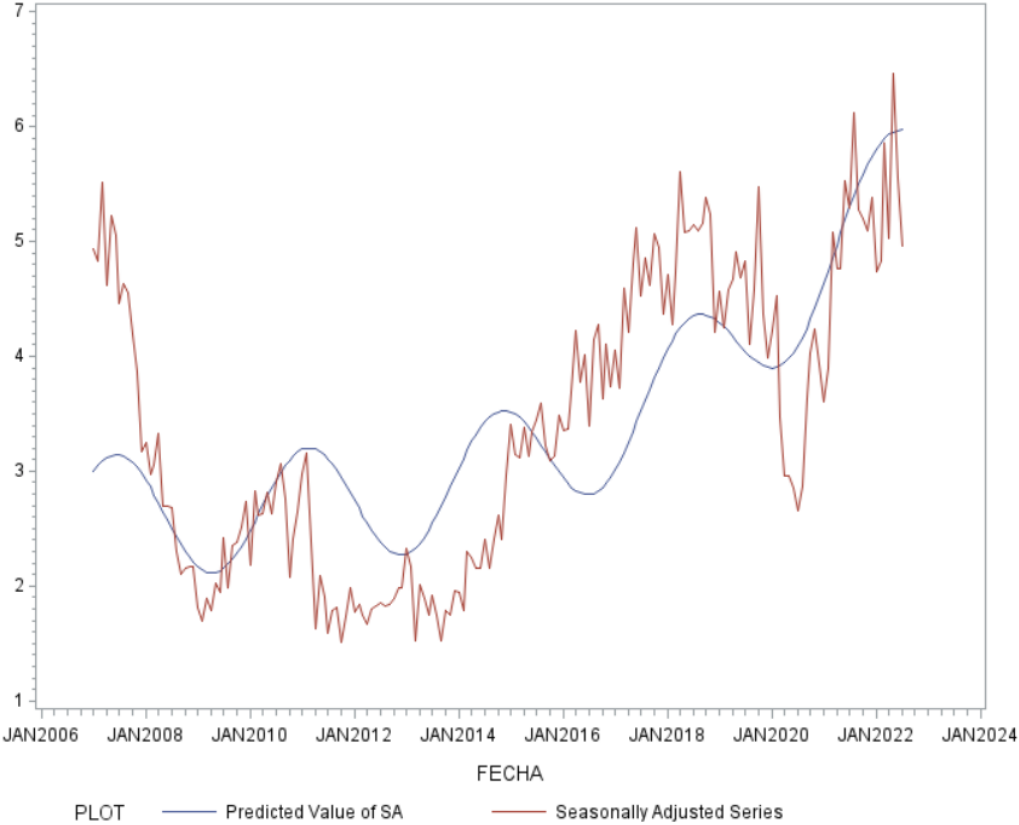
Análisis Detrended

El análisis detrended se centra en estudiar una serie temporal después de haber eliminado su estacionalidad principal. Este enfoque permite examinar las características de la serie que no están relacionadas con la tendencia a largo plazo, como la estacionalidad o patrones cíclicos más sutiles.

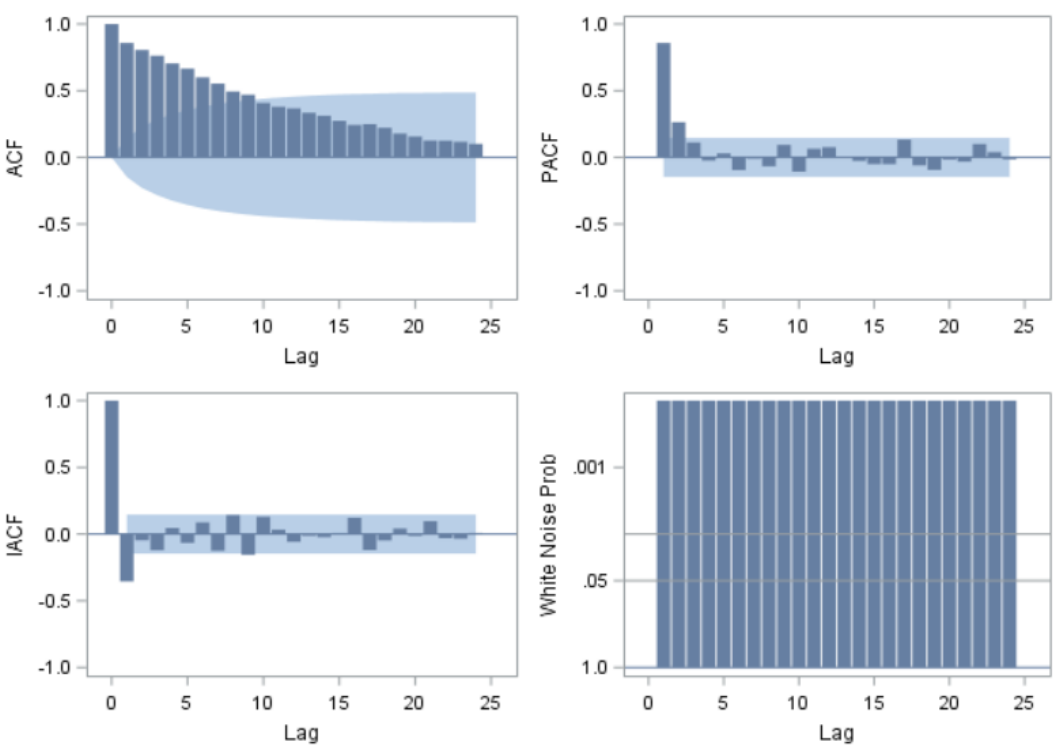
Para modelar esta serie inicialmente se llevó a cabo una regresión (con selección de variables stepwise) con la variable dependiente el componente "seasonally adjusted" (ajustada estacionalmente) de la descomposición multiplicativa, que esencialmente es la serie temporal original con los efectos estacionales quitados y variables explicativas el término cúbico t^{**3} y las funciones trigonométricas seno y coseno. Este enfoque permitió modelar patrones cíclicos y tendencias no lineales en los datos.

Summary of Stepwise Selection									
Step	Variable Entered	Variable Removed	Label	Number Vars In	Partial R-Square	Model R-Square	C(p)	F Value	Pr > F
1	t3			1	0.4681	0.4681	35.6183	162.79	<.0001
2	seno			2	0.0547	0.5228	15.1285	21.10	<.0001
3	coseno			3	0.0319	0.5547	4.0000	13.13	0.0004

Predicted Value of SA



Correlations for residuo



Posteriormente, los residuos del modelo de regresión fueron analizados y modelados utilizando un proceso ARIMA. En este caso, se ajustó un modelo ARIMA específico, expresado matemáticamente como:

$$(1 - 0.93991B)Y_t = (1 - 0.31938B)\varepsilon_t$$

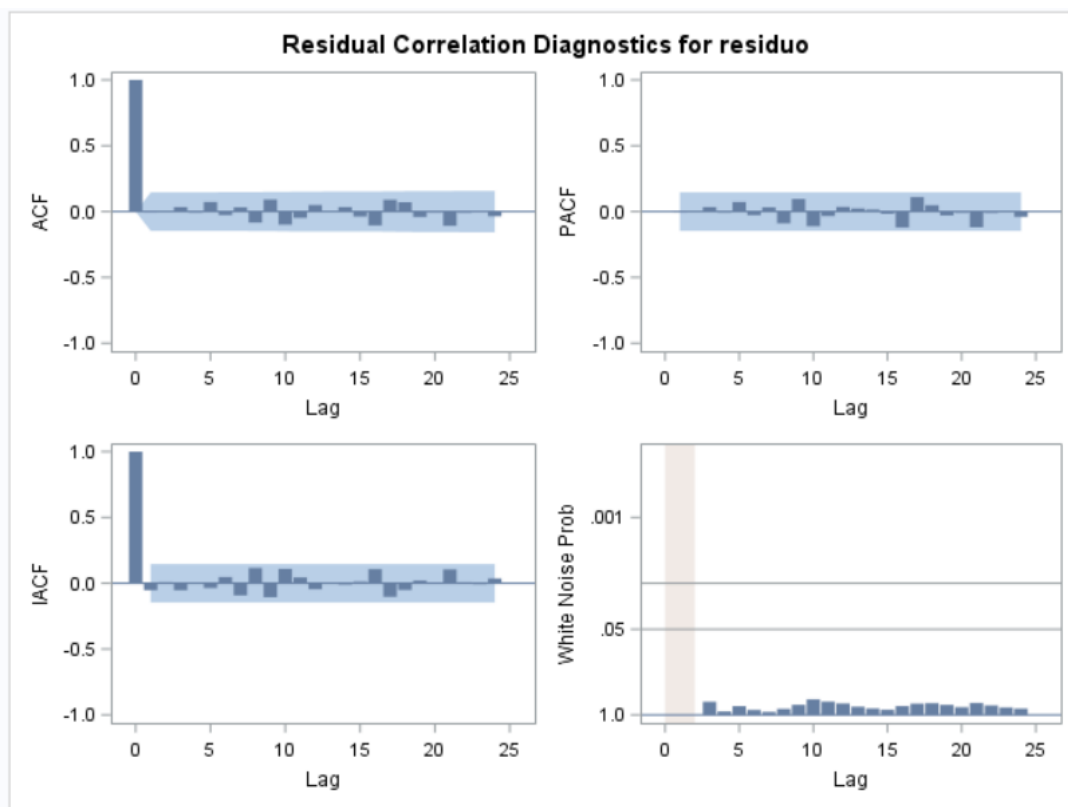
Este modelo representa una combinación de un término autorregresivo de primer orden (AR(1)) y un término de media móvil de primer orden (MA(1)).

