

FACULTAD DE ESTUDIOS ESTADÍSTICOS

**MÁSTER EN MINERÍA DE DATOS E INTELIGENCIA
DE NEGOCIOS**

Curso 2017/2018

Trabajo de Fin de Máster

***ANÁLISIS Y PREDICCIÓN DEL MERCADO
INMOBILIARIO EN LA COMUNIDAD DE MADRID***

Alumno: TERESA ÁLVAREZ MARTÍN

Tutor: JUANA MARÍA ALONSO REVENGA

Septiembre de 2018



UNIVERSIDAD COMPLUTENSE
MADRID

Índice

1. INTRODUCCIÓN	1
1.1. Justificación del proyecto	2
1.2. Objetivos del proyecto y materiales a emplear	4
2. MODELOS TEÓRICOS EMPLEADOS EN EL ANÁLISIS	5
2.1. Conocimientos previos	5
2.2. Modelos Univariantes	7
2.2.1. Modelos basados en métodos de suavizado o alisado	7
2.2.2. Modelos estacionarios	8
2.2.3. Modelos no estacionarios	10
2.3. Metodología Box-Jenkins	12
2.4. Datos atípicos	15
2.5. Métodos de predicción avanzados	16
2.5.1. Redes Neuronales Autorregresivas	18
2.5.2. Extreme Learning Machine	18
2.6. Modelos Multivariantes: Regresión Dinámica	20
2.6.1. Modelo Dinámico entre dos series	20
3. EL MERCADO INMOBILIARIO EN MADRID	21
3.1. Análisis Univariante de las series	21
3.1.1. Hipotecas Constituidas en la Comunidad de Madrid	21
3.1.2. Capital Hipotecado en la Comunidad de Madrid	29
3.1.3. Viviendas Libres Terminadas en la Comunidad de Madrid	36
3.1.4. Transmisiones de Compraventa de Inmuebles en la Comunidad de Madrid	40
3.1.5. Valor Tasado Medio por M2 en la Comunidad de Madrid	44
3.2. Predicción avanzada: Redes Neuronales	52
3.2.1. Red Neuronal Autorregresiva	52
3.2.2. Red Neuronal ELM	55
3.3. Análisis multivariante: Regresión Dinámica	57
3.4. Comparación de modelos	60

4. CONCLUSIONES	61
BIBLIOGRAFÍA	63
I Anexo	65
A. Código SAS	65
B. Código R	81

Índice de figuras

1.	Red Neuronal con 4 inputs y una capa oculta con 3 nodos	16
2.	Elementos de Red Neuronal. Elaboración propia.	17
3.	Representación gráfica del número de hipotecas constituidas en la Comunidad Madrid	22
4.	Representación gráfica del número de hipotecas en Madrid entre 2003 y 2005	22
5.	Principales estadísticos sobre la serie	23
6.	Tabla de la descomposición estacional	23
7.	Gráfico de la Descomposición estacional	24
8.	Suavizado de Winters Multiplicativo	24
9.	Alisado doble de Holt	25
10.	Predicciones de Alisado doble de Holt	25
11.	Funciones de autocorrelación simple parcial: <i>fas</i> (ACF) y <i>fap</i> (PACF)	26
12.	Funciones de autocorrelación simple y parcial en serie con una diferencia normal	26
13.	Funciones de autocorrelación simple y parcial en serie con diferencia normal y estacional	27
14.	Resultados del modelo ARIMA ajustado	27
15.	Predicciones del modelo ARIMA ajustado	28
16.	Gráfico de las predicciones del modelo ARIMA ajustado	28
17.	Comparativa con los datos reales	29
18.	Representación gráfica de la serie del total de capital hipotecado en la Comunidad de Madrid, en miles de euros.	29
19.	Representación gráfica de la serie entre enero de 2003 y enero de 2005	30
20.	Principales estadísticos sobre la serie	30
21.	Suavizado de Winters Multiplicativo	31
22.	Alisado doble de Holt	31
23.	Predicciones de Alisado doble de Holt	31
24.	Funciones de autocorrelación simple parcial: <i>fas</i> (ACF) y <i>fap</i> (PACF)	32
25.	Funciones de autocorrelación simple parcial para la serie transformada logarítmicamente	32
26.	Funciones de autocorrelación simple parcial en serie con diferencia normal y estacional	33
27.	<i>Outliers</i> detectados en la serie	33
28.	Resultados del ARIMA ajustado	34
29.	Predicciones del ARIMA ajustado incluyendo outliers	35

30.	Gráfico de las predicciones del ARIMA ajustado incluyendo outliers	35
31.	Comparativa de las predicciones y la realidad	36
32.	Representación gráfica de la serie del total de viviendas libres construidas en C.Madrid	36
33.	Principales estadísticos sobre la serie	37
34.	Tabla de la descomposición estacional	37
35.	Suavizado de Winters Aditivo	37
36.	Predicciones de Winters Aditivo	38
37.	Funciones de autocorrelación simple parcial en la serie transformada	38
38.	Resultados de ARIMA ajustado	39
39.	Gráfico de las predicciones ARIMA ajustado	39
40.	Predicciones ARIMA ajustado	40
41.	Comparativa de modelos con la realidad	40
42.	Representación gráfica de la serie de transmisiones de compraventa de inmuebles en C.Madrid	41
43.	Representación gráfica de la serie entre enero de 2009 y enero de 2011	41
44.	Principales estadísticos sobre la serie	42
45.	Tabla de la descomposición estacional	42
46.	Suavizado de Winters Multiplicativo	42
47.	Resultados de ARIMA ajustado	43
48.	Predicciones de ARIMA ajustado	43
49.	Gráfico de las predicciones de ARIMA ajustado	44
50.	Comparativa Realidad - ARIMA	44
51.	Representación gráfica de la serie de valor tasado medio del metro cuadrado en los inmuebles de la Comunidad de Madrid.	45
52.	Suavizado de Winters Multiplicativo	45
53.	Alisado doble de Holt	45
54.	Predicciones de Alisado doble de Holt	46
55.	Representación de los correlogramas	46
56.	Representación de los correlogramas con una diferencia regular	47
57.	Resultados del ARIMA ajustado	47
58.	Predicciones del ARIMA ajustado	48
59.	Gráfico de las predicciones del ARIMA ajustado	48
60.	Diferencias propuestas por R	49
61.	ACF y PACF de la serie diferenciada dos veces	49

62.	ARIMA propuesto con la función <i>autoarima</i>	50
63.	Función ACF y PACF de los residuos	51
64.	Gráfico de la predicción del ARIMA modelado	51
65.	Predicciones del ARIMA modelado	52
66.	Modelo Red Neuronal Autorregresiva	53
67.	Gráfico de la predicción de la red NNAR(1,1)	53
68.	Predicción de la red NNAR(1,1)	53
69.	Red construida número 2	54
70.	Predicción de la red NNAR(2,5)	54
71.	Red construida número 3	54
72.	Predicción de la red NNAR(1,4)	54
73.	Red construida número 4	55
74.	Predicción de la red NNAR(6,6)	55
75.	Resultado de la red ELM de la función <i>nnfor</i>	55
76.	Red ELM propuesta	56
77.	Gráfico de la predicción	56
78.	Predicción del modelo ELM propuesto	56
79.	Red ELM 2	57
80.	Predicción del modelo 2 de red ELM propuesto	57
81.	Construcción de tabla donde las series se utilizarán como variables	58
82.	Resultados Regresión Dinámica	59
83.	Predicciones Regresión Dinámica	59
84.	Comparación de la predicción	60
85.	Error cuadrado medio de la predicción y los valores reales	60

1. INTRODUCCIÓN

Inicialmente una serie temporal se puede definir como una sucesión de valores observados a intervalos regulares en el tiempo. Esta sucesión de valores conocidos permite su análisis, y en segunda instancia, la utilización de técnicas de previsión cuyo objetivo es reducir la incertidumbre del mercado y poder aplicar decisiones de negocio acertadas (el famoso *Business Intelligence*). Estas técnicas de previsión se pueden agrupar en dos bloques: métodos cualitativos (empleados en los casos en los cuales el pasado no proporciona información útil para la previsión), y métodos cuantitativos (donde se tienen registradas informaciones del pasado que se utilizarán para previsiones futuras). Es en estos últimos donde entran en juego las series temporales.

Este TFM pretende poner en práctica algunos de los modelos más empleados en el análisis y predicción de las series en un tema de continua actualidad: los precios en la vivienda. Madrid es una de las Comunidades Autónomas que mayor aumento de precios ha sufrido en los últimos años, por lo que se va a tratar de hacer un estudio de distintas series temporales que tienen una incidencia directa en este auge, y que en última instancia, influyen en la decisión de los particulares a la hora de poner precio a sus propiedades. Los distintos conjuntos que se van a estudiar son los siguientes:

- **Número de hipotecas constituidas en la Comunidad de Madrid:** son los datos publicados por el INE para las hipotecas que se han constituido en la Comunidad, entre 2003 y 2017. Se trata de una serie mensual.

- **Capital constituido de hipoteca en la Comunidad de Madrid:** son datos publicados por el INE del capital hipotecado por los ciudadanos de la Comunidad, entre 2003 y 2017. Son datos son mensuales.

- **Viviendas libres construidas terminadas en la Comunidad de Madrid:** datos publicados por el Ministerio de Fomento. Son datos mensuales entre enero de 2003 y diciembre de 2017.

- **Número de transmisiones de propiedad por medio de compraventa en la Comunidad de Madrid:** se trata de datos mensuales con el número de compraventas de propiedades, entre 2007 y 2017, publicados por el INE.

· **Valor tasado medio por metro cuadrado de la vivienda libre en la Comunidad de Madrid.** Se trata de datos trimestrales , entre 2000 y 2017, proporcionados por el Ministerio de Fomento(cuya fuente de información son los datos de los informes de tasación procedentes de las empresas tasadoras de inmueble en España, y que pertenecen a la Asociación Española de Análisis del Valor). Esta serie es la que se va a tomar como serie principal sobre la que se realizarán los modelos avanzados, ya que se considera la serie que más preponderancia tiene a la hora de determinar el precio que los particulares ponen a sus pisos.

Se analizará y se realizarán métodos de predicción univariante sobre cada una de las series de manera individual, para conocerlas a fondo, y posteriormente, se realizará un análisis multivariante de regresión dinámica sobre ésta última serie, que se analizará en función de las anteriores. Este estudio se realizará con SAS. Asimismo, esta última serie será sobre la que se realizará un estudio más avanzado con redes neuronales, con los paquetes que proporciona R para ello. La bondad de los resultados se analizará mediante la comparación entre la predicción de cada uno de los modelos que se obtengan, y los datos publicados para 2018, es decir, se comparará con la realidad.

1.1. Justificación del proyecto

El mercado inmobiliario en España tiene un peso fundamental en la situación del país. La mayoría de los economistas coinciden en señalar como uno de los principales causantes de la crisis económica del año 2008 el desplome de la construcción, motor de la economía nacional desde años atrás. En dicho año, la economía española sufrió de manera imprevista las consecuencias del desplome, agravadas por la crisis financiera que se vivía en el resto del mundo, y que situó al país en una tasa de paro de hasta un 26 % y una economía moribunda. Es por ello que resulta de especial interés su seguimiento, habida cuenta de los problemas que ya se vivieron y que deberían ser suficientes para prevenirnos.

Una década después de aquella crisis, la situación ha vivido un nuevo cambio en la tendencia de los precios con respecto a aquellos años negros, lo cual se explica por varios factores¹, entre los que destacan una mejora en la economía española, con el crecimiento del empleo como punto

¹Garijo, M., 2018: 5 gráficos que explican la situación del mercado inmobiliario tras diez años de crisis, Business Insider España <https://www.businessinsider.es>

fundamental, el descenso de los tipos de interés, y una vuelta en las financiaciones que dan los bancos. También la construcción se ha visto reactivada, aunque aún se sitúa lejos de los años del *boom* inmobiliario.

Sin embargo, este cambio en la tendencia está trayendo consigo una incipiente preocupación por volver a caer en los mismos errores: la tendencia alcista en los precios tanto de compra como de arrendamiento que se está viviendo en las principales Comunidades Autónomas, está volviendo a ser alarmante. Según publica El País, en España el precio del alquiler se ha incrementado un 18,6 % en los últimos cinco años. El alza ha sido de hasta un 47,5 % en Barcelona y de un 38 % en la ciudad de Madrid, según indica el portal de viviendas de Fotocasa. Y cinco provincias —Baleares, Las Palmas, Salamanca, Barcelona y Madrid—, han alcanzado ya este año su máximo histórico, registrado entre 2007 y 2008².

Este aumento desproporcionado en los alquileres (sin reflejo en un aumento de los salarios), en ciudades como Madrid o Barcelona llega a suponer que se destine más del 40 % de los ingresos mensuales en afrontar las rentas, gestándose una situación en la que que cabría poner medidas.

Afortunadamente, este aumento de precios en el alquiler aún no se ha contagiado del todo al precio de venta, que aunque está creciendo, aún no ha llegado a acercarse a máximos históricos. Según informe trimestral de Tinsa Research, el precio medio de la vivienda en España registró un incremento interanual del 3,8 % en el primer trimestre de 2018, aún un 37,2 % menos que en los máximos de 2007, destacando Madrid, Navarra e Islas Baleares como las provincias con mayor recorrido al alza de estos precios³. Sin embargo, es lógico suponer que la situación de 2008 puede repetirse, especialmente en Madrid, cuyo aumento de precios está por encima de la media. Es éste mercado, y este contexto geográfico -Madrid-, en el que centraremos el estudio. Como se ha mencionado, una de las principales causas de mejora que ha propiciado esta tendencia positiva actual en el aumento de precios es la reactivación en las financiaciones, traducidas en un aumento en la obtención de hipotecas, cuyos datos analizaremos. Además, dado el auge en el precio del alquiler, conlleva en muchos casos que las rentas mensuales de arrendamiento sean superiores a la cuota de amortización de una hipoteca, lo que genera que muchos jóvenes y familias opten por la compra en detrimento del alquiler. Este aumento en la demanda de

²López, S., 2018: España vuelve a especular con la vivienda, El País, <https://elpais.com/economia>

³Tinsa Research, Mercados Locales 1er trimestre de 2018 CCAA, Provincias y Capitales

vivienda para adquirir, se analizará a través de los datos de transmisiones de compraventa, pues es obvio su efecto en el precio de la venta.

1.2. Objetivos del proyecto y materiales a emplear

El objetivo fundamental de este TFM consiste en tratar de anticipar los datos de las distintas series que influyen en el precio de venta de la vivienda, para ser conscientes del contexto actual, del contexto en el futuro próximo (por medio de las predicciones que se emplearán), y de lo cerca o lejos que estamos con respecto a los números que se obtuvieron en los momentos previos al estallido de la crisis.

Para ello, se tratarán de ajustar modelos de series temporales a las series con las que se partirá de inicio, ya mencionadas y con un peso importante en la determinación del precio. Con dichos modelos, se realizarán predicciones, comparando con los valores de la realidad, para ver qué métodos funcionan mejor en según qué series.

Los métodos que se emplearán a tal efecto son métodos univariantes, aplicados en orden según la cronología de su aparición. En primer lugar, se realizarán modelos de suavizado, para continuar con los modelos ARIMA, ampliamente conocidos desde su popularización treinta años atrás, y hoy por hoy, los mayormente empleados en la predicción de series. Para la teoría de estos modelos, se emplea la notación de los libros [3] y [6]. Por último, sobre una de las series (tomada como serie de referencia por ser la serie más íntimamente relacionada con los precios de venta que los particulares de manera subjetiva ponen a sus viviendas) se aplicarán redes neuronales, métodos relativamente novedosos que aún no han desbancado a los modelos ARIMA en su uso, pero que se proponen como una alternativa a tener en cuenta, para lo cual se ha tomado como bibliografía la expuesta al final de este trabajo en los puntos [9], [10], [11] y [14].

Por último, se aplicará regresión dinámica multivariante sobre la serie de referencia (valor de tasación medio), tratada en función del resto de las series analizadas según el párrafo anterior.

En la comparación final se podrá visualizar qué método da mejores resultados, y podremos comprobar, según dicho modelo, cuáles son las predicciones que a priori nos esperan para la serie principal.

2. MODELOS TEÓRICOS EMPLEADOS EN EL ANÁLISIS

2.1. Conocimientos previos

Un **proceso estocástico** es una sucesión de variables aleatorias ordenadas y equidistantes cronológicamente $\{z_t\}$, donde t determina el tiempo de observación.

Se considera que un proceso queda caracterizado mediante los momentos de primer y segundo orden:

- El momento de primer orden o media se define como $\mu_t = E(z_t), \forall t$
- Como momentos de segundo orden respecto a la media se suelen considerar no sólo la varianza, sino también las covarianzas para variables en distintos momentos de tiempo, que vienen dadas por: $\gamma_{t,s} = Cov(z_t, z_s) = E(z_t - \mu_t)(z_s - \mu_s)$. En particular, si $t = s$: $\gamma_{t,t} = E(z_t - \mu_t)^2 = Var(z_t) = \sigma_t^2$

También es común utilizar los coeficientes de autocorrelación para dos instantes distintos:

$\rho_{t,t+k} = \frac{Cov(z_t, z_{t+k})}{\sqrt{Var(z_t)Var(z_{t+k})}}$, siendo t el instante inicial, y k el intervalo entre las observaciones (también llamado retardo).

Por otra parte, se dice que un proceso es estacionario en sentido amplio o débil cuando se verifica que :

- 1) $E(z_t) = \mu, \forall t$ es decir, la media es constante.
- 2) $E(z_t - \mu)^2 = \sigma^2 < \infty, \forall t$, es decir, la varianza es finita y constante a lo largo del tiempo.
- 3) $E(z_{t+k} - \mu)(z_t - \mu) = Cov(z_{t+k}, z_t) = \gamma_k, \forall t$ esto es, la autocovarianza entre dos periodos de tiempo diferentes sólo depende del lapso de tiempo transcurrido.

Ésta será la definición que se usará al hablar de estacionariedad.

Por otro lado, una **serie temporal** se puede definir como una sucesión de valores en el tiempo. Generalmente tiene un carácter aleatorio y se la suele considerar como una de las posibles realizaciones de un proceso estocástico.

La teoría clásica considera que una serie de tiempo está formada por 4 componentes teóricas:

-La tendencia: viene a ser el movimiento general a largo plazo de la serie.

-Las variaciones estacionales: son oscilaciones que se producen con un período igual o inferior a un año, y que se reproducen de manera reconocible en los diferentes años.

-Las variaciones cíclicas: son oscilaciones que se producen con un período superior a un año y que se deben, principalmente, a la alternancia de etapas largas (ciclos) en las que se repite el comportamiento de la serie. Es la más difícil de detectar, pues sus variaciones tienen un período no fácilmente identificable y en muchos casos variable.

-La aleatoriedad: son comportamientos irregulares compuestos por fluctuaciones causadas por sucesos impredecibles o no periódicos.

Estos componentes son la base principal en los que se sustenta el análisis de las series temporales. En general, una serie de tiempo puede ser considerada como un modelo aditivo de esas componentes: $X_t = T_t + S_t + C_t + Z_t$.