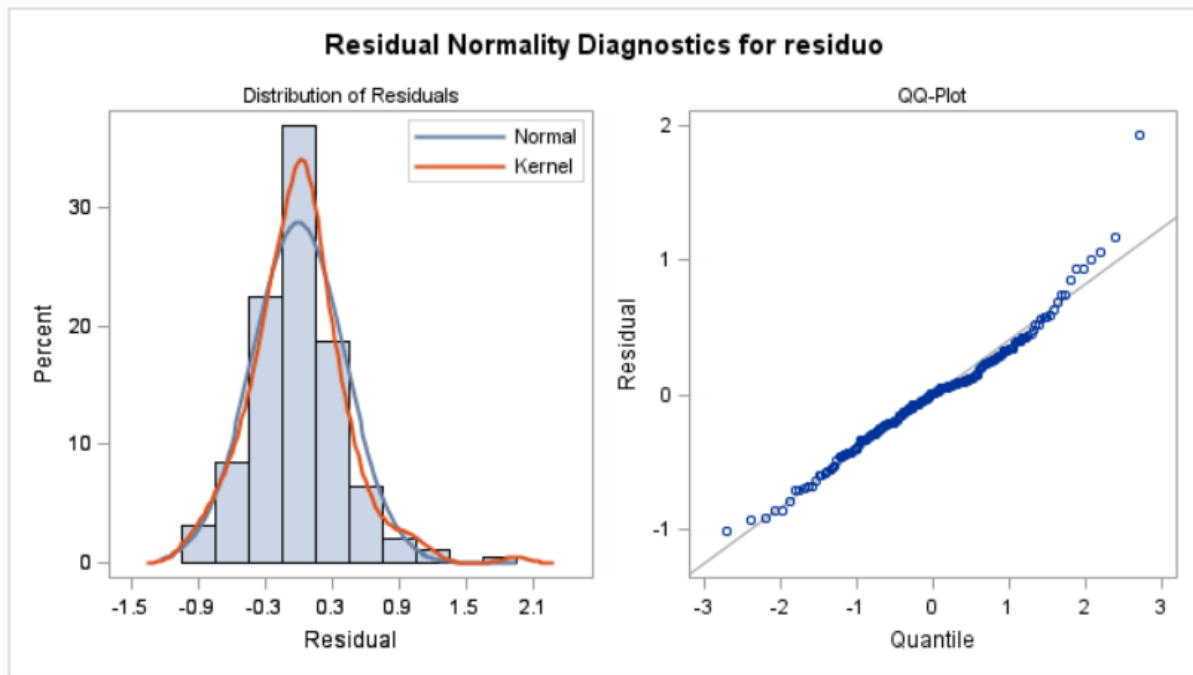
Conditional Least Squares Estimation					
Parameter	Estimate	Standard Error	t Value	Approx Pr > t	Lag
MA1,1	0.31938	0.07950	4.02	<.0001	1
AR1,1	0.93991	0.02891	32.51	<.0001	1

Variance Estimate	0.17368
Std Error Estimate	0.416749
AIC	205.3207
SBC	211.7829
Number of Residuals	187

Correlations of Parameter Estimates		
Parameter	MA1,1	AR1,1
MA1,1	1.000	0.463
AR1,1	0.463	1.000

Autocorrelation Check of Residuals									
To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations					
6	1.43	4	0.8383	-0.005	0.004	0.035	-0.007	0.074	-0.026
12	7.58	10	0.6693	0.032	-0.083	0.092	-0.099	-0.047	0.050
18	13.14	16	0.6626	-0.001	0.034	-0.037	-0.105	0.090	0.071
24	16.25	22	0.8033	-0.041	-0.007	-0.108	-0.008	0.003	-0.035
30	21.30	28	0.8125	0.030	0.072	-0.048	0.067	0.095	-0.029
36	27.92	34	0.7593	0.039	0.016	-0.071	0.023	0.097	-0.108





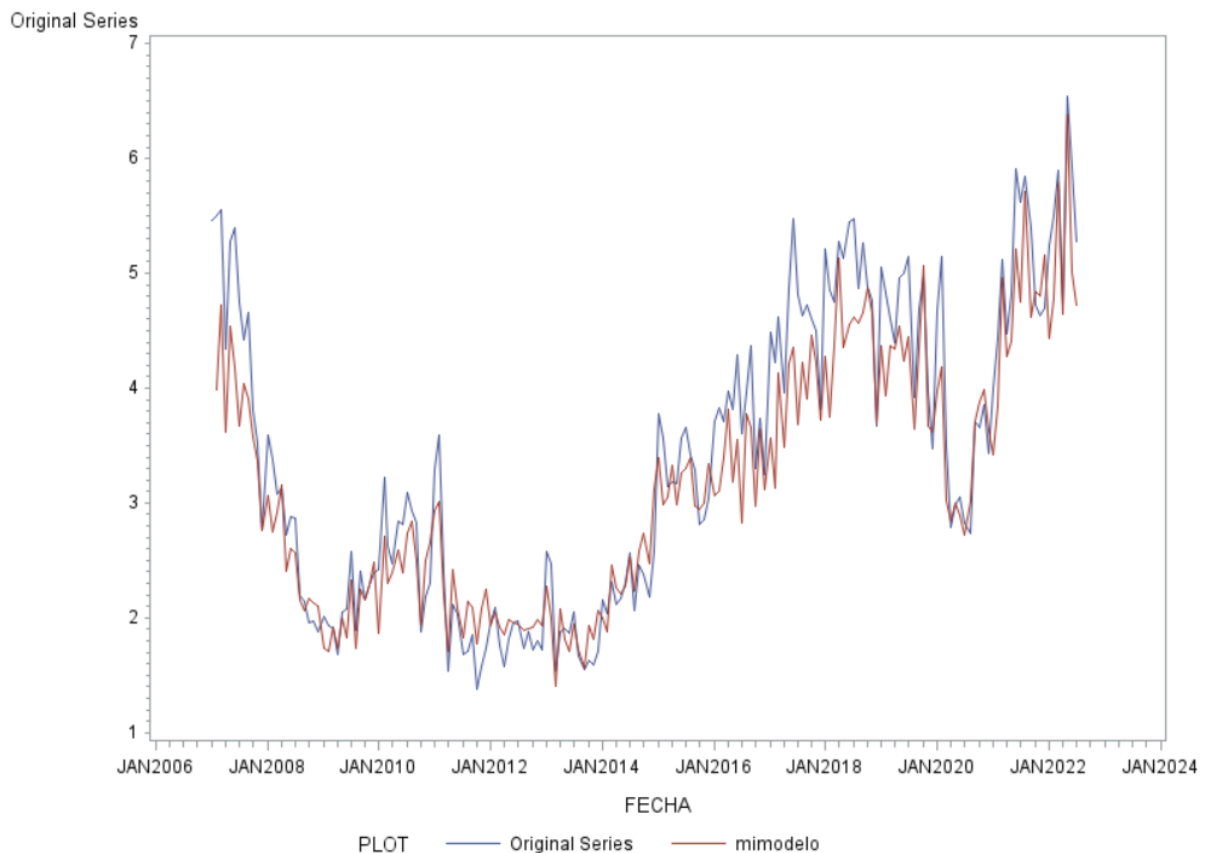
El modelo final que combina ambas etapas de modelización se expresa como:

$$\text{mimodelo} = 0.93991 \times \text{lag}(\text{previsto}) + \text{residuo} - 0.31938 \times \text{lag}(\text{residuo})$$

Aquí, "previsto" corresponde a las predicciones del modelo de regresión que incluye t^{**3} , seno y coseno. En el modelo ARIMA, el término $0.94 \times y(t-1)$ representa la parte autoregresiva aplicada directamente al valor anterior de la serie (en este caso a la variable "previsto") y luego incorpora un ajuste del residuo anterior.

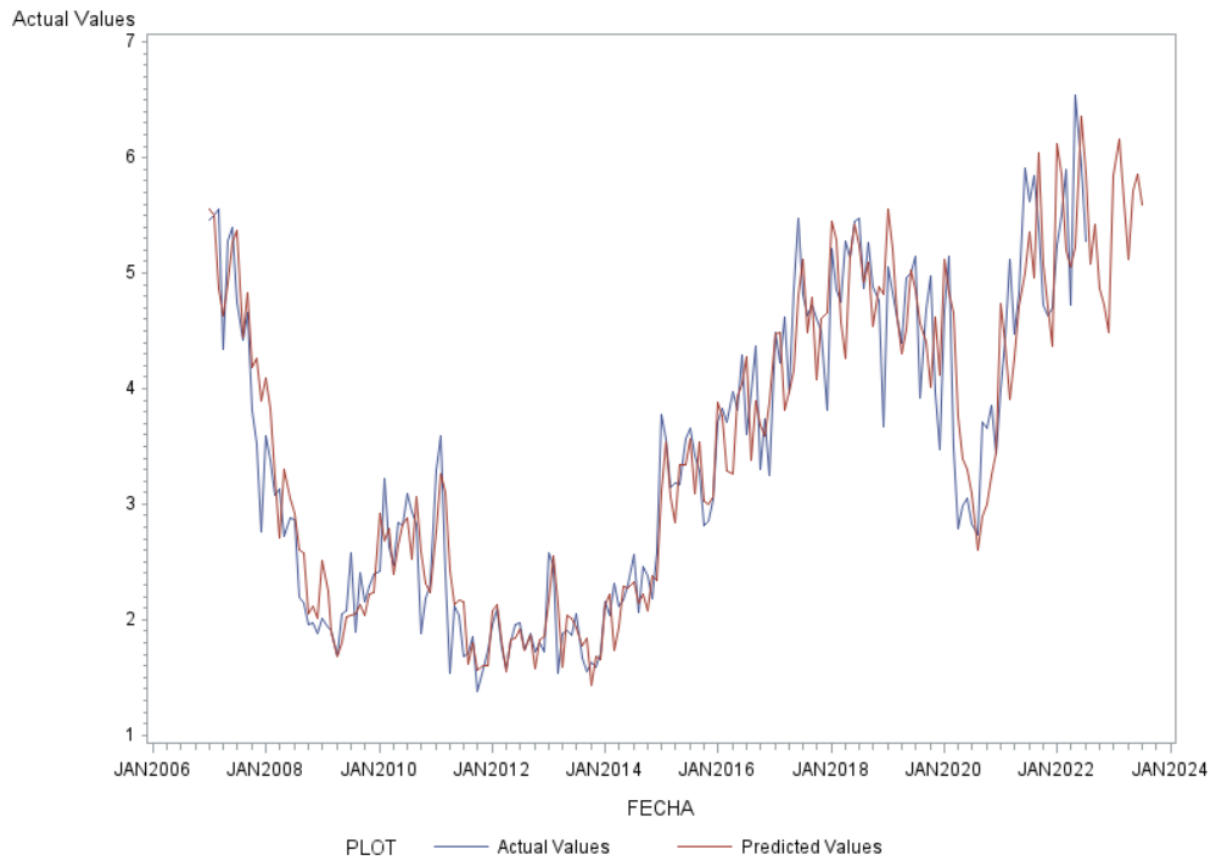
Autoregressive Factors	
Factor 1:	$1 - 0.93991 B^{**}(1)$
Moving Average Factors	
Factor 1:	$1 - 0.31938 B^{**}(1)$

Con este modelo se consigue ruido blanco y como se puede observar en los gráficos, un razonable ajuste a la realidad presentada por los datos.



Suavizado Exponencial

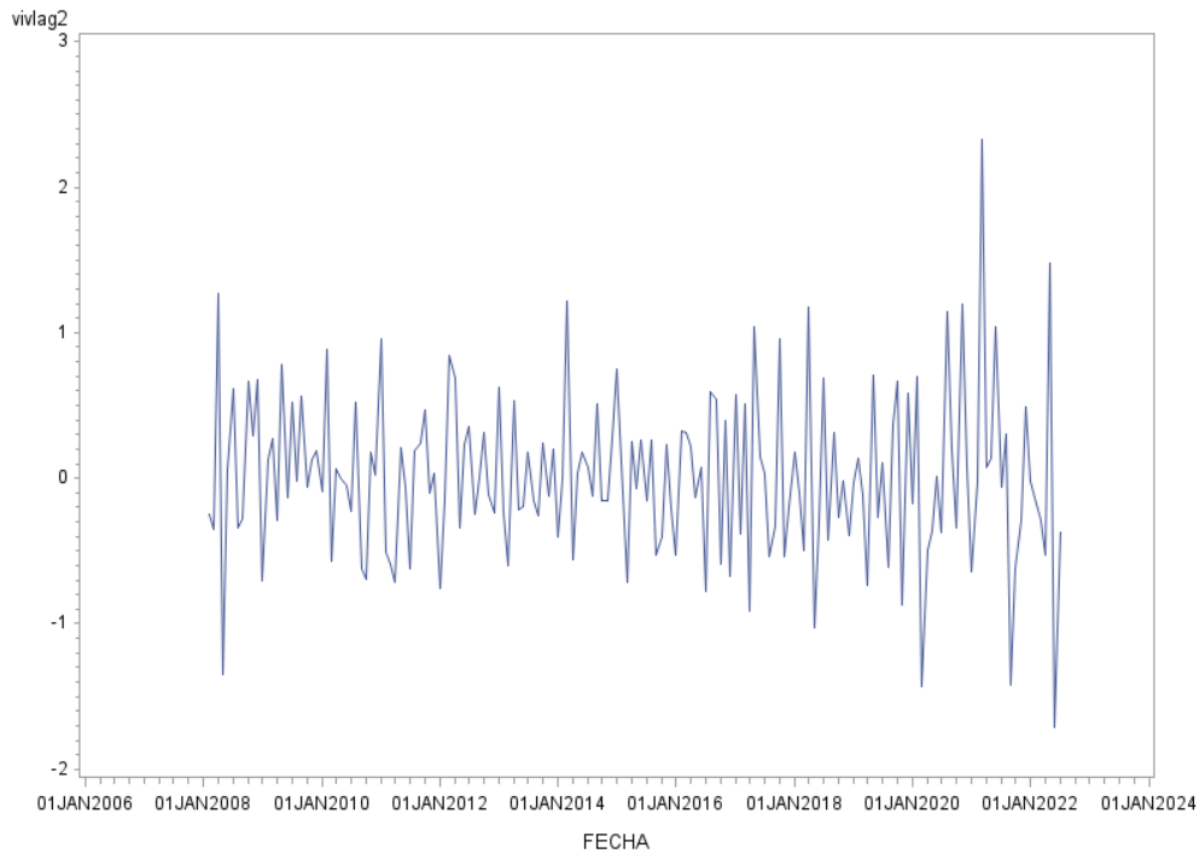
Para el análisis de las ventas de viviendas, se va a emplear un suavizado exponencial, específicamente el método de Holt-Winters (suavizado exponencial triple). Este método es ideal para series temporales con tendencia y estacionalidad, ya que ajusta y pronostica los datos considerando estos tres componentes clave. Dado que se ha identificado tanto tendencia como estacionalidad en la serie temporal, el suavizado exponencial triple de Holt-Winters, en su variante multiplicativa, es la técnica seleccionada para este análisis. Este enfoque es particularmente efectivo cuando la estacionalidad de la serie cambia proporcionalmente a su nivel, como se observa en los datos.



Análisis ARIMA

En el proceso de análisis ARIMA de la serie temporal de las ventas de viviendas, se siguieron varios pasos esenciales para garantizar un modelo preciso y confiable.