O como un modelo multiplicativo, si se considera que las componentes contribuyen al comportamiento de la variable de interés en forma multiplicativa: $X_t = T_t * S_t * C_t * Z_t$

Pues bien, en el estudio de las series temporales son los procesos estacionarios los más comúnmente empleados, pues aunque en el mundo real predominan las series no estacionarias, éstas pueden transformarse en series estacionarias, como se verá más adelante. La herramienta básica con la que se cuenta para un proceso estacionario es la función de autocorrelación, que se define como: $\rho_k = \frac{\gamma_k}{\gamma_0}, k \geq 0$, donde $\rho_k = \rho_{-k}$, pues $\gamma_k = \gamma_{-k}$ por tratarse de un proceso estacionario (es decir, la correlación y la covarianza de dos variables sólo dependen del retardo).

Se denomina **función de autocorrelación simple (fas o ACF en inglés)** a la representación de los coeficientes de autocorrelación en función del retardo. La *fas* proporciona la estructura de dependencia lineal de la serie, y su objetivo es estudiar cómo influye una observación en las siguientes. Pero con la *fas* surge una cadena de influencias entre unas observaciones y las posteriores que no aclara en qué medida influyen individualmente; es decir, con la *fas*, si ρ_1 es significativa, ρ_2 también lo será.

Para subsanar este problema, surge la **función de autocorrelación parcial (fap o PACF en inglés)**, que se define como una función que para cada instante t y cada retardo k , toma un valor igual a la correlación entre z_t y z_{t+k} ajustada por el efecto de los retardos intermedios.

Dicho de otro modo, la *fap* mide la ‘aportación’ que a las variaciones de z_t tiene por ejemplo z_{t-2} , aislados los efectos de z_{t-1} .

Por último, se dice que un proceso es **ergódico** si $\lim_{k \rightarrow \infty} \rho_k = 0$, es decir, si para un k suficientemente grande la dependencia entre las observaciones tiende a 0.

2.2. Modelos Univariantes

2.2.1. Modelos basados en métodos de suavizado o alisado

Estos modelos ajustan la evolución de la serie dando una mayor importancia a los valores más recientes mediante un suavizado exponencial, donde se dan diferentes pesos a las variables, siendo mayor el que se asigna a los últimos datos, y disminuyendo la importancia según vamos hacia atrás en el tiempo. De esta manera se sustituye cada dato de la serie por una media ponderada de las observaciones anteriores, considerando que los pesos de las mismas se reducen exponencialmente. Como principal ventaja se incluye el hecho de que son modelos no excesivamente complejos, cuyo objetivo es eliminar las fluctuaciones aleatorias, para quedarse sólo con la tendencia y la estacionalidad (si la hubiera). Algunos de estos modelos son:

- **De alisado simple** Cuando la serie no tiene tendencia, se puede modelizar como: $X_t = L_t + Z_t$

El cálculo de L_t se realizará en función de un parámetro, α , cuyo valor se sitúa entre 0 y 1, que modula la importancia que tienen las observaciones pasadas sobre el presente. Si este parámetro tiene un valor próximo a 0, entonces las predicciones a lo largo de la serie son muy similares entre sí, y se modifican poco con la nueva información. Cuando $\alpha = 0$, entonces la predicción es una constante a lo largo del tiempo. Si por el contrario, es un valor próximo a 1 la predicción se va adaptando al último valor observado, por lo que se puede decir que los valores alejados en el tiempo no tienen gran influencia en la predicción.

- **De alisado doble de Holt** Cuando la serie presenta tendencia, entonces el método anterior no proporcionará buenos resultados. En ese caso, este método se representa como: $X_t = L_t + b_t + Z_t$

donde se considera que la pendiente de la tendencia varía con el tiempo de manera lineal. Este modelo depende de dos parámetros, el parámetro α ya visto antes, y el parámetro β . Este parámetro β modula la importancia que tienen las observaciones pasadas sobre la pendiente estimada en tiempo t . Éste también oscila entre 0 y 1, de tal manera que para un valor próximo a 0, la pendiente es constante o casi constante. Si tiene un valor próximo a 1, entonces la predicción de la pendiente se va adaptando al último valor observado y por tanto, las observaciones de las pendientes más alejadas de la serie, no tienen apenas influencia sobre la predicción.

- **Modelo de Holt-Winters** Este modelo se puede aplicar en series temporales con tendencia lineal y estacionalidad. Se puede aplicar un modelo aditivo si la estacionalidad no aumenta con el tiempo: $X_t = L_t + b_t + S_t + Z_t$

O multiplicativo, si las variaciones estacionales aumentan con el tiempo: $X_t = (L_t + b_t) * S_t * Z_t$

Este modelo se trata de una extensión del modelo de Holt, donde ahora se aplican 3 parámetros, α , β y γ . Este nuevo parámetro modula la importancia que tienen las observaciones hechas para el mismo periodo de tiempos pasados sobre la predicción en tiempo t . También su valor se sitúa entre 0 y 1, de tal forma que si γ es 0 la predicción en tiempo t va a tomar un valor constante que va a depender de todas las observaciones pasadas dentro de ese mismo periodo. Si por el contrario, es 1, la predicción en tiempo t depende sólo de la observación hecha en tiempo $t-p$, siendo p la frecuencia .

2.2.2. Modelos estacionarios

Los modelos anteriores parten de un enfoque que no tiene en cuenta cómo se genera la serie. Sin embargo, los estadísticos Box y Jenkins popularizaron una serie de modelos que consideraban que la serie temporal viene determinada por un proceso estocástico, siendo ésta una realización particular del mismo. No obstante, para poder efectuar inferencias sobre un proceso es necesario imponer una serie de restricciones a éste: que sea estacionario y ergódico. El proceso estacionario y ergódico más simple de todos es el proceso de ruido blanco, que cuenta con las siguientes propiedades:

1. $E[z_t] = 0, t = 1, 2, \dots$
2. $Var(z_t) = \sigma^2, t = 1, 2, \dots$
3. $Cov(z_t, z_{t-k}) = 0, k = \pm 1, \pm 2, \dots$

donde se ve que la esperanza siempre toma el valor constante cero, la varianza es constante, y las variables del proceso están incorreladas para todos los retardos.

A continuación se analizarán los principales modelos con los que se cuenta para ajustar los procesos estocásticos (de hecho, el término modelo será equivalente a proceso) y donde se hará uso de las funciones de autocorrelación definidas anteriormente.

- **Modelos autorregresivos** Son modelos llamados así por la obtención de z_t mediante regresión sobre los valores desfasados; en estos modelos se impone una dependencia temporal entre las variables del proceso. Un proceso autorregresivo de orden p , o un proceso AR(p) se expresa como sigue:

$$z_t = \phi_1 z_{t-1} + \phi_2 z_{t-2} + \dots + \phi_p z_{t-p} + \varepsilon_t \quad (2.1)$$

donde ε_t es un ‘ruido blanco’ referido al momento actual que se conoce como innovación, ϕ_j son parámetros a estimar, y además aparece la variable desfasada con p de retardo máximo.

Es conveniente mencionar el operador polinomial de retardos L , que aplicado a la variable z_t se define de la siguiente manera: $Lz_t = z_{t-1}$, y $L^k z_t = z_{t-k}$.

Aplicado a la expresión 2.1:

$$\begin{aligned} z_t &= \phi_1 z_{t-1} + \phi_2 z_{t-2} + \dots + \phi_p z_{t-p} + \varepsilon_t = \\ &= (\phi_1 L + \dots + \phi_p L^p) z_t + \varepsilon_t \implies (1 - \phi_1 L - \dots - \phi_p L^p) z_t = \varepsilon_t \end{aligned}$$

Definiendo como $\phi(L)$ a la expresión última entre paréntesis, se obtiene la expresión:
 $\phi(L)z_t = \varepsilon_t$

La *fas* de este modelo se caracteriza por contar con muchos coeficientes no nulos que decrecen hacia cero rápidamente. Por su parte, la *fap* tendrá sólo p valores significativos.

- **Modelos de media móvil** Un proceso de media móvil de orden q , o proceso $MA(q)$ se define como:

$$z_t = \varepsilon_t - \theta_1\varepsilon_{t-1} - \theta_2\varepsilon_{t-2} - \dots - \theta_q\varepsilon_{t-q}$$

En esta expresión z_t se obtiene como un promedio de variables de ruido blanco, con los θ_i de coeficientes de ponderación. Por tanto, son modelos cuyo valor no sólo depende del último valor del proceso de ruido blanco, sino de los q últimos valores. Utilizando como antes el operador polinomial de retardos, $\theta(L) = 1 - \theta_1L - \theta_2L^2 - \dots - \theta_qL^q$, se puede expresar este modelo como $z_t = \theta(L)\varepsilon_t$

En estos modelos, la *fas* se caracteriza por tener los q primeros coeficientes distintos de cero. La *fap* contará con muchos coeficientes no nulos decreciendo rápidamente hacia cero.

- **Modelos mixtos autorregresivos-medias móviles (ARMA)**

Combinan parte $AR(p)$ y $MA(q)$. Su ecuación se define como:

$$z_t - \phi_1z_{t-1} - \dots - \phi_pz_{t-p} = \varepsilon_t - \theta_1\varepsilon_{t-1} - \dots - \theta_q\varepsilon_{t-q}$$

Mediante los ya mencionados operadores de retardo, se tendrá: $\phi_p(L)z_t = \theta_q(L)\varepsilon_t$

En estos modelos, en la *fas* los q primeros valores significativos vienen dados por la parte MA , para a continuación, producirse un decrecimiento provocado por la parte AR . En cuanto a la *fap*, los p primeros valores significativos vienen dados por la parte AR y a continuación se produce un decrecimiento provocado por la parte MA .

2.2.3. Modelos no estacionarios

Dado que no todas las series económicas son estacionarias, cabe mencionar algunos procesos no estacionarios; interesarán aquéllos que se puedan transformar fácilmente en procesos estacionarios, como son los procesos integrados, que mediante la toma de diferencias se convierten en los procesos buscados.

Se dice que un proceso es **integrado** de orden $h \geq 0$, y se representa por $I(h)$, cuando al diferenciarlo h veces se obtiene un proceso estacionario.

Se define el **operador diferencia regular**, ∇ , como $\nabla = 1 - L$, de tal manera que un proceso con una diferencia en la serie z_t será la serie ω_t , donde $\omega_t = \nabla z_t = (1 - L)z_t = z_t - z_{t-1}$; de idéntica forma, un proceso de dos diferencias en z_t será y_t , donde $y_t = \nabla \omega_t = \nabla^2 z_t = (1 - L)^2 z_t = z_t - 2z_{t-1} + z_{t-2}$.

A continuación se exponen algunos de los modelos más importantes para procesos integrados.

- **Paseo aleatorio** Se representa como $z_t = z_{t-1} + \varepsilon_t$. Se trata de un modelo AR(1) con $\phi = 1$. Como se puede ver, es un proceso estocástico muy básico cuya primera diferencia es un ruido blanco.
- **Modelos ARIMA(p,d,q)** Se dice que un proceso z_t es ARIMA(p,d,q) si al tomar diferencias de orden d se llega a un proceso estacionario ARMA(p,q). En un proceso ARIMA(p,d,q), la p es el orden de la parte autorregresiva estacionaria, d el orden de integración y q el orden de la parte de media móvil.

Su ecuación se define como:

$$(1 - \phi_1 L - \dots - \phi_p L^p)(1 - L)^d z_t = (1 - \theta_1 L - \dots - \theta_q L^q) \varepsilon_t$$

o de forma más compacta: $\phi_p(L) \nabla^d z_t = \theta(L)_q \varepsilon_t$

Este tipo de modelos se caracterizan por su *fas*, que tiene coeficientes positivos que decrecen de forma lineal, y que pueden ser distintos de cero incluso para valores altos del retardo.

- **Modelos estacionales** Cuando los datos tienen oscilaciones periódicas, repitiéndose una cierta pauta cada s periodos, se dice que se trata de un modelo estacional.

Un modelo estacional se dice que es multiplicativo, cuando combina componente estacional con componente no estacional. En concreto se estudian los modelos ARIMA estacionales, o modelos ARIMA(P,D,Q)_sx(p,d,q), introducidos por Box y Jenkins para representar muchas de las series temporales que se observan en la práctica.

Estos modelos tratan de forma separada la dependencia regular (asociada a los intervalos de medida de la serie), y la dependencia estacional (asociada a observaciones separadas

por s periodos), para luego incorporar ambas de manera multiplicativa en la construcción del modelo.

Se define el **operador diferencia estacional** como $\nabla_s = 1 - L^s$ (nótese que $\nabla_s \neq \nabla^s = (1 - L)^s$).

Un modelo ARIMA(P,D,Q)_sx(p,d,q) se formula tal que:

$\Phi_P(L^s)\phi_p(L)\nabla_s^D\nabla^d z_t = \Theta_Q(L^s)\theta_q(L)\varepsilon_t$, donde:

$\Phi_P(L^s) = (1 - \Phi_1 L^s - \dots - \Phi_P L^{sP})$ es el operador AR estacional de orden P

$\phi_p(L) = (1 - \phi_1 L - \dots - \phi_p L^p)$ es el operador AR regular de orden p,

$\nabla_s^D = (1 - L^s)^D$ son las diferencias estacionales,

$\nabla^d = (1 - L)^d$ son las diferencias regulares,

$\Theta_Q(L^s) = (1 - \Theta_1 L^s - \dots - \Theta_Q L^{sQ})$ es el operador MA estacional de orden Q,

$\theta_q(L) = (1 - \theta_1 L - \dots - \theta_q L^q)$ es el operador MA regular de orden q, y

ε_t es un proceso de ruido blanco.

2.3. Metodología Box-Jenkins

La metodología Box y Jenkins, nombrada así en honor a los estadísticos George Box y Gwilym Jenkins, se aplica a los modelos ARMA o ARIMA ya estudiados, a fin de encontrar el mejor ajuste de una serie de tiempo y así poder obtener más acertadas previsiones. Dichos estadísticos propusieron realizar este ajuste en cuatro etapas:

- **Identificación del modelo:** Este paso requiere decidir las transformaciones a aplicar, con el fin de obtener una serie estacionaria si ésta no lo fuera, y determinar un modelo ARMA regular (y estacional si fuese necesario). Para ello, en primer lugar, si la serie siguiera una estructura no estacionaria habría que transformar la serie hasta obtener varianza y media constantes. Una vez hecho esto, se procede a determinar el modelo ARMA.
 - **Varianza constante:** Los instrumentos principales que se tienen para detectar si la serie es o no estacionaria en varianza son el gráfico temporal, y el gráfico dispersión-media. Esto es debido a que generalmente, en la mayoría de las series, la variabilidad

tiende a ser mayor conforme mayor es el nivel de la serie, de tal manera que a la desviación típica se la puede considerar como función de la media del tipo $\sigma_t = k\mu_t^\alpha$. De esta forma, transformando la variable z_t en una nueva variable y_t y empleando la transformación de Box-Cox: $y_t = \frac{z_t^{1-\alpha} - 1}{1 - \alpha}$ se llegará a varianzas constantes.

Si α es próxima a 1, la transformación consistirá en tomar logaritmos de z_t (como queda comprobado al tomar límite en la transformación de Box-Cox), si $\alpha = 0$, la varianza es constante y no se requiere modificar la serie, y para valores diferentes, se hace la modificación que establece la transformación Box-Cox.

- Media constante: Para obtener una serie estacionaria en media, se deberán aplicar diferencias regulares y estacionales. Los instrumentos de los que se dispondrá para decidir la diferencia regular serán el gráfico de la serie y las funciones de autocorrelación; es decir, se deberá comprobar si la serie tiene tendencia o deambula de arriba a abajo sin afinidad a ningún nivel, o si la *fas* muestra un decrecimiento lento y lineal; si hay alguna pauta así, es claro indicio de que hay que diferenciar. Se debe partir de la idea de que en caso de duda, se prefiere sobrediferenciar a infradiferenciar, pues los errores son menores. No obstante, en la práctica, el número de diferencias no suele ser mayor que dos.

En cuanto a las diferencias estacionales, se procederá similarmente, comprobando en el gráfico si la serie presenta una pauta repetida cada s observaciones, y si la *fas* presenta coeficientes positivos con decrecimiento lento en los retardos $s, 2s, 3s, \dots$

- Identificación modelo ARMA: Consiste en identificar los órdenes p y q de la estructura ARMA regular, y si hubiera estructura estacional, los órdenes P y Q , según las características de los modelos explicados.

Como instrumentos básicos para ello, en los modelos estacionarios se emplean la función de autocorrelación estimada y la función de autocorrelación estimada parcial, por medio de la comparación con las funciones de autocorrelación teóricas, correspondientes a modelos teóricos con los que puedan guardar cierta similitud.

Por tanto, en esta fase se propondrán posibles modelos que ajustan los datos de las series, en función de los valores que han resultado significativos en la *fas* y *fap*.

- Estimación y selección del modelo ARIMA:** Una vez transformada la serie e identificados los posibles modelos ARMA(p,q) que ajustan los datos a estudiar, el objetivo consistirá en la obtención de estimadores de los parámetros del modelo, esto es, $\phi_1, \phi_2, \dots, \theta_1, \theta_2, \dots$, y la selección de un modelo ARMA entre los modelos estimados. Los principales métodos con los que se cuenta para la estimación de los parámetros son métodos de mínimos cuadrados y de máxima verosimilitud, siendo ésta última la técnica preferida.

En cuanto a la selección de modelos, se debe seleccionar cuál de todos ellos explica mejor la serie observada, para lo que se describen el criterio de Akaike o AIC(Akaike Information Criterium) y el criterio Schwartz o BIC(Bayesian Criterium). Sea k el número de parámetros del modelo M_m ; Según estos criterios, se elige k tal que sea mínimo : $IC = \ln \hat{\sigma}_{MV}^2 + k \frac{C(T)}{T}$ donde $\hat{\sigma}_{MV}^2$ es el estimador de máxima verosimilitud de la varianza, T el tamaño muestral, y $C(T)$ es un término de penalización que varía según el criterio seleccionado: en AIC, $C(T) = 2$ y en BIC, $C(T) = \ln T$.

- Diagnosis del modelo** En esta etapa se trata de detectar posibles discrepancias entre el modelo y la serie observada. Si las hubiera, se deberá repetir el proceso a partir de la etapa de identificación.

Este análisis se deberá hacer a partir de los residuos $\hat{\varepsilon}_t$, de tal manera que si el comportamiento de estos se asemeja al de una serie ruido blanco, con coeficientes incorrelados y distribución normal, existirá adecuación entre modelo y serie. Para verificar esta incorrelación se tiene como procedimiento habitual dibujar dos líneas paralelas a distancia $2/\sqrt{T}$ (siendo T el tamaño de la muestra), del origen en las fas y fap estimadas, y comprobar que los coeficientes \hat{r}_k están dentro de dichas bandas. Si hubiera algún valor próximo a los límites $\pm 2/\sqrt{T}$ en los retardos iniciales, cabe esperar que el modelo es inadecuado.