Inicialmente, se enfrentó el desafío de que la variable `segunda_mano` no era estacionaria, un requisito crucial para la aplicación de ARIMA. Por ello, se optó por trabajar con `vivlag2`, una versión de la serie original que había sido diferenciada dos veces, primero para eliminar la estacionalidad y luego para deshacerse de la tendencia. Esta serie modificada mostraba signos de estacionariedad, lo que la hacía ideal para el análisis.



El siguiente paso fue verificar la estacionariedad y la constancia de la varianza en vivlag2. Tras confirmar que la serie era estacionaria y tenía una varianza constante, se procedió con la construcción del modelo ARIMA.

Augmented Dickey-Fuller Unit Root Tests							
Type	Lags	Rho	Pr < Rho	Tau	Pr < Tau	F	Pr > F
Zero Mean	0	-221.6520	<.0001	-17.4965	<.0001		
Single Mean	0	-221.7309	<.0001	-17.4538	<.0001	152.3174	<.0010
Trend	0	-222.0691	<.0001	-17.4483	<.0001	152.2331	<.0010

KPSS Stationarity Test			
Type	Lags	Eta	Pr > Eta
Single Mean	4	0.0901	0.6361
Trend	4	0.0343	0.8091

Analysis of Variance					
Source	DF	Sum of Squares	Mean Square	F Value	Pr > F
Model	1	0.01747	0.01747	0.16	0.7084
Error	5	0.55706	0.11141		
Corrected Total	6	0.57453			

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > t
Intercept	1	-0.71920	0.51511	-1.40	0.2215
logm_viv	1	-0.06825	0.17234	-0.40	0.7084

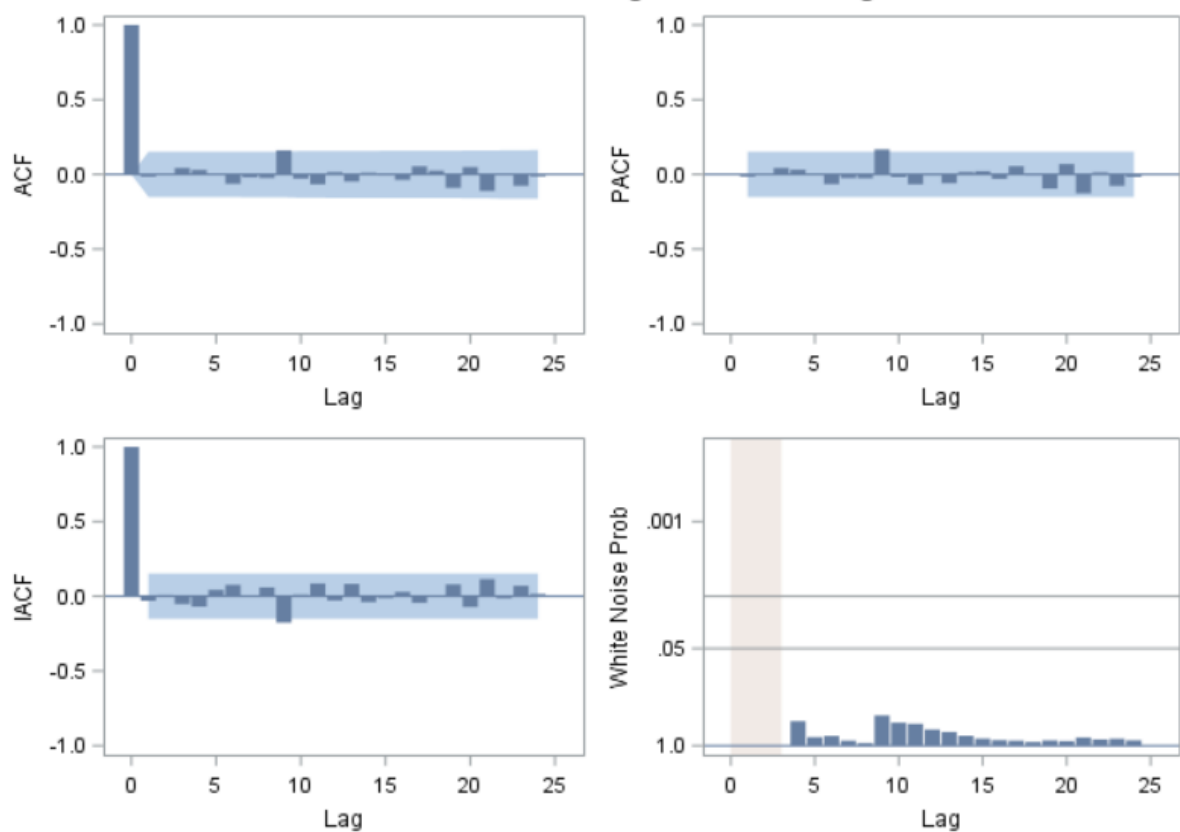
Este modelo ARIMA específico fue elegido con dos retrasos autorregresivos ($p=2$) y un término de media móvil para el retardo 12 ($q=12$). Esta configuración del modelo ARIMA se seleccionó con el objetivo de capturar tanto las dependencias a corto plazo (a través de los retrasos autorregresivos) como la estacionalidad anual (mediante el término de media móvil en el retardo 12).

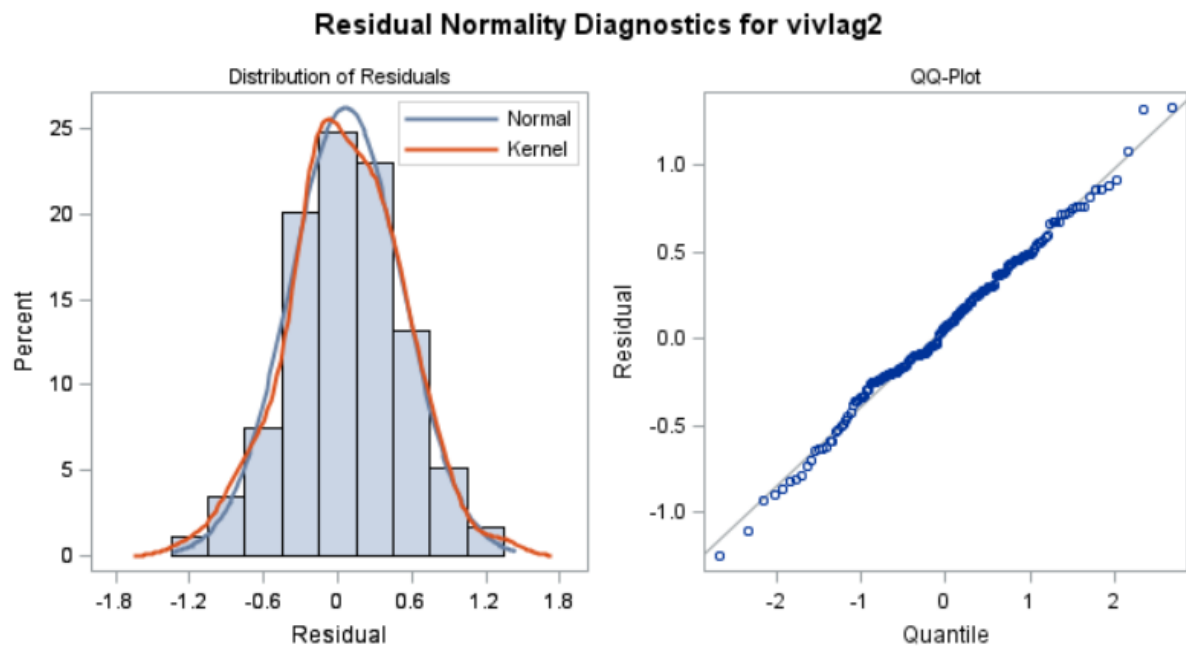
Conditional Least Squares Estimation					
Parameter	Estimate	Standard Error	t Value	Approx Pr > t	Lag
MA1,1	0.69226	0.05874	11.78	<.0001	12
AR1,1	-0.38405	0.07590	-5.06	<.0001	1
AR1,2	-0.19233	0.07690	-2.50	0.0133	2

Variance Estimate	0.215194
Std Error Estimate	0.46389
AIC	229.4626
SBC	238.9398
Number of Residuals	174

Autocorrelation Check of Residuals									
To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations					
6	1.66	3	0.6462	0.007	0.023	0.061	0.051	0.024	-0.041
12	8.25	9	0.5088	0.002	-0.002	0.180	-0.005	-0.043	0.036
18	9.90	15	0.8260	-0.025	0.031	0.013	-0.017	0.070	0.039
24	14.49	21	0.8476	-0.072	0.062	-0.096	0.006	-0.067	-0.005
30	17.96	27	0.9047	0.060	0.035	0.016	0.061	0.065	-0.060

Residual Correlation Diagnostics for vivlag2





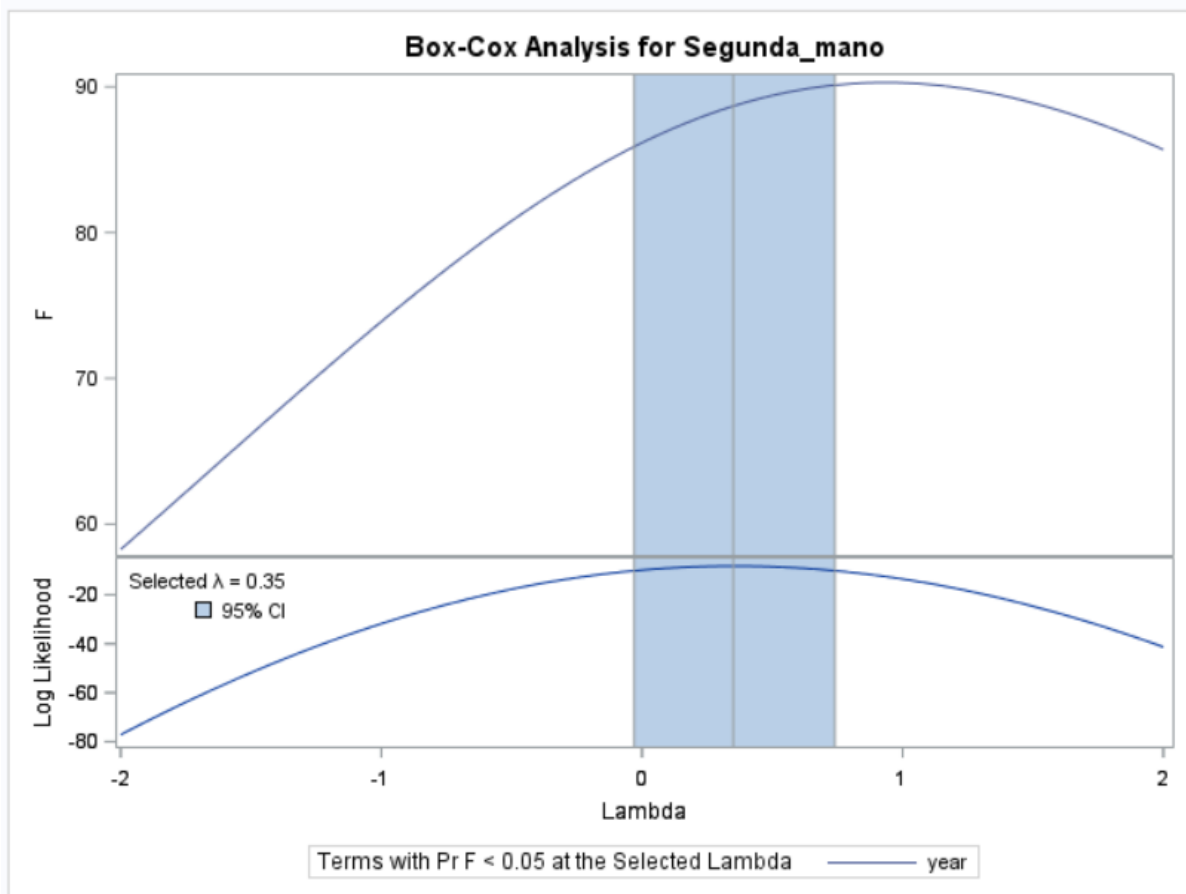
Paralelamente se decidió trabajar con la variable “cruda” segunda_mano, que como ya hemos visto no era estacionaria. Para abordar esto, se probó una serie de transformaciones para conseguir estacionariedad y homogeneidad de varianzas. Primero, se intentó una transformación Box-Cox, utilizando un valor lambda de 0.25, obtenido de una regresión logarítmica, dónde

$\log_{std_viv} = a + b \cdot \log_{m_viv}$ y $\lambda = 1 - \text{pendiente de la regresión}$.

Augmented Dickey-Fuller Unit Root Tests							
Type	Lags	Rho	Pr < Rho	Tau	Pr < Tau	F	Pr > F
Zero Mean	0	-2.0522	0.3240	-1.0326	0.2710		
Single Mean	0	-15.6332	0.0311	-2.8622	0.0523	4.0963	0.0825
Trend	0	-26.8462	0.0135	-4.1843	0.0057	9.2416	<.0010

KPSS Stationarity Test			
Type	Lags	Eta	Pr > Eta
Single Mean	3	2.2758	<.0001
Trend	3	0.5718	0.0001

Parameter Estimates					
Variable	DF	Parameter Estimate	Standard Error	t Value	Pr > t
Intercept	1	-1.73754	0.35041	-4.96	0.0002
logm_viv	1	0.75128	0.28455	2.64	0.0194



Dado este resultado, se intentó aparte una transformación logarítmica ya que una regla general que se puede aplicar es que siempre cuando la lambda es entre -0.3 y 0.3 la mejor transformación es la logarítmica.

Luego se probó una transformación Box-Cox con lambda de 0.35, obtenido a través del procedimiento PROC TRANSREG. Sin embargo, ninguna de estas transformaciones lograron hacer la serie estacionaria.

Como resultado, se decidió aplicar otra transformación: la diferenciación logarítmica, representada por

$\text{viv_loglag} = \text{viv_log} - \text{lag}(\text{viv_log})$

Esta transformación consiste en tomar el logaritmo de segunda_mano y luego restar el valor logarítmico de cada punto con su valor anterior (primera diferenciación).

Esta técnica resultó ser más efectiva para alcanzar la estacionariedad necesaria en la serie.

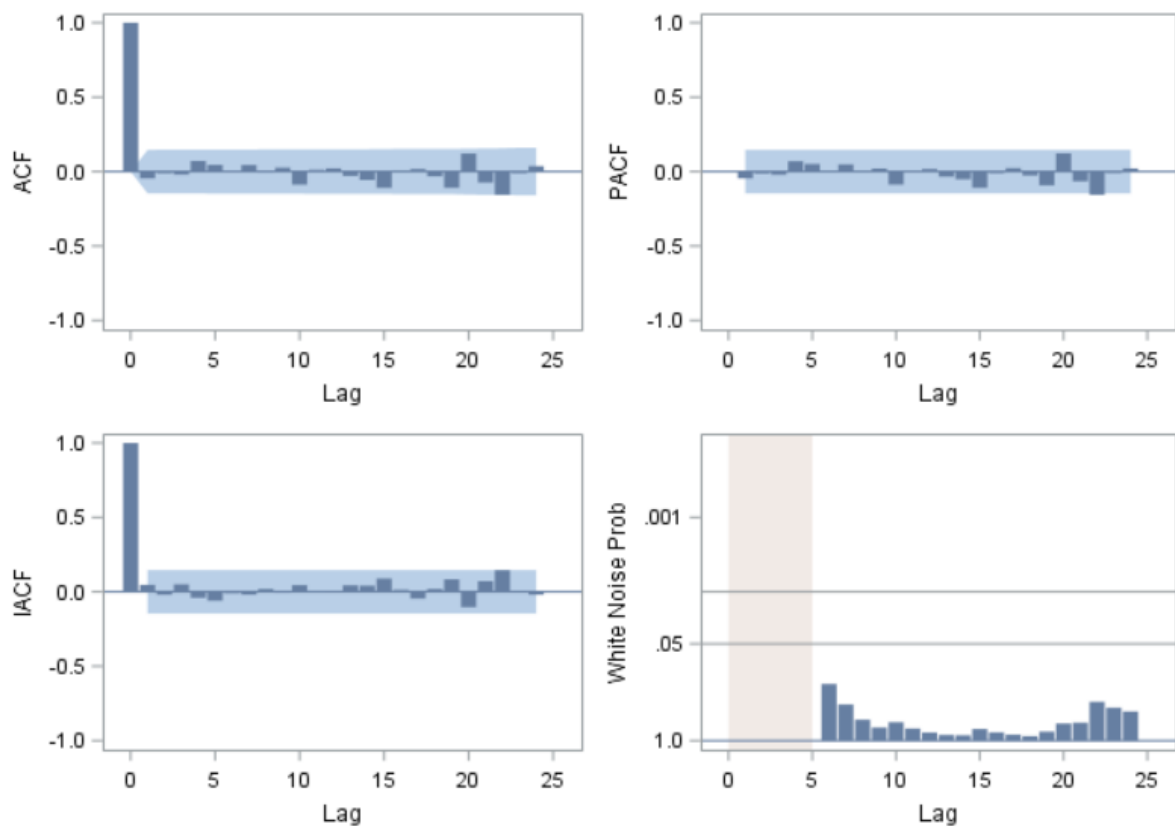
Para el análisis ARIMA, se seleccionó un modelo más refinado que incluye términos autorregresivos en los retardos 1, 2 y 21, y términos de media móvil en los retardos 12 y 24, sin incluir un término constante. Esta configuración específica se eligió para capturar de manera más efectiva tanto las dependencias a corto como a largo plazo, así como la estacionalidad compleja en los datos de ventas de viviendas.