- Predicción:** Las fases anteriores constituyen un proceso iterativo cuyo resultado final es la obtención de un modelo estimado que sea compatible con la estructura de los datos. Una vez conseguido, la fase siguiente consiste en su utilización en la predicción de valores futuros de la variable estudiada.

Sea $z_T = (z_1, z_2, \dots, z_T)$ la realización observada de una serie temporal de longitud T , y se

desea prever un valor futuro $k > 0$ periodos adelante, es decir, z_{T+k} . Para ello, se necesitaría determinar un predictor óptimo, que sera aquél que minimice el error cuadrático medio de predicción de z_{T+k} . Este predictor óptimo $\hat{z}_T(k)$ es la esperanza de la distribución de la variable z_{T+k} condicionada a los valores observados z_T , es decir, $\hat{z}_T(k) = E[z_{T+k}|z_T]$.

Para hacer predicción se debe tener en cuenta que los parámetros de las funciones y todos los ruidos presentes y pasados $\varepsilon_1, \varepsilon_2, \dots$ son conocidos.

2.4. Datos atípicos

Una limitación de la modelización ARIMA univariante está en que, provocados por distintos motivos, aparecen con frecuencia puntos atípicos (outliers), que dan lugar a errores de predicción, en algunos casos elevados: son circunstancias excepcionales que perturban la dinámica de la serie a modelizar. Existen cuatro casos básicos de atípicos:

- **Atípico aditivo (también llamado *additive outlier* o AO):** Se dice que ocurre un AO sobre una serie temporal en el instante h si el valor de la serie se genera en ese instante de manera distinta al resto. Sea $y_t = \Psi(L)\varepsilon_t$; el modelo que seguirá la serie observada z_t afectada por un AO en $t=h$ será:

$$z_t = \begin{cases} y_t & \text{si } t \neq h \\ y_t + \omega_A & \text{si } t = h \end{cases} \quad \text{Es decir: } z_t = \omega_A I_t^{(h)} + \Psi(L)\varepsilon_t, \text{ donde } I_t^{(h)} = 0 \text{ para } t \neq h, \text{ e } I_h^{(h)} = 1.$$

- **Atípico innovativo (también llamado *innovational outlier* o IO):** Se dice que ocurre un IO en una serie temporal en $t=h$ cuando en ese punto, la innovación (componente aleatorio que representa el efecto agregado de todas las variables que influyen sobre la serie) está directamente afectada por una cantidad desconocida debida a un suceso imprevisto.

Se considera que un IO es un incremento del error de previsión de la serie en un punto debido a un cambio de las condiciones externas, pues la innovación viene a ser ε , que representa dicho error de previsión. El modelo para una serie con este atípico viene dado por $z_t = \Psi(L)(\omega_I I_t^{(h)} + \varepsilon_t)$ donde ω_I es la magnitud del IO.

- **Cambio de nivel (también llamado *level shift* o LS):** Una serie sufre un cambio de nivel LS en el instante h si sigue el modelo $z_t = \omega_L S_t^{(h)} + \Psi(L)\varepsilon_t$, donde $S_t^{(h)}$ es la variable

escalón, con $S_t^{(h)} = 1$ si $t \geq h$ y 0 en otro caso, de tal manera que los valores de la serie observada estarán relacionados con la serie sin el LS mediante: $z_t = \begin{cases} y_t & \text{si } t < h \\ y_t + \omega_L & \text{si } t \geq h \end{cases}$

- **Cambio transitorio (también llamado *Temporary change* o **TC**):** El TC se define por: $z_t = \frac{\omega_{TC}}{1-\delta L} I_t^{(h)} + \Psi(L)\varepsilon_t$, donde si $\delta = 1$ se trata de un cambio de nivel, pues $\nabla^{-1} I_t^{(h)} = S_t^{(h)}$, y si $\delta = 0$, se trata de un atípico aditivo. En la práctica normalmente se fija el valor de δ en 0,7.

2.5. Métodos de predicción avanzados

Los modelos ARIMA han sido uno de los modelos más empleados para la predicción de series de tiempo en las décadas recientes, desde que Box y Jenkins las popularizaron hace más de tres décadas. Sin embargo, estudios más recientes proponen otros métodos alternativos: Las redes neuronales artificiales ([10]). Esto ha motivado que haya controversia y debate acerca de qué método es superior como técnica predictiva, pero aunque los modelos ARIMA son aceptados como modelos de resultados satisfactorios en lo que a predicción se refiere, requieren que las series sean lineales, ya que no son capaces de capturar las relaciones no lineales de los datos. Es ahí donde las redes neuronales se convierten en una opción que consigue capturar tanto las relaciones lineales como las no lineales. Estas redes neuronales están basadas en modelos matemáticos del cerebro. Se organizan en capas, donde las entradas o *inputs* son la capa más baja, y los *outputs* o salidas, la capa más alta. Entre ambas capas se sitúan las llamadas capas ocultas que permiten las relaciones no lineales entre los *inputs* y los *outputs*. Sin esa capa oculta, se estaría hablando de un modelo más simple: regresión lineal.

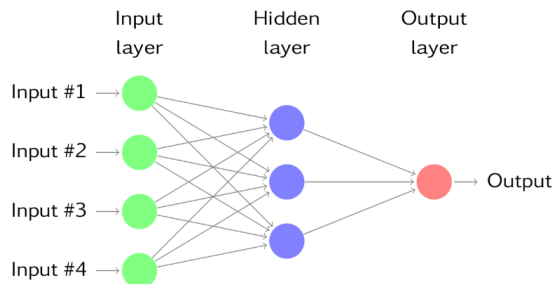


Figura 1: Red Neuronal con 4 inputs y una capa oculta con 3 nodos

Fuente: <https://otexts.org/fpp2/nnetar.html>

El funcionamiento de una red neuronal artificial básica se compone de los siguientes elementos: Entradas o **inputs**, un conjunto de **pesos** (o coeficientes correspondiente a cada entrada), una **función de agregación** Σ (consistente en una combinación lineal de los inputs y los pesos), una **función de activación** f y un conjunto de salidas o **outputs**.

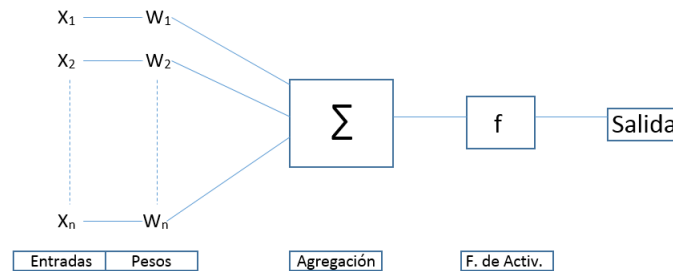


Figura 2: Elementos de Red Neuronal. Elaboración propia.

Es decir, cada neurona recibe una serie de *inputs* ponderados, y una vez realizado el sumatorio $\sum_{i=1}^n X_i w_i$, se aplica una función de activación, resultando una salida. En un primer momento, los pesos son inicializados aleatoriamente pero se actualizan mediante algoritmos de aprendizaje que minimizan una función de coste, representando el conocimiento de la red.

La función de activación en su forma más simplificada es binaria: la neurona se activa, o no. Sin embargo, una vez añadimos capas ocultas, la red se convierte en no lineal. Un ejemplo de ello es el de la figura 1. A esta red se la llama Red neuronal multicapa feed-forward. Es un tipo de red donde los outputs de cada capa de nodos, son los inputs de la capa siguiente. De esta manera, los inputs en la capa oculta vendrán dados por la combinación lineal ([9]):

$$z_j = b_j + \sum_{i=1}^n X_i w_{i,j}$$

A continuación, en la capa oculta se aplica la función f de activación, que puede ser la sigmoide: $s_z = \frac{1}{1 + e^{-z}}$, para obtener el input de la siguiente capa.

Como ya se ha comentado, los pesos toman valores aleatorios al inicio, y estos se actualizan utilizando los datos observados. Por tanto, hay un elemento de aleatoriedad en las predicciones producidas por una red neuronal, por lo que generalmente se entrena varias veces usando diferentes puntos de partida aleatorios, haciendo media a los resultados.

Existen otros tipos de redes tales como las redes recurrentes ([14]), donde existe la retroali-

mentación, es decir, no sólo las salidas de cada capa son las entradas de la capa siguiente sino que además se puede formar ciclos entre ellas, conectándose un nodo consigo mismo, o con otro nodo de la misma capa o con nodos de la capa anterior.

2.5.1. Redes Neuronales Autorregresivas

La red de feed-forward con capa oculta única es el modelo más utilizado en la predicción de series mediante redes. En este caso, los valores rezagados del conjunto de datos se usan como inputs en la red neuronal. Este tipo de red aplicado a las series temporales, se conoce como NNAR (de las siglas neural network autorregresive model, o red neuronal autorregresiva). Por tanto, la notación que se emplea es NNAR(p,k), donde p indica el número de retardos (que serán los inputs), y k el número de nodos de la capa oculta única. De esta manera, un NNAR(p,0) será equivalente a un ARIMA(p,0,0), pero sin las restricciones aplicadas para la condición de estacionariedad. La representación matemática sería:

$$y_t = \alpha_0 + \sum_{j=1}^k \alpha_j g \left(\beta_{0j} + \sum_{i=1}^p \beta_{ij} y_{t-i} \right) + \epsilon_t$$

donde y_t es el output de la red, α_j y β_{ij} son los pesos, p , los inputs, k el número de nodos de la capa oculta y g es la función de activación. Es decir, y_t es una función tal como:

$$y_t = f(y_{t-1}, y_{t-2} + \dots + y_{t-i}, w) + \epsilon_t$$

donde w es un vector formado por todos los parámetros y f una función según la estructura de la red neuronal y los pesos.

En el caso de series estacionales, se tendría NNAR(p, P, k)_s, donde como inputs estarían los retardos ($y_{t-1}, y_{t-2} + \dots + y_{t-p}, y_{t-s}, y_{t-2s} + \dots + y_{t-Ps}$).

2.5.2. Extreme Learning Machine

De la misma manera que las redes neuronales autorregresivas, la idea del modelo conocido como ELM es la de crear un nuevo algoritmo de aprendizaje para las redes neuronales feed-

forward de una única capa oculta en su aplicación a las series de tiempo ([8]). La formulación ELM consiste en dar solución a un sistema de ecuaciones lineales planteado en términos de los pesos desconocidos que conectan la capa oculta con la capa de salida. Matemáticamente, este algoritmo se plantea de la siguiente manera:

Sea un conjunto de datos $(x_{i1}, x_{i2}, \dots, x_{im})$ como vector input de la red. Y sea $(y_{i1}, y_{i2}, \dots, y_{ik})$ el vector output. $w_i = (w_{i1}, w_{i2}, \dots, w_{is})$ será el vector de pesos que conecta el nodo i-ésimo de la capa oculta con el nodo s-ésimo de los inputs, b_i el sesgo, $\beta_i = (\beta_{i1}, \beta_{i2}, \dots, \beta_{ik})$ el vector de pesos que conecta la capa oculta y los outputs, y L , el número de nodos de la capa oculta. Entonces, para M muestras distintas arbitrarias como datos de entrada y $\forall j = 1, 2, \dots, M$:

$$y_j = \sum_{i=1}^L \beta_i G(w_i, b_i, x_j)$$

siendo $G(w_i, b_i, x_j) = g(w'_i x + b_i)$ con g como función de activación. Esta expresión se puede poner en forma matricial como: $H\beta = Y$, siendo:

$$H_{M \times L} = \begin{bmatrix} G(w_1, b_1, x_1) & \dots & G(w_L, b_L, x_1) \\ \vdots & & \\ G(w_1, b_1, x_M) & \dots & G(w_L, b_L, x_M) \end{bmatrix} \quad \beta_{L \times k} = \begin{bmatrix} \beta'_1 \\ \vdots \\ \beta'_L \end{bmatrix} \quad Y_{M \times k} = \begin{bmatrix} y'_1 \\ \vdots \\ y'_M \end{bmatrix}$$

donde H , llamada matriz de salida de la capa oculta, no tiene por qué ser cuadrada. La columna i-ésima de H es el output del i-ésimo nodo oculto para los inputs x_1, x_2, \dots, x_M .

Este algoritmo resuelve el modelo anterior planteado mediante 3 pasos, para un conjunto de datos de entrenamiento (M muestras distintas de los datos) y una función g de activación:

1) Se asignan aleatoriamente los vectores de peso que conectan los inputs y la capa oculta w_i , y el sesgo b_i .

2) Se calcula H .

3) Se calculan los pesos β_i , a partir de: $\beta = H^+ Y$, donde H^+ es la pseudoinversa o inversa generalizada de Moore-Penrose ([8]).

De esta manera, este modelo mejora a otros algoritmos por el hecho de que es más flexible en la función de activación, dado que no impone diferenciabilidad, tiene un aprendizaje y entrenamiento rápido y da un paso más en el modelado avanzado de series temporales.

2.6. Modelos Multivariantes: Regresión Dinámica

Un modelo de regresión dinámica es una representación de la relación existente entre dos o más series temporales, donde se supone conocida la dirección de causalidad, es decir, existe una cierta variable y_t dependiente o *endógena* y una (o unas) variable x_t explicativa o *exógena*. Este tipo de modelos se conocen como modelos de **función de transferencia** o modelos econométricos dinámicos, y su utilidad es describir cómo se transmiten los efectos desde una variable a otra.

2.6.1. Modelo Dinámico entre dos series

Se parte de la suposición de que sólo existe relación unidireccional de la serie x a la serie y , siendo ambas estacionarias. Esta relación se puede expresar mediante:

$$y_t = v_0x_t + v_1x_{t-1} + v_2x_{t-2} + \dots + n_t \quad (2.2)$$

donde los coeficientes v_i describen la relación dinámica entre las series, y n_t es un proceso estacionario que recoge el efecto de todas las otras variables que pueden tener efecto sobre y_t , y se denomina **perturbación** de la relación. De forma simplificada con el operador de retardos será:

$$y_t = v(L)x_t + n_t$$

con $v(L) = v_0 + v_1L + v_2L^2 + \dots$ que se denomina **función de transferencia**. Los coeficientes v_i determinan la **función de respuesta a impulsos**.

Se denomina **ganancia** de la función de transferencia a $g = v(1) = \sum_{i=0}^{\infty} v_i$, representando el efecto a largo plazo que experimenta y_t cuando x_t aumenta en una unidad y permanece constante a continuación.

Por su parte, la perturbación n_t sigue un proceso ARMA estacionario tal que $\phi(L)n_t = \theta(L)\varepsilon_t$, siendo ε_t un proceso de ruido blanco.

El modelo formado por la ecuación (2.2) y la ecuación de la perturbación se conoce como **modelo de regresión dinámica**.

No obstante, se puede plantear una forma alternativa de escribir la función de transferencia, útil en aquellos casos donde el proceso x_t tiene alta autocorrelación y sería muy complicado separar los efectos de $x_t, x_{t-1}, x_{t-2}, \dots$, además de contener infinitos parámetros a estimar. Este método alternativo se representa por $v(L) = \frac{\omega_m(L)}{\delta_a(L)}L^b$, donde:

- $\omega_m(L) = \omega_0 + \omega_1(L) + \dots + \omega_m L^m$ es el numerador de la función de transferencia y se puede considerar la parte MA de un modelo ARMA, con orden m;

- $\delta_a = 1 - \delta_1 L - \dots - \delta_a L^a$ es el denominador, considerándose la parte AR de orden a. Para que el sistema aporte una respuesta finita a una modificación finita de x_t se tiene que cumplir que expresando $\delta_a(L) = (1 - \alpha_1 L) \dots (1 - \alpha_a L)$, los coeficientes α_i sean menores que la unidad;

- L^b indica que la relación puede establecerse con un retardo inicial b.

El modelo se puede generalizar para p variables explicativas, mediante:

$$y_t = \frac{\omega_1(L)L^{b_1}}{\delta_1(L)}x_{1t} + \dots + \frac{\omega_p(L)L^{b_p}}{\delta_p(L)}x_{pt} + n_t$$

3. EL MERCADO INMOBILIARIO EN MADRID

Como se indicó, en este TFM se va a tratar de hacer un estudio de distintas series temporales que tienen importancia en el mercado inmobiliario, a las que se aplicarán métodos de predicción sobre ellas, mediante SAS fundamentalmente. La predicción se comparará finalmente con los valores reales para decidir el mejor modelo.

3.1. Análisis Univariante de las series

3.1.1. Hipotecas Constituidas en la Comunidad de Madrid

En primer lugar, se representa la serie:

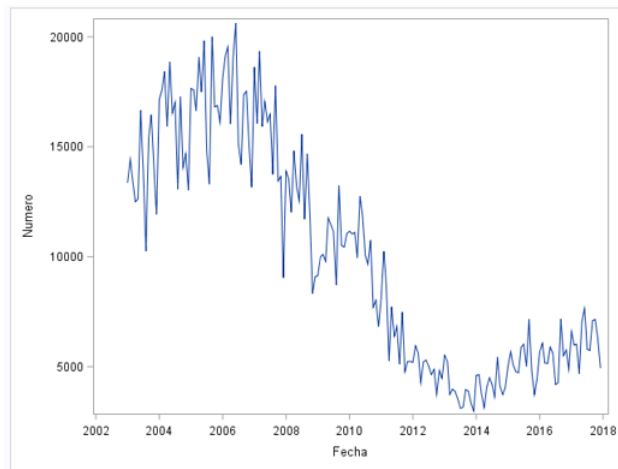


Figura 3: Representación gráfica del número de hipotecas constituidas en la Comunidad Madrid

Esta representación de la figura 3 aporta información de la tendencia de la serie, viendo un crecimiento en el número de hipotecas constituidas que vio su punto más álgido alrededor del año 2006; sin embargo, a partir de 2008 toma un descenso impactante que alcanza su punto más bajo en 2014; desde ahí, y de manera lenta, se vuelve a observar una tendencia positiva.

Si se fija un periodo menor, es posible observar el comportamiento anual, en la figura 4.

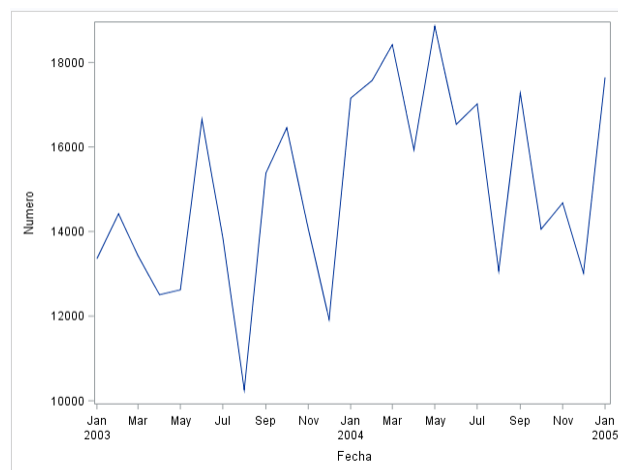


Figura 4: Representación gráfica del número de hipotecas en Madrid entre 2003 y 2005

En esta figura 4, se ve que en agosto y diciembre se sucede un descenso en el número de hipotecas y un aumento en septiembre y enero, por lo que podría tener componente estacional.