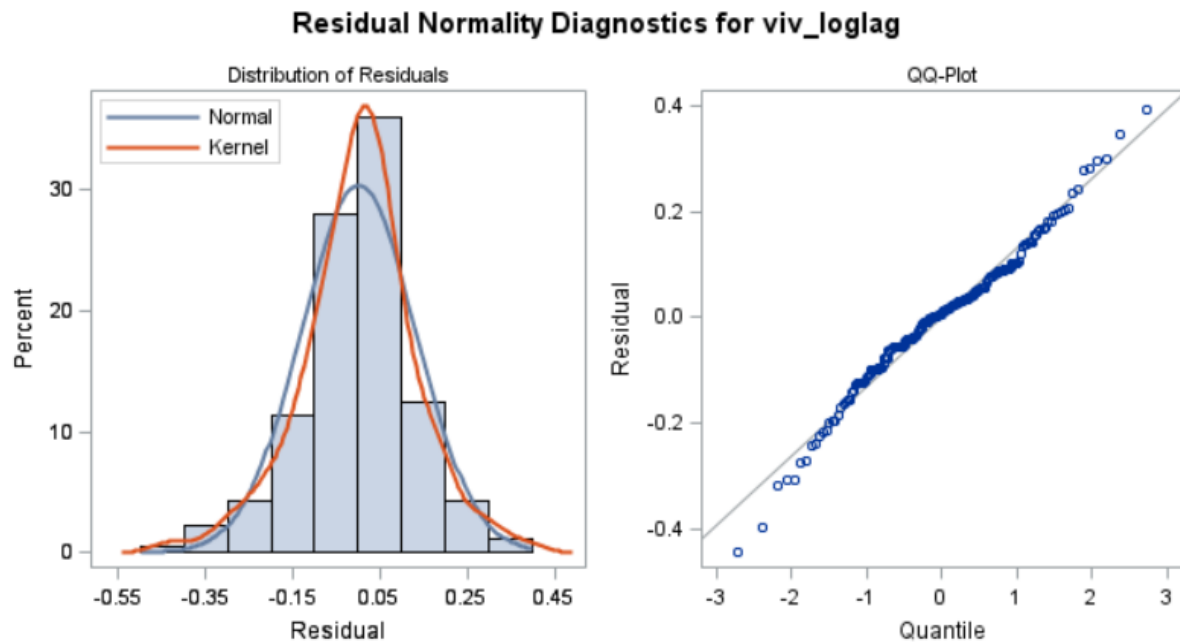
Conditional Least Squares Estimation					
Parameter	Estimate	Standard Error	t Value	Approx Pr > t	Lag
MA1,1	-0.30864	0.07380	-4.18	<.0001	12
MA1,2	-0.31961	0.07818	-4.09	<.0001	24
AR1,1	-0.26589	0.07309	-3.64	0.0004	1
AR1,2	-0.21024	0.07221	-2.91	0.0040	2
AR1,3	-0.21103	0.07511	-2.81	0.0055	21

Variance Estimate	0.017667
Std Error Estimate	0.132917
AIC	-217.93
SBC	-201.801
Number of Residuals	186

Autocorrelation Check of Residuals									
To Lag	Chi-Square	DF	Pr > ChiSq	Autocorrelations					
6	1.85	1	0.1733	-0.045	-0.011	-0.020	0.071	0.045	-0.002
12	4.01	7	0.7785	0.044	-0.006	0.026	-0.087	0.012	0.022
18	7.55	13	0.8714	-0.029	-0.056	-0.109	0.008	0.017	-0.032
24	19.78	19	0.4080	-0.109	0.122	-0.074	-0.155	-0.010	0.037
30	24.16	25	0.5104	0.002	-0.009	-0.096	0.089	0.022	-0.046
36	32.00	31	0.4168	-0.003	0.073	-0.082	0.012	0.042	0.141

Residual Correlation Diagnostics for viv_loglag





También se tomaron en cuenta las proposiciones del método automático de selección de valores para los parámetros para un modelo ARIMA (minic).

ARMA(p+d,q) Tentative Order Selection Tests		
SCAN		
p+d	q	BIC
6	5	-1.08608
0	12	-1.35834
12	1	-1.33614

Examen de Outliers

Se aplicó una búsqueda específica de outliers con el criterio de seleccionar los seis más significativos, basándose en un umbral de significancia de 0.001. Esta búsqueda reveló la existencia de dos outliers prominentes en las observaciones 171 y 151.

Outlier Details					
Obs	Time ID	Type	Estimate	Chi-Square	Approx Prob>ChiSq
171	01-MAR-2021	Additive	1.05544	13.44	0.0002
159	01-MAR-2020	Additive	-0.89044	11.16	0.0008

Para abordar estos datos atípicos, se crearon variables ficticias específicas para cada outlier identificado. Estas variables toman el valor de 1 en los instantes temporales correspondientes a los outliers y 0 en el resto, permitiendo así modelar el impacto específico de estos puntos atípicos en la serie.

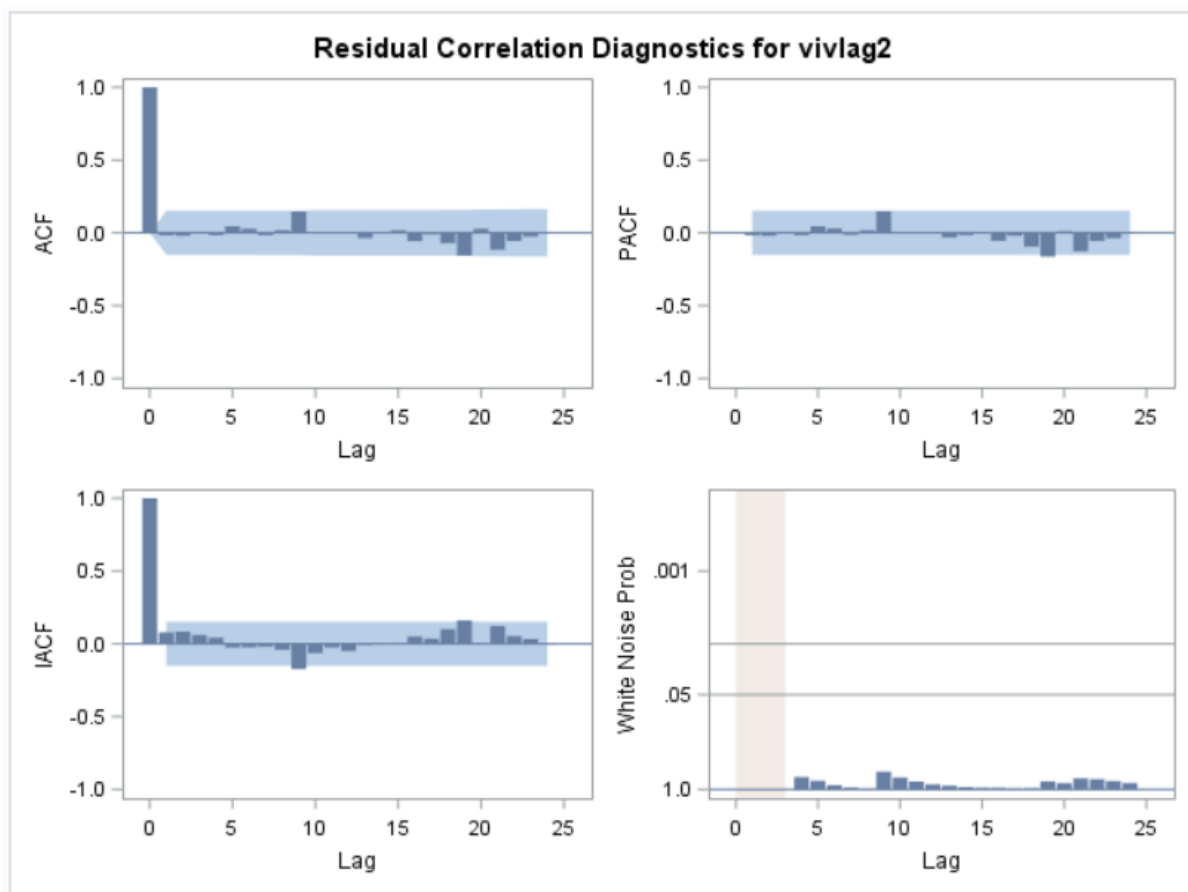
Posteriormente, se ajustó nuevamente el modelo ARIMA incluyendo estas variables ficticias como parte de la entrada. Esta técnica, conocida como incorporación de variables de intervención, permite al modelo contabilizar y ajustar su comportamiento en presencia de estos outliers. De este modo, se mejora la precisión del modelo y se reduce la distorsión que estos puntos atípicos podrían causar.

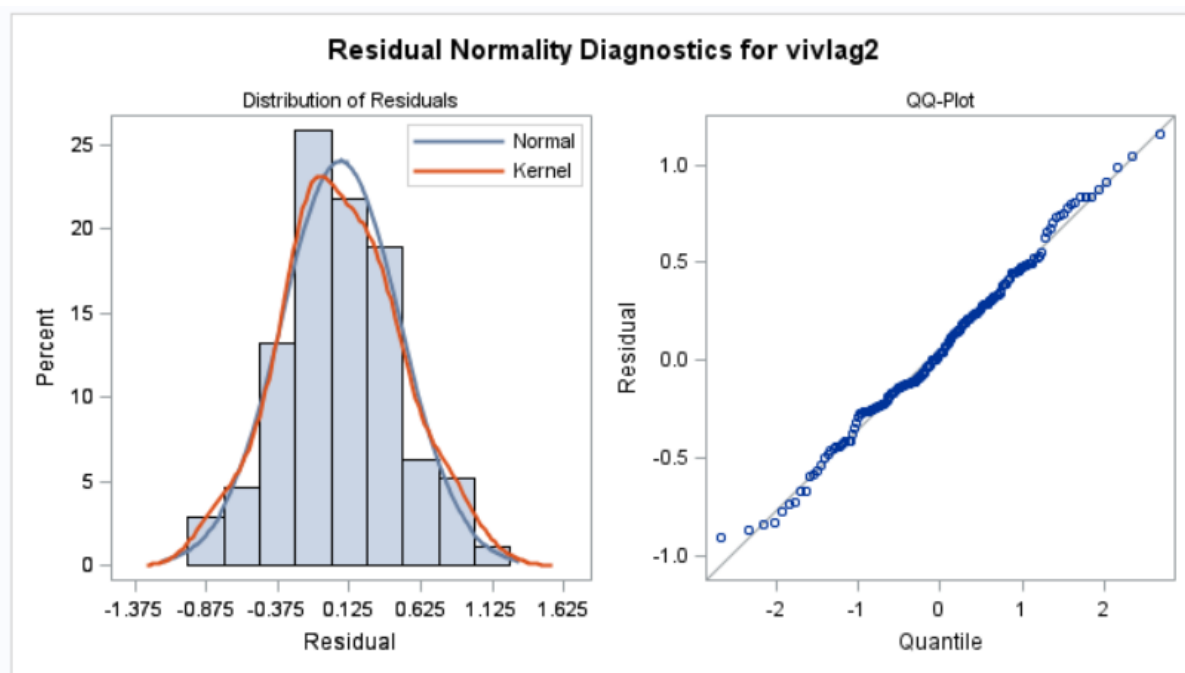
Una revisión adicional reveló otro outlier en la observación 177, que también se trató mediante una variable ficticia. Tras este ajuste, se comprobó que no quedaban más outliers significativos en la serie. Al ser todos los parámetros significativos y los residuos, ruido blanco, no hizo falta revisar el modelo ARIMA.

Una observación curiosa es que los dos primeros outliers (159 y 171) producidos tienen cierta correlación. Se puede notar que el primero ocurre al inicio de la pandemia de COVID-19 en marzo de 2020 y el otro, al inicio de la estabilización de la crisis sanitaria tras un año, cuando se empezó a recuperar el mercado inmobiliario, resultando en un aumento notable en las ventas de viviendas, en comparación con el año pasado.

Variance Estimate	0.181084
Std Error Estimate	0.425539
AIC	202.3542
SBC	221.3085
Number of Residuals	174

Conditional Least Squares Estimation							
Parameter	Estimate	Standard Error	t Value	Approx Pr > t	Lag	Variable	Shift
MA1,1	0.59141	0.06816	8.68	<.0001	12	vivlag2	0
AR1,1	-0.44640	0.07627	-5.85	<.0001	1	vivlag2	0
AR1,2	-0.25866	0.07760	-3.33	0.0011	2	vivlag2	0
NUM1	-1.55568	0.38056	-4.09	<.0001	0	I_159	0
NUM2	2.10873	0.39738	5.31	<.0001	0	I_171	0
NUM3	-1.10716	0.38684	-2.86	0.0047	0	I_177	0





Determinación de Variables Escalón

En el trabajo se realizó un análisis de escalón utilizando el procedimiento PROC AUTOREG en SAS, aplicando el test de Bai y Perron para detectar múltiples cambios estructurales, específicamente enfocado en los residuos del modelo. Este método es crucial para evaluar la estabilidad de la serie temporal, lo cual es un indicador clave de la fiabilidad de las predicciones futuras.

Los resultados del test proporcionan estadísticas para diferentes números de cambios estructurales potenciales, que sugirieron que la inclusión de un solo cambio estructural era suficiente, ya que los p-valores altos para cambios estructurales adicionales indicaron la ausencia de evidencia estadística de más cambios en la estructura de los datos.

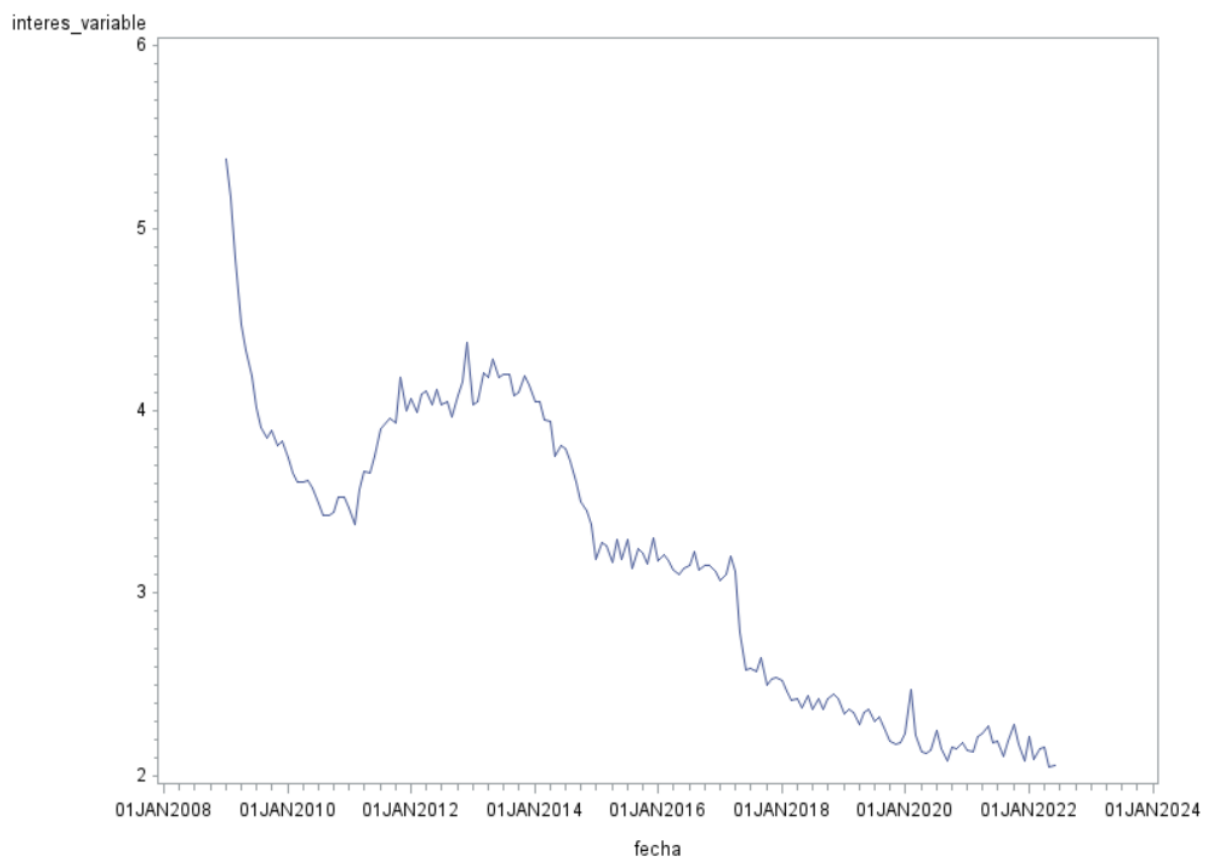
Bai and Perron's Multiple Structural Change Tests			
supF(l+1 l) Tests			
l	New Break	supF(l+1 l)	Pr > supF(l+1 l)
0	37	4.51191308	0.9434
1	46	3.10356084	0.9978
2	150	2.20032126	0.9998
3	37	3.16791974	0.9972
4	113	5.28642868	0.8597
5	42	1.91364135	0.9999