Haciendo la descomposición estacional de la serie, mediante las figuras 5, 6, 7 se puede ver por un lado los estadísticos calculados para cada elemento del periodo (siendo la longitud del

periodo 12 meses), así como la descomposición estacional. Se puede observar que según la media, el mes en el que más hipotecas se constituyen es septiembre, y el mes que menos es diciembre. También es posible ver los coeficientes estacionales en 6, donde por ejemplo en septiembre de 2004 fue de 1.1662, indicando que las hipotecas realizadas en este mes fueron un 16.62 por ciento superior a la media global anual.

En la figura 7 se ve la serie y tendencia (TREND-CYCLE), la serie sin tendencia (SEASONALLY IRREGULAR), la componente irregular (IRREGULAR), y la serie desestacionalizada (SEASONALLY ADJUSTED). Se observa muy claramente la situación vivida en todo el país, donde antes de 2008 se llegaron a números altísimos en el número de hipotecas que los Bancos concedían a la gente, y el increíble descenso cuando la crisis económica se empezó a hacer patente.

Estadísticos estacionales para la variable Numero						
Índice estacional	N	Mínimo	Máximo	Suma	Media	Desviación estándar
1	15	4611.000	18622.00	159861.0	10657.40	5366.5845
2	15	4632.000	19160.00	163174.0	10878.27	5178.1347
3	15	3732.000	19537.00	158694.0	10579.60	5778.0215
4	15	3140.000	19064.00	145273.0	9684.87	5572.3237
5	15	3880.000	19114.00	161382.0	10758.80	5611.3723
6	15	3537.000	20623.00	164443.0	10962.87	5861.2957
7	15	3106.000	17020.00	149192.0	9946.13	5147.4825
8	15	3156.000	14191.00	126187.0	8412.47	4096.4949
9	15	3959.000	20006.00	169662.0	11310.80	5469.3706
10	15	3743.000	17523.00	142466.0	9497.73	5161.5477
11	15	3336.000	16879.00	134030.0	8935.33	4776.0999
12	15	2989.000	16114.00	121205.0	8080.33	4127.9349

Figura 5: Principales estadísticos sobre la serie

Descomposición estacional para la variable Numero						
Obs	Tiempo	Estación	Ciclo de tendencia	Estacional	Irregular	Ajustado estacionalmente
1	JAN2003	1	.	1.064557	.	12549.82
2	FEB2003	2	.	1.099932	.	13110.82
3	MAR2003	3	.	1.032740	.	13004.24
4	APR2003	4	.	0.933248	.	13399.44
5	MAY2003	5	.	1.074473	.	11747.15
6	JUN2003	6	.	1.076497	.	15470.55
7	JUL2003	7	13902.38	0.998434	0.998804	13885.75
8	AUG2003	8	14191.96	0.857064	0.842937	11962.93
9	SEP2003	9	14531.46	1.166274	0.907500	13187.30
10	OCT2003	10	14882.29	0.945201	1.169779	17408.99
11	NOV2003	11	15284.83	0.905084	1.016766	15541.10
12	DEC2003	12	15539.67	0.846497	0.906245	14082.75

Figura 6: Tabla de la descomposición estacional

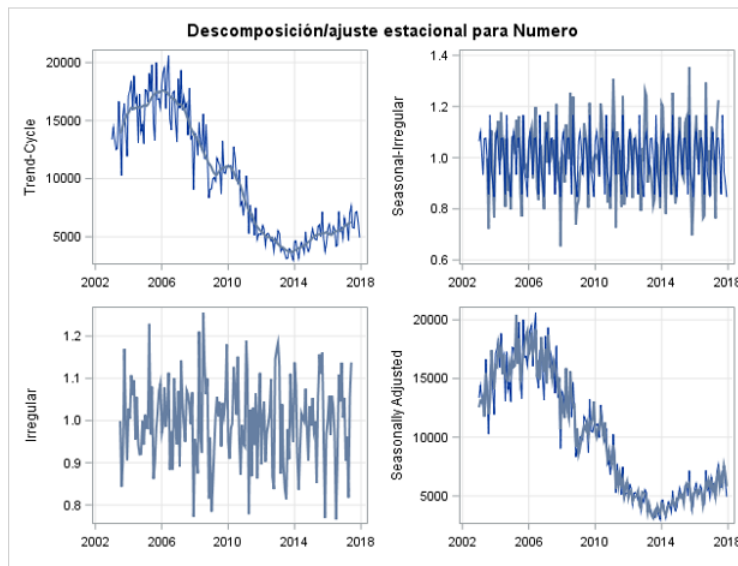


Figura 7: Gráfico de la Descomposición estacional

Una vez vista la serie y analizado su comportamiento de tendencia y estacionalidad, se va a tratar de emplear algunos de los modelos vistos en la teoría para ajustar la serie y predecirla.

En primer lugar, se va a realizar un modelo de suavizado. Dado que tiene tendencia, y presumiblemente, hay estacionalidad, se procede a hacer un suavizado de Winters Multiplicativo en la figura 8

Método de Winters (Multiplicativo) Estimador del parámetro				
Parámetro	Estimador	Error estándar	Valor t	Aprox Pr > t
Peso de nivel	0.33949	0.03795	8.94	<.0001
Peso de tendencia	0.06152	0.01838	3.35	0.0010
Peso estacional	0.02001	0.01466	1.37	0.1739

Figura 8: Suavizado de Winters Multiplicativo

Fijándonos en el p-valor, el parámetro de estacionalidad no es significativamente distinto de cero, por lo que se rechaza que la serie sea estacional, y por tanto, se rechaza este modelo. Haciendo un suavizado lineal con tendencia (alisado doble de Holt), se observa en la figura 9 que esta vez sí hay unos parámetros significativamente distintos de cero, pues su p-valores son menores de 0,05. En este caso, los parámetros α y β toman unos valores de 0,24 y 0,058.

Suavizado exponencial lineal Estimador del parámetro				
Parámetro	Estimador	Error estándar	Valor t	Aprox Pr > t
Peso de nivel	0.24422	0.03422	7.14	<.0001
Peso de tendencia	0.05755	0.01856	3.10	0.0022

Figura 9: Alisado doble de Holt

Mediante este modelo, se tienen las predicciones de la figura 10

Predicciones para la variable Numero					
Obs	Tiempo	Predicciones	Error estándar	Límites de confianza al 95%	
181	JAN2018	6301.4455	1693.9175	2981.4283	9621.4628
182	FEB2018	6337.3496	1749.5011	2908.3905	9766.3087
183	MAR2018	6373.2536	1809.2954	2827.0998	9919.4075
184	APR2018	6409.1577	1873.1999	2737.7533	10080.5621
185	MAY2018	6445.0618	1941.1007	2640.5744	10249.5492
186	JUN2018	6480.9658	2012.8748	2535.8037	10426.1280
187	JUL2018	6516.8699	2088.3944	2423.6921	10610.0477
188	AUG2018	6552.7740	2167.5295	2304.4942	10801.0537
189	SEP2018	6588.6780	2250.1505	2178.4641	10998.8920
190	OCT2018	6624.5821	2336.1302	2045.8509	11203.3132
191	NOV2018	6660.4861	2425.3452	1906.8969	11414.0754
192	DEC2018	6696.3902	2517.6766	1761.8347	11630.9457

Figura 10: Predicciones de Alisado doble de Holt

Se va a tratar de ajustar ahora otro de los modelos vistos en la teoría, y uno de los principales métodos de predicción aplicados a las series temporales: los ARIMA y la metodología de Box Jenkins. En primer lugar, se van a representar las funciones de autocorrelación que tiene la serie sin hacer ajustes, mediante la figura 11. En esta figura, se observa un decrecimiento muy lento en la ACF, a la par que se observa que la serie no es estacionaria, pues no es estable en media. Por tanto, se debe realizar al menos una diferenciación a fin de obtener una serie estacionaria.

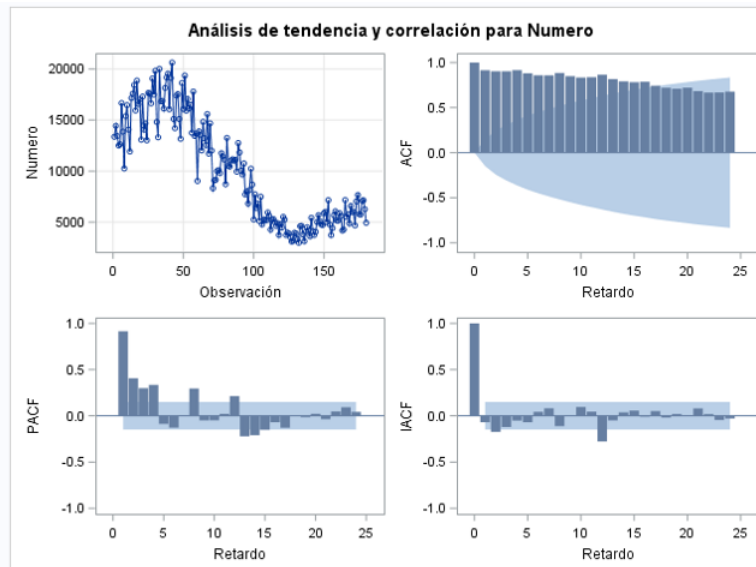


Figura 11: Funciones de autocorrelación simple parcial: *fas* (ACF) y *fap* (PACF)

Se comienza por hacer una diferenciación normal, en la figura 12, donde observando la ACF, vemos que algunas barras son significativas, especialmente la del retardo 12, por lo que se realiza una diferencia estacional, cuyos resultados se ven en la figura 13. A partir de esta figura 13 se puede estimar un modelo ARIMA. Para empezar, se observa una barra significativa en la ACF y otra en el retardo 12, lo que hace suponer que hay componente MA tanto en el retardo 1 como en el 12 estacional. También se observan dos barras significativas en la PACF que sugieren componente AR, también en su parte estacional.

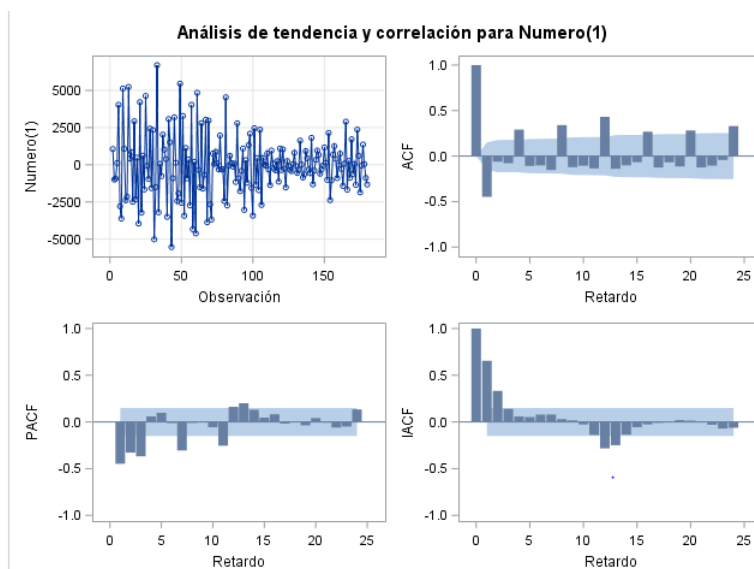


Figura 12: Funciones de autocorrelación simple y parcial en serie con una diferencia normal

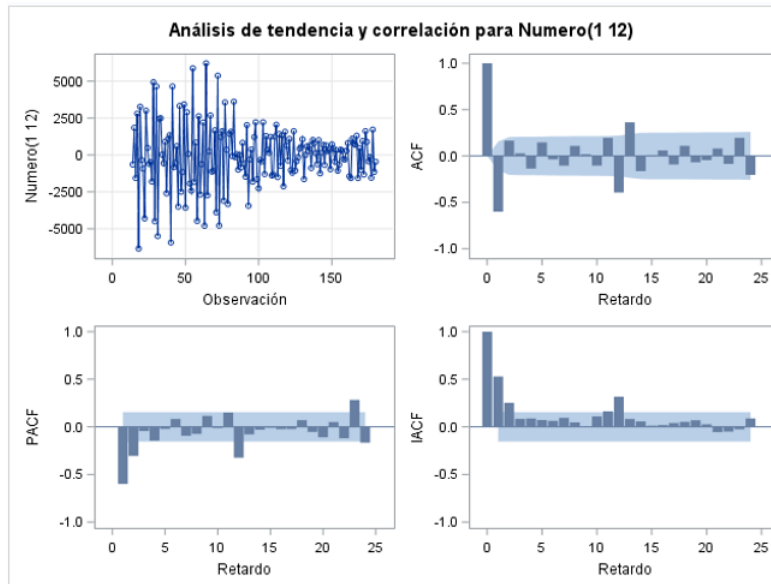


Figura 13: Funciones de autocorrelación simple y parcial en serie con diferencia normal y estacional

Dado que el sobreajuste es evitable, por el principio de parsimonia se prefiere comenzar por modelos simples. Tras probar varios modelos y sus resultados de AIC y p-valores, se obtiene el ARIMA (1, 1, 1), (1, 1, 0)₁₂, cuyos resultados se ven en la figura 14.

Estimación por mínimos cuadrados condicional					
Parámetro	Estimador	Error estándar	Valor t	Aprox Pr > t	Retardo
MA1,1	0.51862	0.09744	5.32	< .0001	1
AR1,1	-0.26018	0.10992	-2.37	0.0191	1
AR2,1	-0.33760	0.07473	-4.52	< .0001	12

Varianza Estimación	2351623
Error Std Estimación	1533.5
AIC	2926.891
SBC	2936.245
Número de residuales	167

* AIC y SBC no incluyen determinante de la log.

Correlaciones de las estimaciones de parámetro			
Parámetro	MA1,1	AR1,1	AR2,1
MA1,1	1.000	0.726	-0.070
AR1,1	0.726	1.000	-0.011
AR2,1	-0.070	-0.011	1.000

Comprobación de autocorrelación de residuales										
Para retardo	Chi-cuadrado	DF	Pr > ChiSq	Autocorrelaciones						
6	2.12	3	0.5475	-0.006	-0.006	0.047	-0.049	0.079	-0.035	
12	10.57	9	0.3063	-0.116	0.072	0.079	-0.076	-0.049	-0.119	
18	16.96	15	0.3215	0.182	-0.006	-0.028	0.017	0.003	0.021	
24	42.64	21	0.0035	-0.079	-0.015	0.059	0.057	0.104	-0.326	
30	45.34	27	0.0149	-0.037	0.069	-0.027	0.051	0.015	-0.060	

Figura 14: Resultados del modelo ARIMA ajustado

Se observa que el p-valor de los parámetros son menores que 0,05, por lo que se rechaza la hipótesis nula, y por tanto, se acepta que son significativamente distintos de 0. Además, se ve que los retardos 6, 12 y 18 están incorrelados, pues sus p-valores son mayores a 0,05. Por tanto, se acepta este modelo ARIMA.

Este modelo, siguiendo lo explicado en la teoría, quedaría como sigue:

$$(1 + 0,3376L^{12})(1 + 0,26018L)(1 - L^{12})(1 - L)z_t = (1 - 0,51862L)\varepsilon_t$$

$$\Rightarrow z_t = 0,73982z_{t-1} + 0,26018z_{t-2} + 0,6624z_{t-12} - 0,4901z_{t-13} -$$

$$0,172324z_{t-14} + 0,3376z_{t-24} - 0,2498z_{t-25} - 0,08784z_{t-26} - 0,51862\varepsilon_{t-1} + \varepsilon_t$$

Las predicciones de este modelo y el gráfico, son las de la figura 15 y 16

Predicciones para la variable Numero				
Obs	Predicción	Error Std	Límites de confianza al 95%	
181	7267.6795	1533.5002	4262.0743	10273.2847
182	6791.9061	1570.5672	3713.6509	9870.1612
183	6570.6281	1699.7347	3239.2093	9902.0469
184	5658.3341	1792.4733	2145.1510	9171.5172
185	7478.9881	1887.1160	3780.3087	11177.6674
186	7785.2730	1975.6071	3913.1541	11657.3918
187	6078.0204	2060.7056	2039.1117	10116.9290
188	6070.8592	2142.3252	1871.9790	10269.7394
189	7951.1232	2220.9727	3598.0966	12304.1497
190	7417.1025	2296.9225	2915.2171	11918.9880
191	6920.3098	2370.4417	2274.3293	11566.2902
192	5743.9065	2441.7479	958.1684	10529.6445

Figura 15: Predicciones del modelo ARIMA ajustado

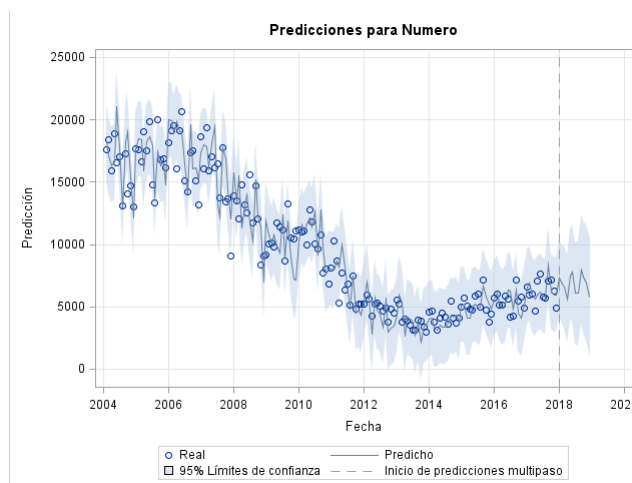


Figura 16: Gráfico de las predicciones del modelo ARIMA ajustado

A la entrega de este trabajo, se han conocido los valores que ha tenido la serie entre enero y mayo de 2018. Haciendo una comparativa de las predicciones con Suavizado y ARIMA, se obtiene el siguiente gráfico de la figura 17. Se observa que en abril ha habido un valor que ha distorsionado la predicción, pero en enero, marzo y mayo se ha ajustado bastante con ARIMA.

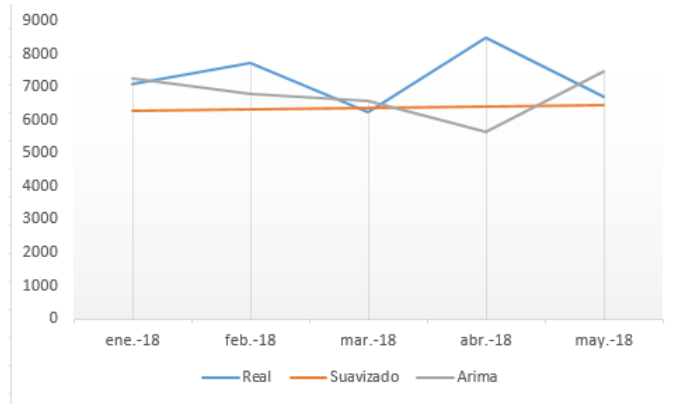


Figura 17: Comparativa con los datos reales

3.1.2. Capital Hipotecado en la Comunidad de Madrid

Al igual que antes, se comienza representando la serie completa, y un periodo menor (figuras 18 y 19), para presentar una idea previa acerca del comportamiento de la serie.

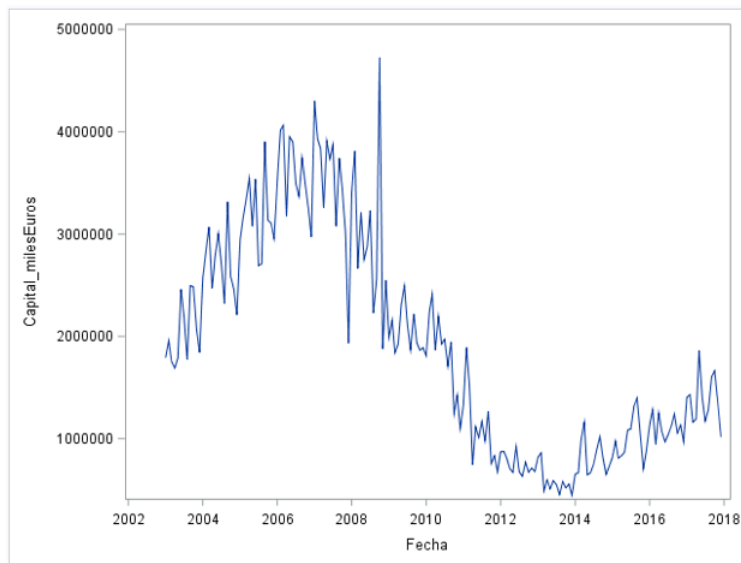


Figura 18: Representación gráfica de la serie del total de capital hipotecado en la Comunidad de Madrid, en miles de euros.

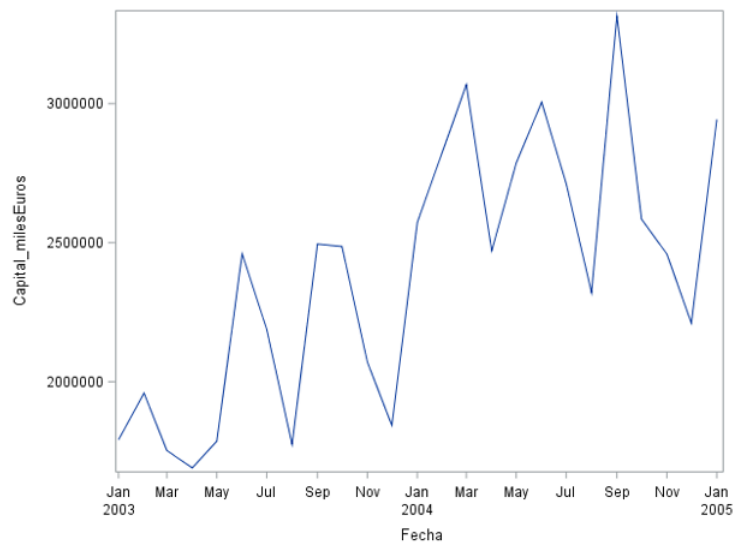


Figura 19: Representación gráfica de la serie entre enero de 2003 y enero de 2005

Se puede observar en la figura 18 que la tendencia fue positiva hasta 2006-2008, que pasó a ser negativa. Además, se observan picos que habría que tener en cuenta en el análisis, pues son **outliers que pueden influir en las predicciones**. Se observa también, en la figura 19, que suele haber un descenso en el capital hipotecado en los meses de agosto y diciembre, y repuntes en septiembre y enero.

Analizando los estadísticos mediante la figura 20 se observa que no hay un fuerte componente estacional, pues los valores en la tabla no presentan grandes diferencias. Se ve también que los meses más fuertes en cuanto a capital hipotecado son febrero y septiembre.

Estadísticos estacionales para la variable Capital_milesEuros						
Índice estacional	N	Mínimo	Máximo	Suma	Media	Desviación estándar
1	15	653144.0	4303037	29343675	1956245	1142042.2
2	15	666834.0	4016308	32091684	2139446	1169315.5
3	15	489810.0	4064854	29716874	1981125	1178003.6
4	15	597788.0	3546420	27625668	1841711	1044811.9
5	15	510059.0	3951228	29515155	1967677	1156773.6
6	15	589472.0	3902589	30590571	2039371	1175544.8
7	15	552613.0	3874075	28725603	1915040	1092662.0
8	15	454828.0	3367364	25734432	1715629	882734.6
9	15	581145.0	3902701	31809267	2120618	1131867.8
10	15	519717.0	4719936	29576341	1971756	1280029.0
11	15	556717.0	3260368	25032858	1668857	947541.6
12	15	447334.0	2973306	22847454	1523164	864117.4

Figura 20: Principales estadísticos sobre la serie

Se realizará en primer lugar la aplicación de los modelos de suavizado. Presuponiendo tendencia y estacionalidad, se realiza en primer lugar el método de Winters, lo que da los resultados

de la figura 21, donde se observa nuevamente que el parámetro de la estacionalidad presenta un p-valor que sugiere aceptar la hipótesis nula, es decir, no es significativo.

Método de Winters (Multiplicativo) Estimador del parámetro				
Parámetro	Estimador	Error estándar	Valor t	Aprox Pr > t
Peso de nivel	0.19322	0.03074	6.29	<.0001
Peso de tendencia	0.12933	0.03467	3.73	0.0003
Peso estacional	0.02826	0.01852	1.53	0.1288

Figura 21: Suavizado de Winters Multiplicativo

Por tanto, habría que hacer un suavizado lineal de Holt para series con tendencia sin estacionalidad, lo cual da los resultados de la figura 22. En este caso los p-valores indican que los parámetros sí son significativos. α toma un valor de 0,16, y β un valor de 0,14.

Suavizado exponencial lineal Estimador del parámetro				
Parámetro	Estimador	Error estándar	Valor t	Aprox Pr > t
Peso de nivel	0.15714	0.02851	5.51	<.0001
Peso de tendencia	0.14112	0.03952	3.57	0.0005

Figura 22: Alisado doble de Holt

Las predicciones de este modelo son los de la figura 23

Predicciones para la variable Capital_milesEuros					
Obs	Tiempo	Predicciones	Error estándar	Límites de confianza al 95%	
181	JAN2018	1441608.290	383342.33	690271.124	2192945.455
182	FEB2018	1455463.333	389456.68	692142.264	2218784.401
183	MAR2018	1469318.375	397042.35	691129.661	2247507.090
184	APR2018	1483173.418	406194.84	687046.162	2279300.674
185	MAY2018	1497028.461	416984.28	679754.298	2314302.623
186	JUN2018	1510883.504	429455.56	669166.077	2352600.930
187	JUL2018	1524738.546	443629.74	655240.230	2394236.863
188	AUG2018	1538593.589	459506.51	637977.384	2439209.794
189	SEP2018	1552448.632	477067.35	617413.813	2487483.451
190	OCT2018	1566303.675	496279.10	593614.506	2538992.843
191	NOV2018	1580158.717	517097.52	566666.205	2593651.229
192	DEC2018	1594013.760	539470.54	536670.922	2651356.598

Figura 23: Predicciones de Alisado doble de Holt

Se va ahora a ajustar el modelo ARIMA. Para ello, se representan los correlogramas en primer lugar (figura 24).

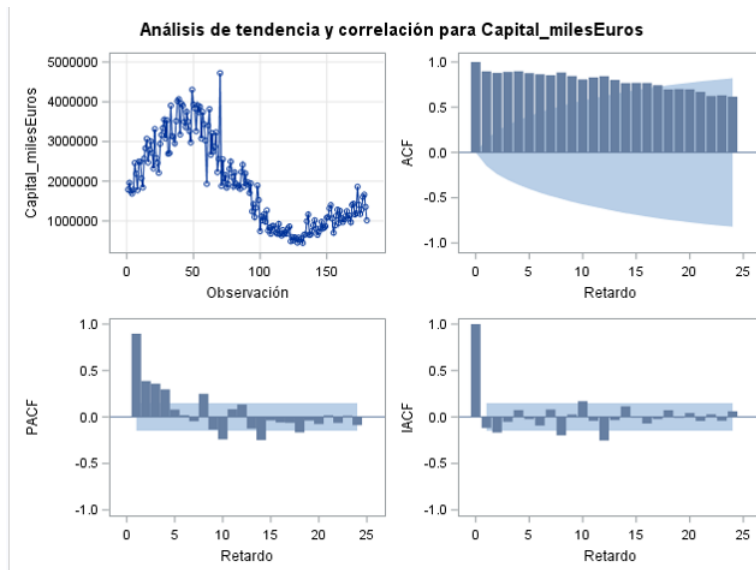


Figura 24: Funciones de autocorrelación simple parcial: *fas* (ACF) y *fap* (PACF)

En este caso, se ve gráficamente que la serie no es estacionaria, ya que no es estable ni en media, ni en varianza, por lo que se va a proceder a hacer una transformación logarítmica para estabilizar la varianza antes de proceder a tomar diferencias en la figura 25, en la cual se ve que hay que proceder a hacer una diferencia regular y otra estacional, ya que parece que en la PACF hay barras significativas en el retardo 12. Haciendo esto, obtenemos los correlogramas de la figura 26

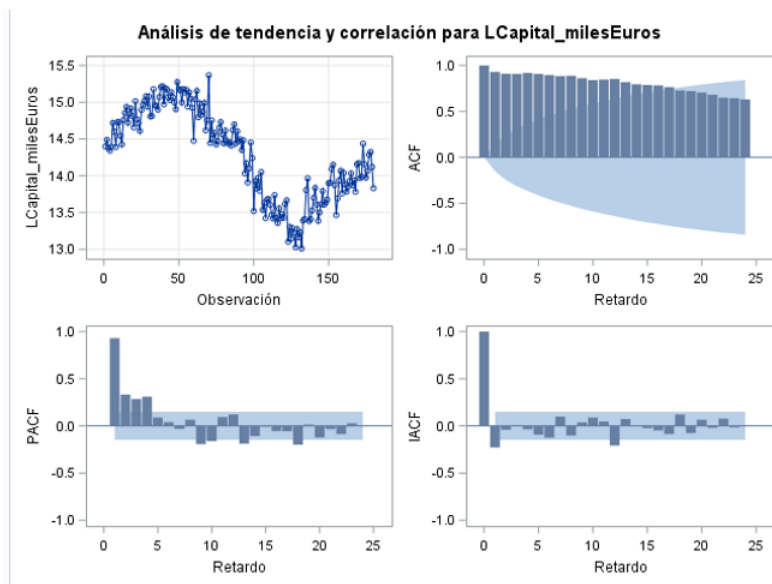


Figura 25: Funciones de autocorrelación simple parcial para la serie transformada logarítmicamente

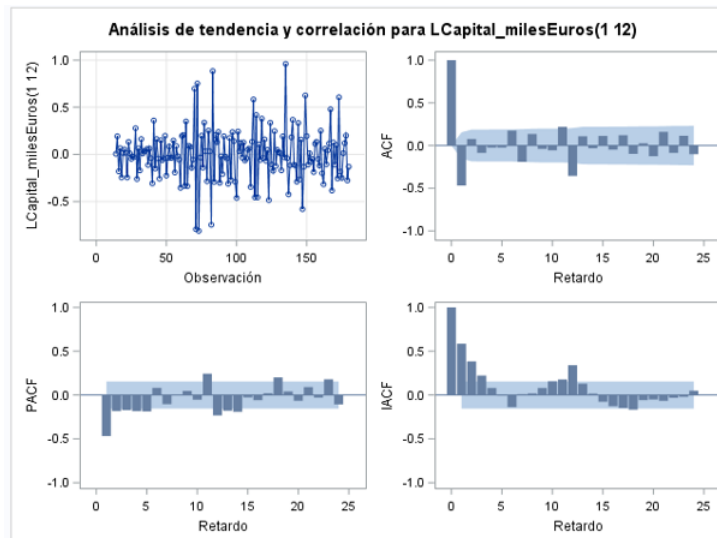


Figura 26: Funciones de autocorrelación simple parcial en serie con diferencia normal y estacional

Se observa al menos una barra significativa en el retardo 1 y una barra en el retardo 12 de la ACF, y una barra en la PACF. Sin embargo, en las representaciones gráficas se habían visto posibles **outliers**, por lo que antes de ajustar el modelo, con la serie transformada, se comprueban los datos atípicos que detecta SAS en la modelización ARIMA, mediante la figura 27. Hay que tener en cuenta estos datos pues pueden afectar a la predicción.

Resumen de detección de valores extremos	
Número máximo buscado	3
Número encontrado	3
Significación empleada	0.05

Detalles de valores extremos					
Obs	ID de tiempo	Tipo	Estimador	Chi-cuadrado	Approx Prob>ChiSq
70	10-2008	Aditivo	0.73944	39.21	<.0001
100	04-2011	Desplazamiento	-0.51315	23.37	<.0001
136	04-2014	Aditivo	0.53035	20.33	<.0001

Figura 27: *Outliers* detectados en la serie

Según se detecta, los 3 *outliers* más significativos, están en la observación de octubre de 2008, abril de 2011, y abril de 2014, siendo uno de salto y 2 aditivos. Por tanto, antes de modelizar, se deben incorporar en el fichero de datos estos 3 puntos, para que se tengan en cuenta como variables regresoras y poder proceder a continuación a ajustar el modelo. Se realiza un ARIMA $(0, 1, 1), (0, 1, 1)_{12}$ según las barras que se veían anteriormente. Los resultados se ven en la figura 28.

Estimación por mínimos cuadrados condicional							
Parámetro	Estimador	Error estándar	Valor t	Aprox Pr > t	Retardo	Variable	Desplazamiento
MA1,1	0.55039	0.06699	8.22	<.0001	1	LCapital_milesEuros	0
MA2,1	0.78963	0.05202	15.18	<.0001	12	LCapital_milesEuros	0
NUM1	0.72169	0.13755	5.25	<.0001	0	AO102008	0
NUM2	-0.49969	0.13069	-3.82	0.0002	0	LS042011	0
NUM3	0.52827	0.14143	3.74	0.0003	0	AO042014	0

Varianza Estimación	0.027142
Error Std Estimación	0.164748
AIC	-123.466
SBC	-107.876
Número de residuales	167

* AIC y SBC no incluyen determinante de la log.

Correlaciones de las estimaciones de parámetro						
Variable Parámetro	LCapital_milesEuros MA1,1	LCapital_milesEuros MA2,1	AO102008 NUM1	LS042011 NUM2	AO042014 NUM3	
LCapital_milesEuros MA1,1	1.000	-0.073	-0.048	0.027	0.181	
LCapital_milesEuros MA2,1	-0.073	1.000	0.035	-0.032	-0.020	
AO102008 NUM1	-0.048	0.035	1.000	-0.004	-0.010	
LS042011 NUM2	0.027	-0.032	-0.004	1.000	0.042	
AO042014 NUM3	0.181	-0.020	-0.010	0.042	1.000	

Comprobación de autocorrelación de residuales									
Para retardo	Chi-cuadrado	DF	Pr > ChiSq	Autocorrelaciones					
6	5.58	4	0.2327	-0.021	0.000	-0.012	0.062	0.096	0.135
12	10.12	10	0.4303	-0.106	0.031	0.071	-0.005	0.074	0.052
18	20.84	16	0.1849	0.066	-0.002	0.123	0.109	0.158	-0.034
24	34.66	22	0.0420	0.057	0.059	0.196	-0.092	0.072	-0.111
30	37.07	28	0.1174	0.054	0.055	-0.054	0.039	0.037	-0.018

Figura 28: Resultados del ARIMA ajustado

Se comprueba que los residuos están incorrelados prácticamente en todos los retardos y que los p-valores de los parámetros del ARIMA y los datos atípicos incorporados, son menores que 0,05, por lo que son significativamente distintos de cero. El modelo sería:

$$(1 - L^{12})(1 - L)z_t = (1 - 0,78963L^{12})(1 - 0,55039L)\varepsilon_t$$

$$\Rightarrow z_t = z_{t-1} + z_{t-12} - z_{t-13} + \varepsilon_t - 0,78963\varepsilon_{t-12} - 0,55039\varepsilon_{t-1} + 0,43460\varepsilon_{t-13} \quad (3.1)$$

Hay que tener en cuenta añadir los parámetros de los datos atípicos, siendo entonces:

$$z'_t = 0,72169I_t^{(h)} + 0,52827I_t^{(h_2)} - 0,49969S_t^{(h_3)} + z_t$$

donde :

z_t es la expresión 3.1,

$I_t^{(h)} = 0$ para $t \neq h$, e $I_h^{(h)} = 1$ siendo h el instante temporal de octubre 2008,

$I_t^{(h_2)} = 0$ para $t \neq h_2$, e $I_{h_2}^{(h_2)} = 1$ cuando h_2 es abril de 2014

y $S_t^{(h_3)}$ es la variable escalón, con $S_t^{(h_3)} = 1$ si $t \geq h_3$ y 0 en otro caso, siendo h_3 abril de 2011.

Las predicciones se obtienen para la serie transformada en logaritmo. Por ello, hay que realizar una conversión por medio del exponente. La predicción de este modelo se observa en la figura 29, y en la figura siguiente, el gráfico.

Fecha	FORECAST	FORECAST_T
01-2018	14.2062415	1478060.7757
02-2018	14.298887622	1621541.2148
03-2018	14.145103002	1390401.3471
04-2018	14.143967118	1388822.9086
05-2018	14.232904787	1518000.8354
06-2018	14.229331396	1512586.1049
07-2018	14.186097627	1448584.7837
08-2018	14.197566732	1465294.3943
09-2018	14.369515609	1740208.7037
10-2018	14.212473301	1487300.5167
11-2018	14.107161043	1338635.0629
12-2018	14.011598486	1216633.8768

Figura 29: Predicciones del ARIMA ajustado incluyendo outliers

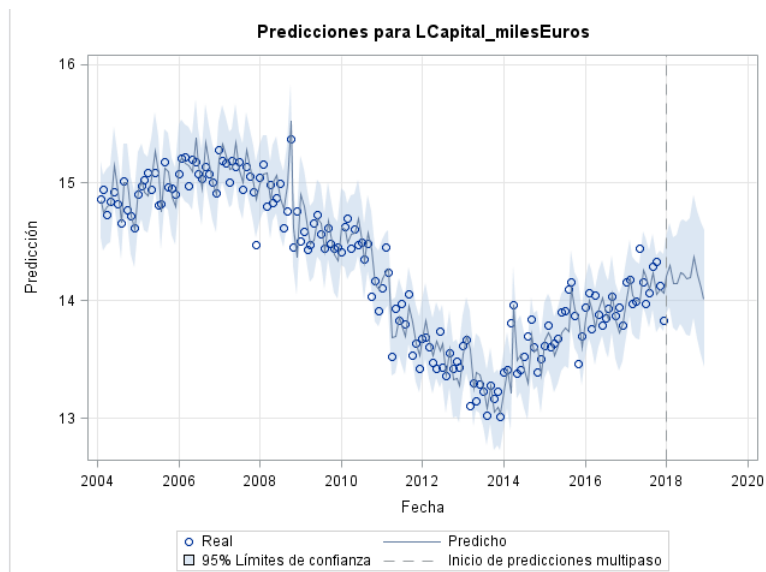


Figura 30: Gráfico de las predicciones del ARIMA ajustado incluyendo outliers

A la entrega de este trabajo, se conocen los valores que ha tenido la serie entre enero y mayo de 2018. Haciendo una comparativa de las predicciones con Suavizado y ARIMA, se obtiene el siguiente gráfico de la figura 31.