La recomendación de un único cambio estructural y la estabilidad subsiguiente de los residuos del modelo sugieren

que las predicciones futuras basadas en este modelo pueden considerarse confiables, siempre y cuando se haya tenido en cuenta adecuadamente el cambio estructural detectado.

Funciones de Transferencia y Análisis de Correlación Cruzada

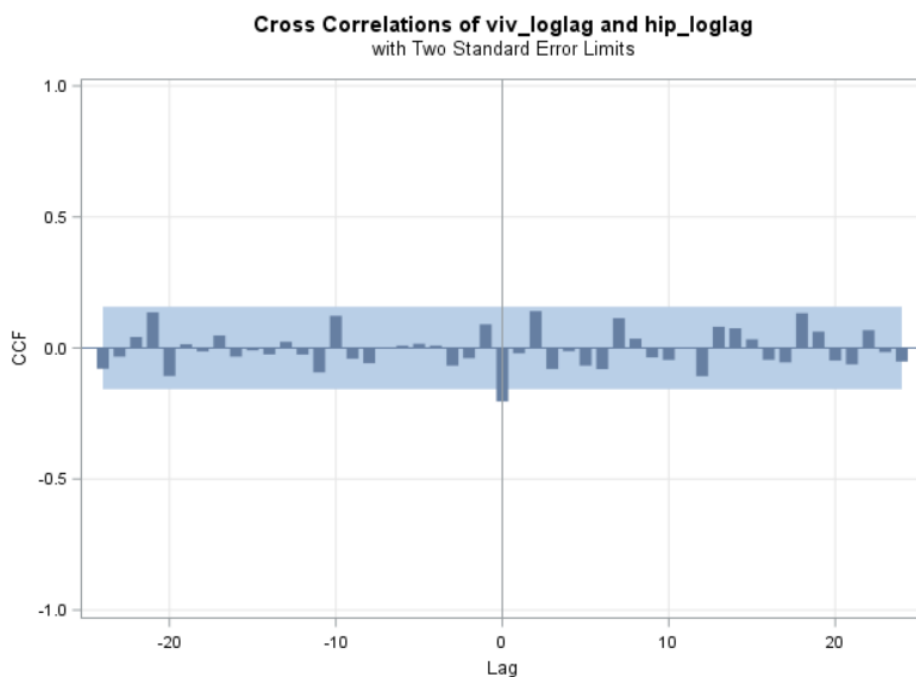
En el análisis realizado, se utilizó la base de datos "hipotecas" para examinar cómo la variable `interes_variable` podría influir en la variable de estudio principal. Mediante el uso de funciones de transferencia, se buscó comprender la relación dinámica entre estas dos variables.



Inicialmente, se aplicó una transformación consistente a `interes_variable`, específicamente la diferenciación del logaritmo de la serie. Esta técnica se empleó para estabilizar la serie y hacerla más adecuada para el análisis de series temporales, facilitando la identificación de relaciones estables a lo largo del tiempo.



Posteriormente, se llevó a cabo un análisis de correlación cruzada para explorar cómo las variaciones en las tasas de interés de las hipotecas se relacionan temporalmente con la variable principal. Los resultados revelaron que el único instante significativo en la correlación cruzada es el instante cero, sugiriendo que cualquier cambio en las tasas de interés tiene un efecto inmediato en la variable de interés.



Para incorporar estos hallazgos en el modelo, se ajustó un modelo ARIMA utilizando funciones de transferencia. El modelo ajustado fue el siguiente:

Autoregressive Factors	
Factor 1:	$1 + 0.31615 B^{**}(1) + 0.22118 B^{**}(2)$
Factor 2:	$1 - 0.40648 B^{**}(12)$
Factor 3:	$1 + 0.23779 B^{**}(21)$

Input Number 1	
Input Variable	hip_loglag
Overall Regression Factor	-0.62216

Este modelo comprende varios factores autorregresivos que capturan diferentes dinámicas en los datos:

- Factor AR de primer y segundo orden.
- Factor estacional de orden 12.
- Factor adicional de orden 21.

La inclusión de la variable transformada hip_loglag como un factor de entrada, con un coeficiente de regresión de -0.62216, que se interpreta de forma que por cada unidad de incremento en la tasa de interés de las hipotecas (tal como está transformada en el modelo como se ha comentado anteriormente), se espera, en promedio, una disminución de 0.62216 unidades en las ventas de casas de segunda mano.

Conditional Least Squares Estimation							
Parameter	Estimate	Standard Error	t Value	Approx Pr > t	Lag	Variable	Shift
AR1,1	-0.31615	0.07907	-4.00	<.0001	1	viv_loglag	0
AR1,2	-0.22118	0.08031	-2.75	0.0066	2	viv_loglag	0
AR2,1	0.40648	0.07642	5.32	<.0001	12	viv_loglag	0
AR3,1	-0.23779	0.08633	-2.75	0.0066	21	viv_loglag	0
NUM1	-0.62216	0.28786	-2.16	0.0322	0	hip_loglag	0

Variance Estimate	0.017784
Std Error Estimate	0.133355
AIC	-186.927
SBC	-171.52
Number of Residuals	161

Capacidad Predictiva



En este estudio, se realizó un análisis detallado de series temporales para evaluar la capacidad predictiva de un modelo ARIMA. Se comenzó dividiendo el conjunto de datos en dos subconjuntos: train y test, utilizando los primeros 175 registros para el entrenamiento y los últimos 12 para la prueba. Posteriormente, se aplicó el procedimiento ARIMA al conjunto de entrenamiento, eligiendo el modelo elegido para ajustar a la variable `vivlag2`, comentado anteriormente.

Luego, se realizaron predicciones para los siguientes 12 meses. Para alinear las fechas de las predicciones con las observaciones reales, se ajustaron las fechas en, estableciéndolas desde agosto de 2021 hasta julio de 2022. Finalmente, se fusionaron los datos de prueba con las predicciones ajustadas y se visualizaron los resultados mediante un gráfico, comparando las observaciones reales (del conjunto test) con las predicciones del modelo.

Conclusiones

El estudio ha realizado un análisis exhaustivo de las ventas de viviendas de segunda mano en el mercado español, centrándose en dos enfoques principales de series temporales: el método `detrended` y el modelo ARIMA. Se han aplicado dos transformaciones clave, `vivlag2` y `viv_loglag`, que han sido esenciales para capturar las tendencias, la estacionalidad y la esencia de los datos.

Trabajando con `vivlag2`, se construyó un modelo robusto, que mostró una mejora significativa en términos de eficiencia y precisión, como lo indican los valores de AIC y BIC (pasando de un AIC de 229.4626 y un BIC de 238.9398 a un AIC de 202.3542 y un SBC de 221.3085). Este avance fue especialmente notable después de analizar y corregir los outliers en los datos, lo que elevó la precisión y relevancia del modelo.

Además, la inclusión de una variable relacionada con las hipotecas ha proporcionado una perspectiva adicional, ofreciendo la posibilidad de comprender aún mejor la dinámica del mercado de viviendas. Este enfoque integral ha permitido no solo un análisis detallado de las ventas de viviendas de segunda mano, sino también la exploración de factores económicos externos que podrían influir en estas tendencias.