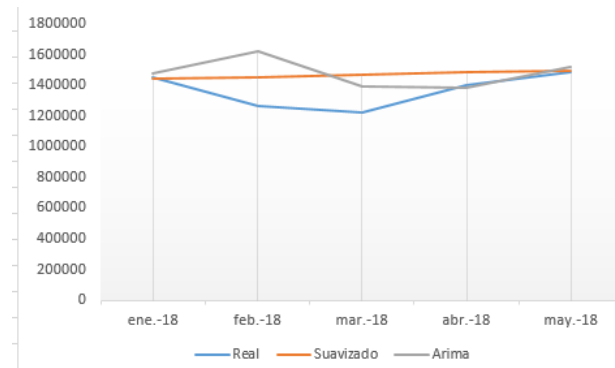


Figura 31: Comparativa de las predicciones y la realidad

Se observa que en febrero el dato del ARIMA ha sido muy poco acertado. sin embargo, el resto de meses se ha acercado bastante a la realidad. El dato de febrero llama la atención, pues tal y como se veía en la figura 20, es el mes con la media más alta de capital hipotecado. Por tanto, en futuros estudios habría que valorar si este dato se considera outlier.

3.1.3. Viviendas Libres Terminadas en la Comunidad de Madrid

En primer lugar, se representa la serie para hacer una valoración visual de la tendencia y posible estacionalidad, mediante la figura 32. Se puede observar una alta variabilidad en los

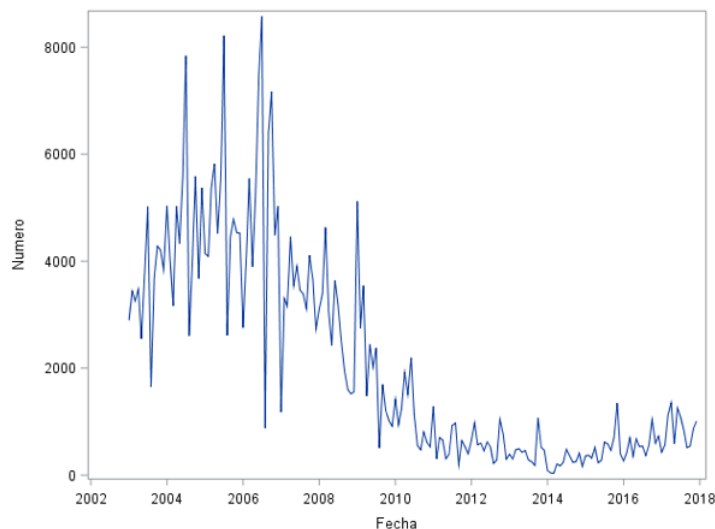


Figura 32: Representación gráfica de la serie del total de viviendas libres construidas en C.Madrid

datos, una varianza que habrá que tener en cuenta a la hora de ajustar modelos ARIMA. La tendencia, sigue el patrón visto hasta el momento: positiva hasta 2008, es decir, hasta el

inicio de la crisis que desembocó en el descenso del hasta entonces, motor de la economía: la construcción. Se observa también que en agosto los números suelen decaer, y en julio, aumentar. Fijémonos en los estadísticos para valorar este hecho, mediante las figuras 33 y 34, donde se observa una mayor estacionalidad que en el resto de las series pues el componente estacional presenta mayores desviaciones que en casos anteriores.

Estadísticos estacionales para la variable Numero						
Índice estacional	N	Mínimo	Máximo	Suma	Media	Desviación estándar
1	15	83.0000	5121.000	29116.00	1941.067	1769.9785
2	15	41.0000	4096.000	29030.00	1935.333	1646.8817
3	15	36.0000	5548.000	33791.00	2252.733	1930.7772
4	15	210.0000	5821.000	33350.00	2223.333	1905.9304
5	15	171.0000	5320.000	29499.00	1966.600	1772.9710
6	15	240.0000	7452.000	37909.00	2527.267	2333.6371
7	15	285.0000	8582.000	44275.00	2951.667	3040.5929
8	15	220.0000	3386.000	18280.00	1218.667	1049.8443
9	15	183.0000	6403.000	28222.00	1881.467	1981.6571
10	15	252.0000	7171.000	34850.00	2323.333	2220.9297
11	15	413.0000	4532.000	28751.00	1916.733	1650.9955
12	15	163.0000	5372.000	27962.00	1864.133	1897.8893

Figura 33: Principales estadísticos sobre la serie

Descomposición estacional para la variable Numero						
Obs	Tiempo	Estación	Ciclo de tendencia	Estacional	Irregular	Ajustado estacionalmente
1	JAN2003	1	.	0.942106	.	3071.843
2	FEB2003	2	.	0.843994	.	4079.414
3	MAR2003	3	.	1.029751	.	3162.900
4	APR2003	4	.	1.072481	.	3240.151
5	MAY2003	5	.	0.892986	.	2857.826
6	JUN2003	6	.	1.136734	.	3321.799
7	JUL2003	7	3597.000	1.303199	1.070482	3850.525
8	AUG2003	8	3709.958	0.730412	0.608902	2259.000
9	SEP2003	9	3729.875	0.802989	1.235707	4609.032
10	OCT2003	10	3790.750	1.297766	0.870616	3300.286
11	NOV2003	11	3929.667	1.099602	0.973370	3825.022
12	DEC2003	12	4080.250	0.847982	1.109545	4527.221

Figura 34: Tabla de la descomposición estacional

Se procede a realizar un modelado de Winters, pero aditivo, ya que en las figuras se veía que podía haber estacionalidad, y de haberla, no aumenta con el tiempo.

Método de Winters (Aditivo) Estimador del parámetro				
Parámetro	Estimador	Error estándar	Valor t	Aprox Pr > t
Peso de nivel	0.03091	0.01436	2.15	0.0327
Peso de tendencia	0.42204	0.20285	2.08	0.0389
Peso estacional	0.21355	0.03160	6.76	<.0001

Figura 35: Suavizado de Winters Aditivo

Se observa que esta vez los p-valores son menores de 0,05 en todos los parámetros, por lo que se acepta que son significativamente distintos de cero. En este caso, el valor de α es 0,031,

β es 0,42 y γ tiene un valor de 0,21. Las predicciones de este modelo vienen dadas por la figura 36.

Predicciones para la variable Numero					
Obs	Tiempo	Predicciones	Error estándar	Límites de confianza al 95%	
181	JAN2018	944.7356	891.4140	-802.4037	2691.8750
182	FEB2018	953.5852	892.2749	-795.2414	2702.4118
183	MAR2018	1232.2960	893.7206	-519.3642	2983.9563
184	APR2018	1243.3930	895.8995	-512.5377	2999.3237
185	MAY2018	1037.0318	898.9565	-724.8906	2798.9542
186	JUN2018	1351.6782	903.0326	-418.2332	3121.5895
187	JUL2018	1485.9938	908.2629	-294.1688	3266.1565
188	AUG2018	926.7708	914.7756	-866.1564	2719.6980
189	SEP2018	1030.5901	922.6900	-777.8490	2839.0292
190	OCT2018	1378.1259	932.1155	-448.7869	3205.0387
191	NOV2018	1316.4294	943.1502	-532.1111	3164.9699
192	DEC2018	1194.4208	955.8800	-679.0696	3067.9111

Figura 36: Predicciones de Winters Aditivo

A continuación, y como se ha hecho antes, se va a ajustar el ARIMA. En primer lugar se observa que la varianza es muy poco constante, por lo que convendría transformarla logarítmicamente. Asimismo, se observa que no es estable en media, y como se vio antes, se puede suponer estacional. Por tanto, se realizan paso por paso las transformaciones, siguiendo la metodología de Box-Jenkins y se grafican los correlogramas para esta serie una vez realizadas las tres transformaciones dichas (figura 37).

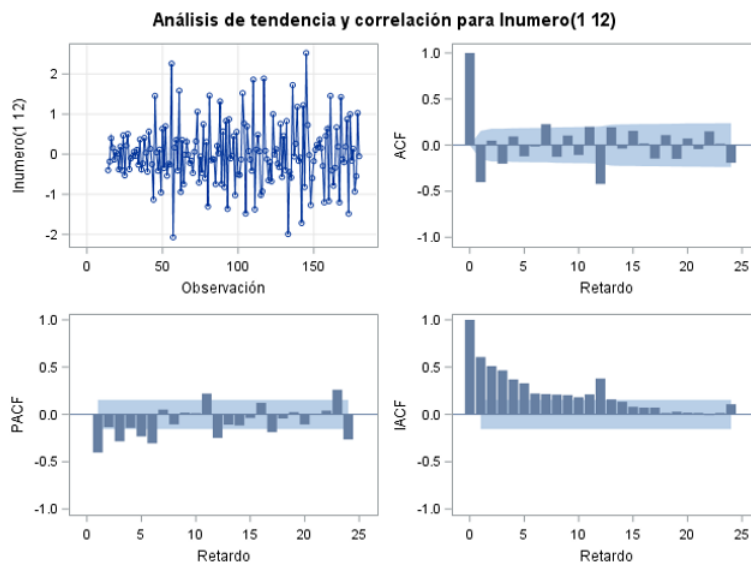


Figura 37: Funciones de autocorrelación simple parcial en la serie transformada

En ella, se observan barras significativas en la ACF y al menos una en PACF. Probando

varios modelos, se comprueba que el que mejor resultados aporta en cuanto a los contrastes es el ARIMA $(0, 1, 1), (0, 1, 1)_{12}$, que se observa sus resultados en la figura 38. Todos los residuos están incorrelados. El modelo sería: $(1 - L^{12})(1 - L)z_t = (1 - 0,84481L^{12})(1 - 0,66004L)\varepsilon_t$
 $\Rightarrow z_t = z_{t-1} + z_{t-12} - z_{t-13} + \varepsilon_t - 0,66004\varepsilon_{t-1} - 0,84481\varepsilon_{t-12} + 0,55761\varepsilon_{t-13}$

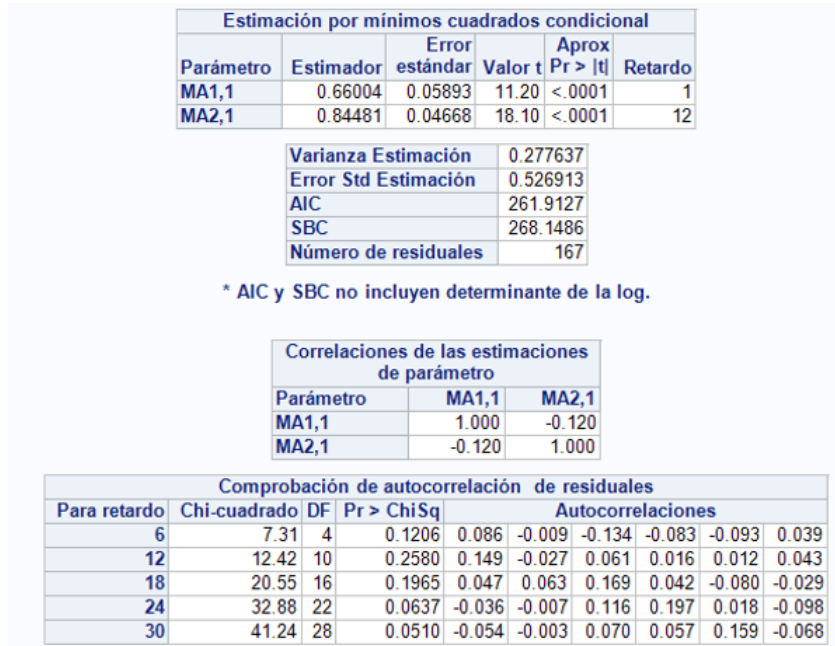


Figura 38: Resultados de ARIMA ajustado

Sus predicciones, transformando el logaritmo, y el gráfico, se ven en las siguientes figuras

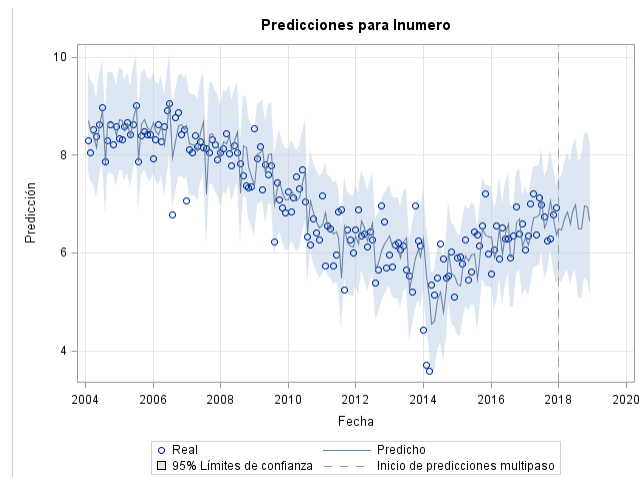


Figura 39: Gráfico de las predicciones ARIMA ajustado

Fecha	FORECAST	FORECAST_T
01-2018	6.4964151954	662.76149894
02-2018	6.4755209005	649.05723383
03-2018	6.6963936492	809.48128142
04-2018	6.8496315911	943.53323663
05-2018	6.5893985702	727.34329241
06-2018	6.8468155157	940.87991361
07-2018	6.9925482285	1088.4916706
08-2018	6.4934711575	660.81317341
09-2018	6.4847331681	655.0641489
10-2018	6.9552473684	1048.6379058
11-2018	6.9375622497	1030.2556452
12-2018	6.6598224673	780.41237613

Figura 40: Predicciones ARIMA ajustado

Comparando con los datos reales, que tenemos de enero a marzo de 2018, se comprueba que funciona mejor el suavizado que el ARIMA, en la figura 41.

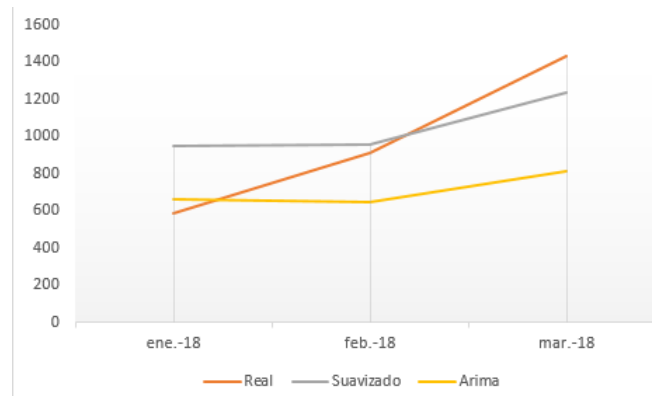


Figura 41: Comparativa de modelos con la realidad

3.1.4. Transmisiones de Compraventa de Inmuebles en la Comunidad de Madrid

Siguiendo la metodología empleada en los análisis anteriores, se empieza por representar la serie para hacer un primer análisis visual de la misma, tanto de manera completa, como en un intervalo menor para valorar la estacionalidad.

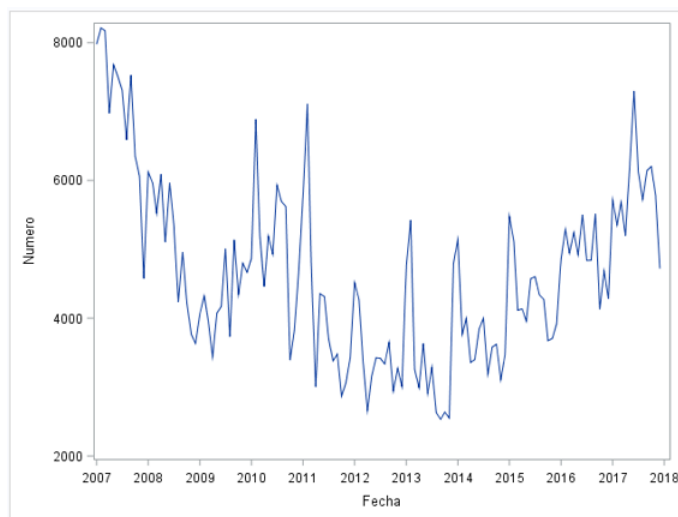


Figura 42: Representación gráfica de la serie de transmisiones de compraventa de inmuebles en C.Madrid

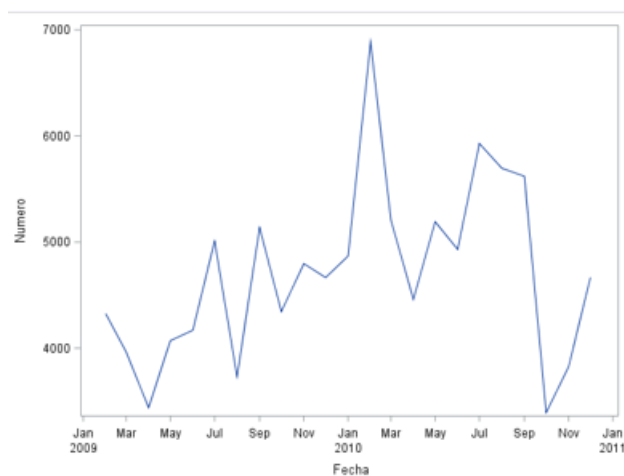


Figura 43: Representación gráfica de la serie entre enero de 2009 y enero de 2011

En las figuras 42 y 43 se observa que no hay una tendencia clara. A diferencia de las series anteriores, ésta presenta datos desde 2007, y se puede ver que ya al principio de la serie hay cierto descenso, para a mitad de la serie, pasar a tener un nivel mayor. Tampoco en el intervalo tomado en la representación de 2 años hay una pauta clara de descenso o auge en un mes concreto. Fijándonos en la descomposición estacional, en la tabla de la figura 44 sí parece que hay una mayor media de compraventas en enero, y los números de estacionalidad sí presentan variaciones a principios de año.

Estadísticos estacionales para la variable Numero						
Indice estacional	N	Mínimo	Máximo	Suma	Media	Desviación estándar
1	11	4067.000	7978.000	59264.00	5387.636	1050.9640
2	11	3765.000	8214.000	61717.00	5610.636	1344.5434
3	11	3259.000	8174.000	52975.00	4815.909	1382.3915
4	11	2660.000	6973.000	47552.00	4322.909	1407.1867
5	11	3159.000	7678.000	51674.00	4697.636	1335.6412
6	11	2906.000	7504.000	54437.00	4948.818	1491.0836
7	11	3288.000	7309.000	53560.00	4869.091	1252.8803
8	11	2627.000	6590.000	47686.00	4335.091	1250.6558
9	11	2532.000	7528.000	52414.00	4764.909	1433.3787
10	11	2637.000	6347.000	44382.00	4034.727	1240.9560
11	11	2551.000	6058.000	44605.00	4055.000	1136.0703
12	11	3007.000	4788.000	45166.00	4106.000	636.5306

Figura 44: Principales estadísticos sobre la serie

Descomposición estacional para la variable Numero						
Obs	Tiempo	Estación	Ciclo de tendencia	Estacional	Irregular	Ajustado estacionalmente
1	JAN2007	1	.	1.180079	.	6760.567
2	FEB2007	2	.	1.221528	.	6724.363
3	MAR2007	3	.	1.014015	.	8061.022
4	APR2007	4	.	0.909885	.	7663.609
5	MAY2007	5	.	0.994742	.	7718.581
6	JUN2007	6	.	1.050739	.	7141.643
7	JUL2007	7	7000.208	1.044535	0.999595	6997.375
8	AUG2007	8	6829.167	0.923122	1.045342	7138.816
9	SEP2007	9	6625.042	1.016189	1.118193	7408.072
10	OCT2007	10	6477.833	0.850247	1.152375	7464.891
11	NOV2007	11	6333.875	0.866122	1.104284	6994.398
12	DEC2007	12	6162.500	0.928798	0.799655	4927.875

Figura 45: Tabla de la descomposición estacional

Haciendo el modelado de suavizados, mediante el modelo de Winters, arroja unos resultados que indican que no hay tendencia, pero sí estacionalidad. Ninguno de los métodos de suavizado estudiados daría resultados satisfactorios, por lo que no se modelará de esta forma.

Método de Winters (Multiplicativo) Estimador del parámetro				
Parámetro	Estimador	Error estándar	Valor t	Aprox Pr > t
Peso de nivel	0.50038	0.04865	10.29	<.0001
Peso de tendencia	0.0010000	0.0072997	0.14	0.8913
Peso estacional	0.13551	0.04175	3.25	0.0015

Figura 46: Suavizado de Winters Multiplicativo

Se procede con ARIMA. Dado que la varianza en este caso no es tan dispar como en anteriores ocasiones, no se harán logaritmos. Sin embargo, para estabilizar la media, se tomarán diferencias. Tras probar varios modelos, se obtiene como modelo que mejores resultados da el ARIMA $(0, 1, 1, (0, 1, 1)_{12})$ tomando una diferencia normal y otra estacional.

Estimación por mínimos cuadrados condicional					
Parámetro	Estimador	Error estándar	Valor t	Aprox Pr > t	Retardo
MA1,1	0.53589	0.07900	6.78	<.0001	1
MA2,1	0.56355	0.08349	6.75	<.0001	12

Varianza Estimación	503156.9
Error Std Estimación	709.3356
AIC	1902.001
SBC	1907.559
Número de residuales	119

* AIC y SBC no incluyen determinante de la log.

Correlaciones de las estimaciones de parámetro		
Parámetro	MA1,1	MA2,1
MA1,1	1.000	-0.068
MA2,1	-0.068	1.000

Comprobación de autocorrelación de residuales									
Para retardo	Chi-cuadrado	DF	Pr > ChiSq	Autocorrelaciones					
6	5.26	4	0.2618	0.043	-0.082	0.015	-0.103	0.115	0.096
12	8.80	10	0.5513	0.009	-0.104	-0.051	0.111	0.026	0.020
18	11.48	16	0.7792	0.073	0.046	0.005	-0.059	-0.049	-0.077
24	18.69	22	0.6641	-0.060	0.076	-0.145	-0.118	0.065	0.020

Figura 47: Resultados de ARIMA ajustado

Para este caso, el modelo resultante es: $(1 - L^{12})(1 - L)z_t = (1 - 0,56355L^{12})(1 - 0,53589L)\varepsilon_t$
 $\Rightarrow z_t = z_{t-1} + z_{t-12} - z_{t-13} + \varepsilon_t - 0,53589\varepsilon_{t-1} - 0,56355\varepsilon_{t-12} + 0,30200\varepsilon_{t-13}$

Sus predicciones:

Fecha	FORECAST
01-2018	6681.8334733
02-2018	6531.1181215
03-2018	6303.6125927
04-2018	6053.9243247
05-2018	6457.3719686
06-2018	7174.8188309
07-2018	6532.9829315
08-2018	6210.2312615
09-2018	6597.573925
10-2018	6153.0285709
11-2018	6074.9904959
12-2018	5675.5021577

Figura 48: Predicciones de ARIMA ajustado

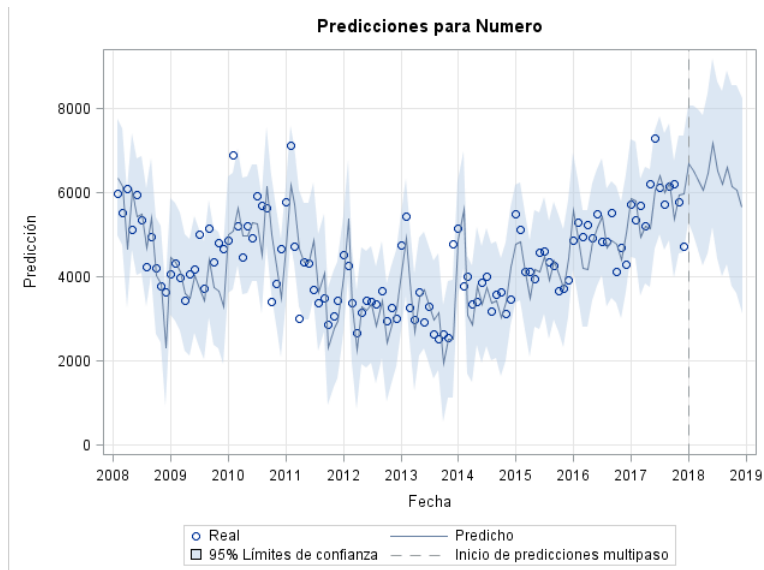


Figura 49: Gráfico de las predicciones de ARIMA ajustado

Haciendo la comparativa con los datos reales, que los tenemos hasta junio de 2018, se obtiene la figura 50. Son datos que pueden tomarse por adecuados.

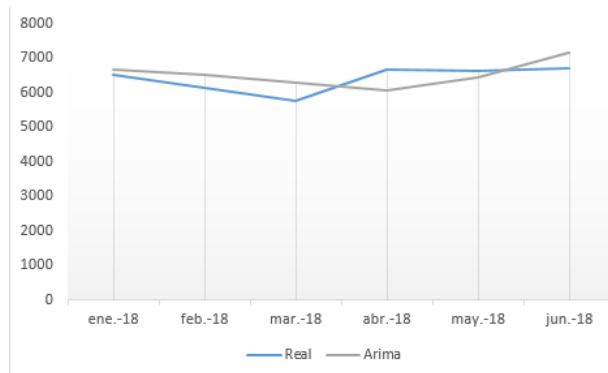


Figura 50: Comparativa Realidad - ARIMA

3.1.5. Valor Tasado Medio por M2 en la Comunidad de Madrid

Esta serie es la que se va a considerar la serie principal en este Trabajo. Se trata de datos trimestrales que tienen la representación que se ve en la figura 51.

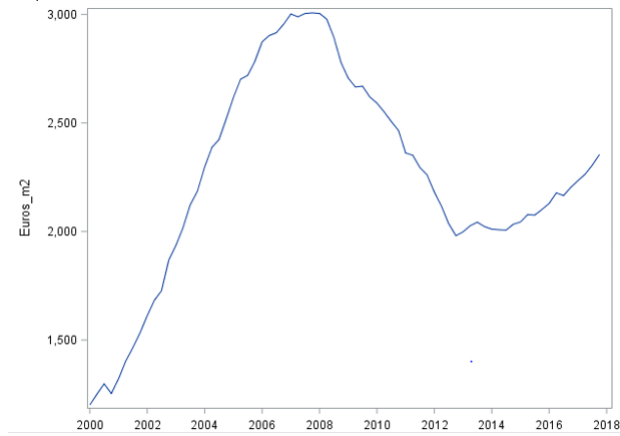


Figura 51: Representación gráfica de la serie de valor tasado medio del metro cuadrado en los inmuebles de la Comunidad de Madrid.

Al tratarse de datos trimestrales, se observa que la serie presenta un trazo más suave, sin las variaciones que presentaban las series anteriores. Una vez más, se puede comprobar la tendencia alcista hasta 2008, y la caída abrupta hasta 2014, que comienza a subir de manera lenta. En esta representación, cabe pensar que no hay estacionalidad. Se procede a modelar por medio de suavizados. Comenzando con el modelo de Winters, se confirma que la estacionalidad no parece presente en la serie, ya que da un p-valor menor de 0,05 en la figura 52.

Método de Winters (Multiplicativo) Estimador del parámetro				
Parámetro	Estimador	Error estándar	Valor t	Aprox Pr > t
Peso de nivel	0.88102	0.06050	14.56	<.0001
Peso de tendencia	0.45552	0.10269	4.44	<.0001
Peso estacional	0.99900	0.57790	1.73	0.0883

Figura 52: Suavizado de Winters Multiplicativo

Haciendo un suavizado doble de Holt, se comprueba en la figura 53 que esta vez los p-valores son menores de 0,05. Los parámetros α y β que se estudiaban en la teoría, toman unos valores de 0,999 y 0,41 respectivamente. Nótese el valor de α , que es prácticamente de 1 su valor, lo que indica que según este método, los valores alejados en el tiempo prácticamente no tienen influencia en la predicción.

Suavizado exponencial lineal Estimador del parámetro				
Parámetro	Estimador	Error estándar	Valor t	Aprox Pr > t
Peso de nivel	0.99900	0.08392	11.90	<.0001
Peso de tendencia	0.40894	0.09692	4.22	<.0001

Figura 53: Alisado doble de Holt

Las predicciones con este modelo de alisado doble de Holt, son las de la figura 54

Predicciones para la variable Euros_m2					
Obs	Tiempo	Predicciones	Error estándar	Límites de confianza al 95%	
73	2018:1	2394.9508	35.2550	2325.8523	2464.0493
74	2018:2	2435.0179	60.8712	2315.7126	2554.3232
75	2018:3	2475.0850	88.3431	2301.9357	2648.2343
76	2018:4	2515.1521	118.1330	2283.6156	2746.6886

Figura 54: Predicciones de Alisado doble de Holt

A continuación se realiza el modelo ARIMA. Se representan los correlogramas sin hacer ningún ajuste, mediante la figura 55.

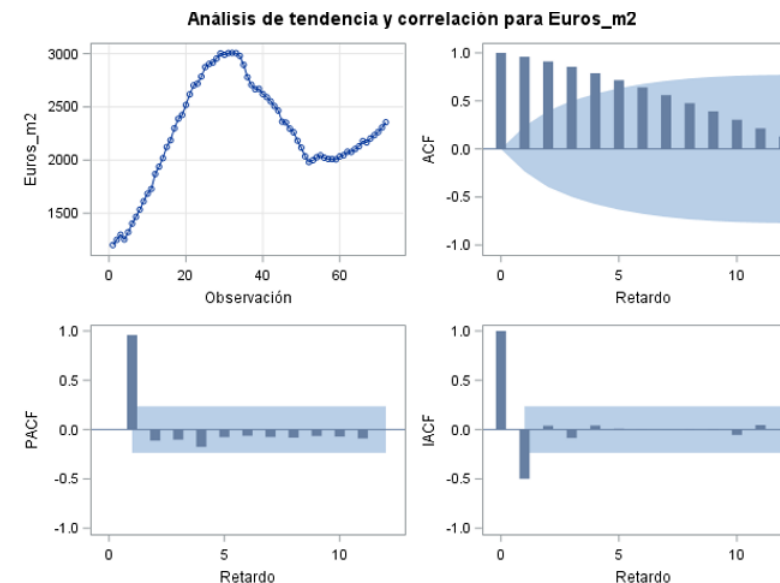


Figura 55: Representación de los correlogramas

No parece necesario aplicar diferencias estacionales ni hacer transformación logarítmica, por lo que se comienza haciendo una sola diferencia regular, para estabilizar la media.

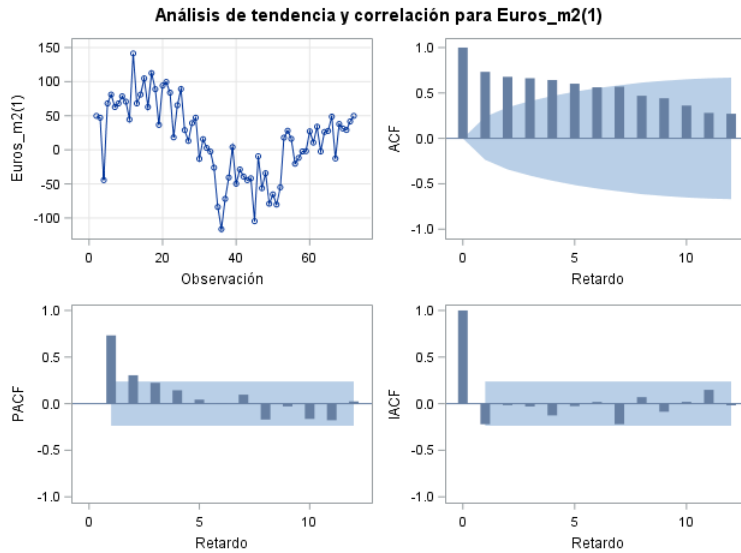


Figura 56: Representación de los correlogramas con una diferencia regular

Parece que el modelo aún no es estable del todo en media, pero como por el principio de parsimonia se requiere empezar por modelos sencillos para no caer en el sobreajuste, y viendo que en PACF hay una barra significativa que hace suponer que haya que ajustar de manera autorregresiva, se ajusta el modelo ARIMA(1,1,0), cuyos resultados se ven en la figura 57.

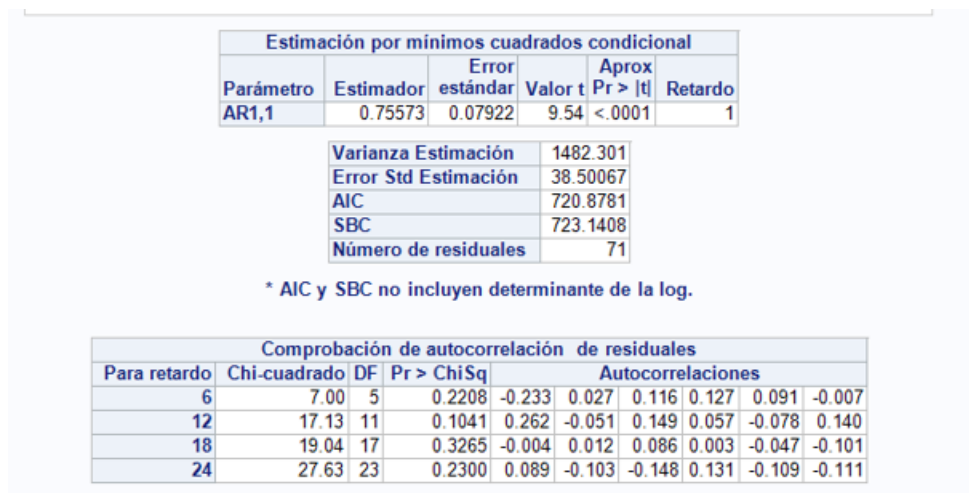


Figura 57: Resultados del ARIMA ajustado

Se ve que los residuos están incorrelados, pues el p-valor es mayor a 0,05 y el parámetro AR toma un valor de 0,75573. El modelo resultante sería entonces:

$$(1 - 0,75573L)(1 - L)z_t = \varepsilon_t \Rightarrow z_t = 1,75573z_{t-1} - 0,75573z_{t-2} + \varepsilon_t$$

Las predicciones vendrían dadas por:

Fecha	FORECAST
2018Q1	2392.4599324
2018Q2	2420.8452144
2018Q3	2442.2969101
2018Q4	2458.5086654
2019Q1	2470.7604246
2019Q2	2480.0194839
2019Q3	2487.0168609
2019Q4	2492.30501

Figura 58: Predicciones del ARIMA ajustado

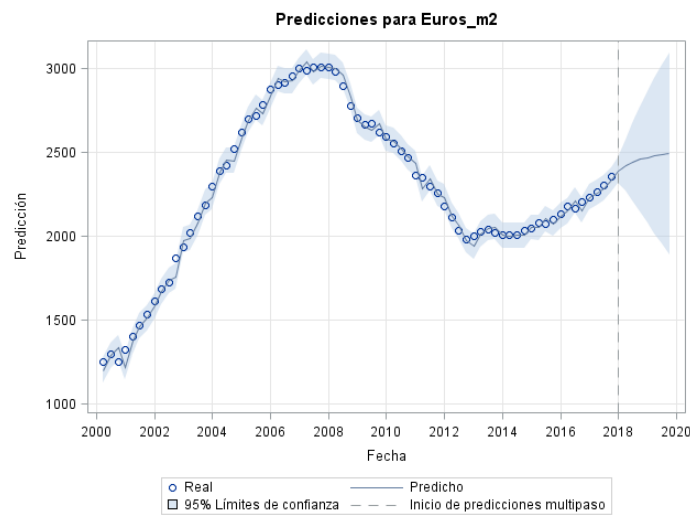


Figura 59: Gráfico de las predicciones del ARIMA ajustado

Estas predicciones se compararán después con el dato real, y con el resto de métodos.

ARIMA con R

Esta serie se va a modelizar también mediante ARIMA con los paquetes que proporciona el software R.

En primer lugar, y tras importar los datos con clase *ts*, se puede utilizar la función *ndiffs* del paquete *forecast* que propone las diferencias necesarias para transformar la serie en estacionaria. De esta manera, R sugiere que se tomen 2 diferencias regulares y ninguna estacional:

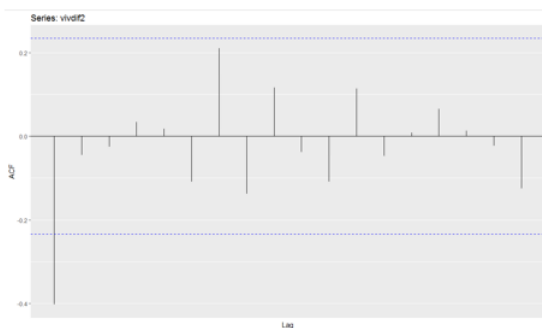
```
f <- ndiffs(viv.ts) -> para diferencias regulares
```

```
s <- nsdiffs(viv.ts) -> para estacionales
```

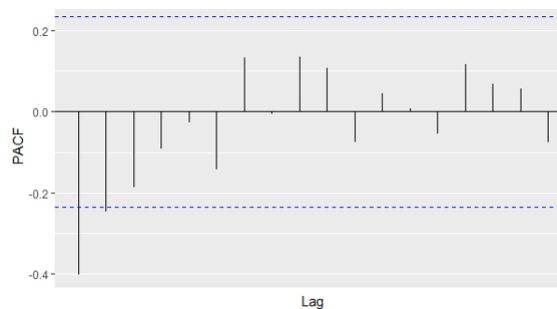
values	
f	2
s	0

Figura 60: Diferencias propuestas por R

De hecho, si se diferencia la serie una sola vez y se realiza el Test de Dickey Fuller Aumentado, el cual tiene como hipótesis nula que la serie **no** es estacionaria, para un p-valor de 0,05 el resultado es de 0.9396736, es decir, no se puede rechazar la hipótesis. Sin embargo, diferenciando dos veces, el valor del test es de 0.01, por lo que ahora sí se rechaza la hipótesis nula: la serie es estacionaria. Ya se había visto cuando se ajustaba en SAS que la representación de los correlogramas admitían una segunda diferenciación, por lo que se va a tener en cuenta en esta modelización con R. Los correlogramas asociados a esta serie diferenciada están en las figuras 61a y 61b.



(a) ACF



(b) PACF

Figura 61: ACF y PACF de la serie diferenciada dos veces

Se observa que en ACF hay una barra significativa, y en PACF, también. Empleando la función *autoarima* de R, del paquete *forecast*, aparece el modelo ARIMA que sugiere R. Este modelo se obtiene según el criterio AIC:

```

> autoarima <- auto.arima(viv.ts)
> autoarima
Series: viv.ts
ARIMA(0,2,1)

Coefficients:
      ma1
      -0.5840
s.e.    0.0929

sigma^2 estimated as 1261: log likelihood=-348.92
AIC=701.84  AICC=702.02  BIC=706.33

```

Figura 62: ARIMA propuesto con la función *autoarima*

El modelo que propone es el ARIMA (0,2,1), es decir, es coherente con el resultado de diferenciar dos veces, y ajusta la parte de medias móviles con un parámetro de -0,5840. Este modelo tiene un AIC de 701,84. No obstante se van a probar varios modelos, entre ellos el que se puso mediante SAS, para comparar los AIC:

```

arima1<- arima(viv.ts , c(1,1,0)) #serie original ,una diferencia y ar1
arima2<- arima(vivdif , c(1,0,1)) #serie diferenciada y ar1 ma1
arima3<- arima(vivdif , c(1,0,0)) #serie diferenciada y ar1 (SAS)
arima4<- arima(vivdif , c(0,0,1)) #serie diferenciada y ma1
arima5<- arima(vivdif2 , c(0,0,1)) #serie diferenciada 2 veces y ma1
arima6<- arima(vivdif2 , c(1,0,1)) #serie diferenciada 2 veces ,ar1 ,ma1
autoarima<- auto.arima(viv.ts) #modelo propuesto de autoarima

```

El AIC de cada uno de estos modelos es el siguiente:

```

arima1: 722.7514;          arima2: 713.9971;          arima3: 723.4929;
arima4: 748.7482;          arima5: 703.8360;          arima6: 705.7769;
autoarima: 701.8360

```

Por lo tanto, efectivamente el modelo autoarima da un AIC menor. Se representan los correlogramas de los residuos de este modelo en la figura 63 .

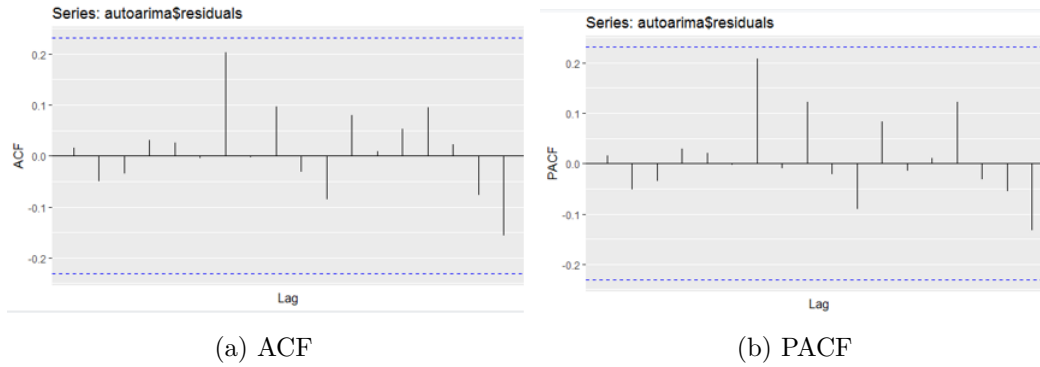


Figura 63: Función ACF y PACF de los residuos

Se observa que están incorrelados. Haciendo el Test de contraste de Ljung-Box con R mediante la función `Box.test`, cuya hipótesis nula es que los residuos son ruido blanco (y por tanto, incorrelados), se acepta este hecho ya que da un valor de 0.888834. El modelo sería: $(1 - L)^2 z_t = (1 + 0,584L)\varepsilon_t \Rightarrow z_t = z_{t-2} + \varepsilon_t + 0,584\varepsilon_{t-1}$.

Por último, como se ve en las siguientes imágenes, se realizan predicciones con este modelo, que para 2018 tomaría los valores:

Q1: 2395,168; Q2: 2435,436; Q3: 2475,703; Q4:2515,971

Esta será una de las predicciones que se tendrá en cuenta en la comparación con los valores reales.

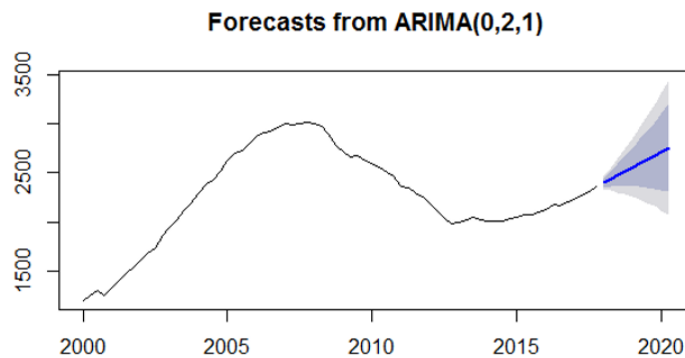


Figura 64: Gráfico de la predicción del ARIMA modelado

	↑ forec.mean ↓	X80. ↓	X95. ↓	X80..1 ↓	X95..1 ↓
1	2395.168	2349.659	2325.568	2440.677	2464.767
2	2435.436	2356.545	2314.783	2514.326	2556.088
3	2475.703	2360.922	2300.161	2590.484	2651.246
4	2515.971	2362.215	2280.821	2669.727	2751.121
5	2556.239	2360.434	2256.782	2752.043	2855.696
6	2596.507	2355.702	2228.228	2837.311	2964.785
7	2636.774	2348.157	2195.373	2925.391	3078.176
8	2677.042	2337.932	2158.419	3016.152	3195.665
9	2717.310	2325.148	2117.551	3109.471	3317.069
10	2757.578	2309.915	2072.936	3205.240	3442.219

Figura 65: Predicciones del ARIMA modelado

3.2. Predicción avanzada: Redes Neuronales

Para modelizar mediante redes neuronales la serie *Valor de tasación medio por metro cuadrado en la Comunidad de Madrid*, se empleará el software RStudio.

3.2.1. Red Neuronal Autorregresiva

En primer lugar se tratará de modelizar una red neuronal autorregresiva mediante la función *nnetar* del paquete *forecast*. Este paquete usa redes neuronales feed-forward con una capa oculta y los retardos de la serie univariante como inputs de la misma. Dado que en esta serie no se consideraba la existencia de estacionalidad, se programa la red sin este hecho, y se realizan 100 repeticiones como número de redes a ajustar, con los diferentes pesos iniciales aleatorios. Cuando se realizan las predicciones, estos pesos se promedian. La función de activación empleada es la que usa R por defecto, esto es, la logística, y el número de nodos de la capa oculta viene determinado por el parámetro *size*, que por defecto toma el valor redondeado de $k = (p+P+1)/2$, donde p es el número de retardos no estacionales utilizados como inputs, que en el caso de series temporales no estacionales será el número óptimo de retardos (según el AIC) para un modelo lineal AR (p). Por tanto, el valor de *size* será por defecto $(p + 1)/2$. El modelo que resulta es:

```

Series: viv.ts
Model: NNAR(1,1)
Call: nnetar(y = viv.ts, P = 0, repeats = 100, act.fct = "logistic")

Average of 100 networks, each of which is
a 1-1-1 network with 4 weights
options were - linear output units

sigma^2 estimated as 2774.932

```

Figura 66: Modelo Red Neuronal Autorregresiva

Una red NNAR(1, 1) tendrá un nodo en la capa oculta, y un retardo como input, ya que se ha considerado que 1 era lo óptimo. Es decir, tendrá 4 parámetros, pues el número de parámetros viene determinado por: $h(k + 1) + h + 1$, donde h es 1, pues hay un nodo en la capa oculta, y k es 1, pues se trata del número de nodos inputs.

La predicción que da esta red viene dada por la figura 67 y 68

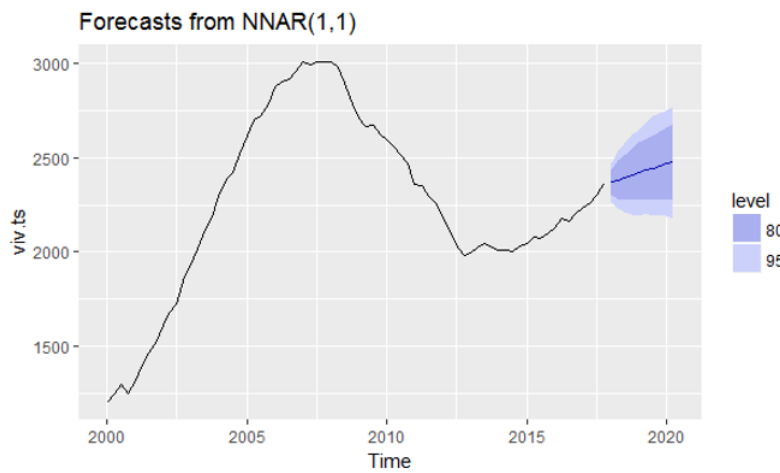


Figura 67: Gráfico de la predicción de la red NNAR(1,1)

	Point Forecast	Lo 80	Hi 80	Lo 95	Hi 95
2018 Q1	2368.583447	2300.983158	2437.663357	2267.143310	2471.139597
2018 Q2	2381.932966	2280.152374	2480.358472	2233.666792	2533.192106
2018 Q3	2394.950918	2276.059816	2510.356221	2214.720720	2575.442144
2018 Q4	2407.639574	2272.447443	2545.517457	2200.165950	2623.999318
2019 Q1	2420.001166	2278.926761	2578.778753	2189.409594	2644.307021
2019 Q2	2432.037943	2274.225985	2592.681353	2202.684928	2675.652013
2019 Q3	2443.752210	2274.222430	2616.309258	2191.241471	2719.460406
2019 Q4	2455.146373	2279.602425	2635.036364	2191.152983	2736.058825
2020 Q1	2466.222970	2276.517300	2658.752581	2189.475837	2749.636467
2020 Q2	2476.984698	2281.210969	2675.417897	2180.216119	2766.316941

Figura 68: Predicción de la red NNAR(1,1)

Se construyen ahora otras redes donde no se utilicen los parámetros por defecto; en concreto,

se modificarán los nodos ocultos y el valor de p.

- Red 2: 5 nodos ocultos y 2 nodos inputs:

```
Series: viv.ts
Model: NNAR(2,5)
Call: nnetar(y = viv.ts, p = 2, P = 0, size = 5, repeats = 100, act.fct = "logistic")

Average of 100 networks, each of which is
a 2-5-1 network with 21 weights
options were - linear output units

sigma^2 estimated as 1124.589
```

Figura 69: Red construida número 2

	Point Forecast	Lo 80	Hi 80	Lo 95	Hi 95
2018 Q1	2391.902810	2351.397242	2434.880342	2328.965543	2456.432616
2018 Q2	2419.484344	2334.960831	2514.861650	2283.535368	2554.907587
2018 Q3	2440.081131	2305.460948	2578.963230	2230.961144	2654.474379
2018 Q4	2455.438330	2266.714060	2657.307715	2175.833749	2738.964640
2019 Q1	2466.838615	2230.618305	2712.289483	2114.891466	2832.312237
2019 Q2	2475.242625	2194.587939	2775.641380	2062.085807	2907.478867
2019 Q3	2481.375658	2149.221316	2821.551131	2016.209357	2986.852918
2019 Q4	2485.786334	2131.831806	2868.465290	1996.211328	3033.141133
2020 Q1	2488.889142	2111.019618	2905.070576	1981.667040	3050.590891
2020 Q2	2490.996477	2090.449972	2932.155648	1969.453876	3049.551841

Figura 70: Predicción de la red NNAR(2,5)

- Red 3: 4 nodos ocultos y 1 nodo inputs:

```
Series: viv.ts
Model: NNAR(1,4)
Call: nnetar(y = viv.ts, p = 1, P = 0, size = 4, repeats = 100, act.fct = "tanh")

Average of 100 networks, each of which is
a 1-4-1 network with 13 weights
options were - linear output units

sigma^2 estimated as 2521.22
```

Figura 71: Red construida número 3

	Point Forecast	Lo 80	Hi 80	Lo 95	Hi 95
2018 Q1	2360.435411	2295.769021	2430.605751	2263.946209	2464.619358
2018 Q2	2365.746478	2276.903472	2456.216720	2234.798459	2516.356196
2018 Q3	2370.843694	2259.893871	2475.778792	2203.477708	2552.217771
2018 Q4	2375.737325	2246.175942	2502.709797	2180.560924	2574.499312
2019 Q1	2380.437341	2235.883466	2516.378685	2167.291217	2598.423382
2019 Q2	2384.953367	2227.605796	2540.948596	2160.098153	2619.436715
2019 Q3	2389.294650	2226.617053	2549.781136	2147.573357	2631.782617
2019 Q4	2393.470038	2223.793066	2557.583113	2149.436065	2652.776320
2020 Q1	2397.487971	2212.519748	2572.163003	2130.315036	2668.463910
2020 Q2	2401.356476	2203.924021	2584.664105	2114.728888	2692.705063

Figura 72: Predicción de la red NNAR(1,4)

- Red 4: 6 nodos ocultos y 6 nodos inputs:

```

Model: NNAR(6,6)
Call: nnetar(y = viv.ts, p = 6, P = 0, size = 6, repeats = 100, act.fct = "logistic")

Average of 100 networks, each of which is
a 6-6-1 network with 49 weights
options were - linear output units

sigma^2 estimated as 263.8831

```

Figura 73: Red construida número 4

	Point Forecast	Lo 80	Hi 80	Lo 95	Hi 95
2018 Q1	2418.717726	2398.358850	2441.417238	2387.376510	2452.752649
2018 Q2	2490.353734	2459.506427	2520.490299	2444.915758	2539.610264
2018 Q3	2570.012112	2533.011321	2599.887444	2511.015261	2618.142082
2018 Q4	2645.835566	2608.617736	2675.474804	2581.980390	2689.567393
2019 Q1	2710.646392	2671.942074	2738.706380	2656.262006	2753.017942
2019 Q2	2769.452182	2729.251470	2801.948946	2705.528647	2818.872291
2019 Q3	2825.640630	2786.434229	2854.261355	2767.817448	2873.545412
2019 Q4	2874.771754	2839.222482	2900.862585	2819.751325	2918.457607
2020 Q1	2915.368787	2877.366827	2941.248865	2859.394445	2957.916656
2020 Q2	2950.992600	2911.344636	2976.615935	2892.333624	2992.662248

Figura 74: Predicción de la red NNAR(6,6)

Estos resultados se tendrán en cuenta en la comparación con los valores reales y el resto de métodos empleados.

3.2.2. Red Neuronal ELM

Para este tipo de redes, R proporciona la función *elmdel* paquete del *nnfor*, que pretende ajustar una red ELM (Extreme Learning Machine) a una serie temporal. En primer lugar, se especifica el método *stepwise* como método de estimación de los pesos de la salida, y 100 repeticiones a entrenar, cuyo resultado será un ensamblado de las 100 repeticiones. El resto de parámetros se dejarán por defecto, entre ellos, la función de activación (logística) y el número de retardos que se toman como inputs, cuyo valor por defecto es de 1 a la frecuencia de la serie, en este caso, 4. El número de nodos en la capa oculta, se deja por defecto, de manera que se ajusta automáticamente. El resultado que arroja esta función es:

```

ELM fit with 40 hidden nodes and 100 repetitions.
Series modelled in differences: D1.
univariate lags: (1,2,4)
Forecast combined using the median operator.
Output weight estimation using: step.
MSE: 328.7361.

```

Figura 75: Resultado de la red ELM de la función *nnfor*

Indica que utiliza 3 retardos como inputs, el 1, 2 y 4, y que toma 40 nodos en la capa oculta.

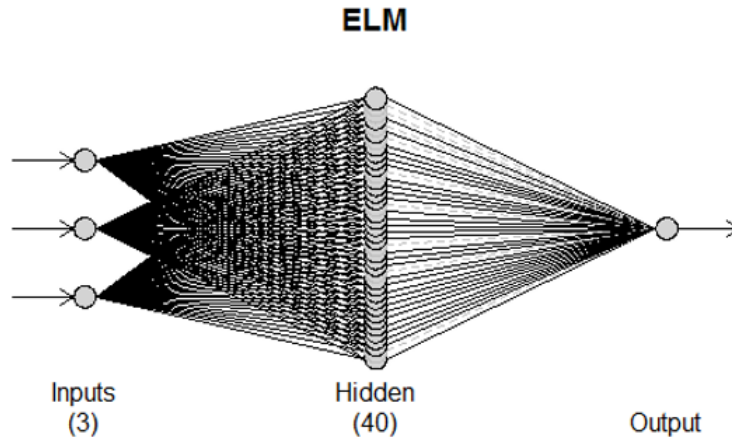


Figura 76: Red ELM propuesta

Las predicciones que da este modelo se ven en las siguientes imágenes 77 y 78.

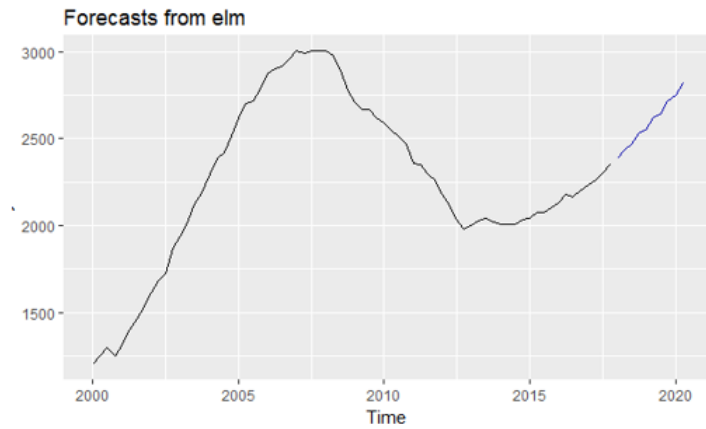


Figura 77: Gráfico de la predicción

	Qtr1	Qtr2	Qtr3	Qtr4
2018	2388.919733	2437.286195	2469.720150	2532.337301
2019	2553.067290	2617.464782	2641.679826	2716.359411
2020	2748.811448	2820.074174		

Figura 78: Predicción del modelo ELM propuesto

Cambiando el método por "LASSO", quedaría el siguiente modelo:

```

ELM fit with 65 hidden nodes and 100 repetitions.
Series modelled in differences: D1.
Univariate lags: (1,2,4)
Forecast combined using the median operator.
Output weight estimation using: lasso.
MSE: 1259.7021.

```

Figura 79: Red ELM 2

Es decir, este modelo se propone con más nodos, en concreto 65. Se observa un mayor Mean Squared Error con este método. Las predicciones en la figura 80

	Qtr1	Qtr2	Qtr3	Qtr4
2018	2389.880415	2420.390840	2448.973136	2477.628652
2019	2503.982920	2528.682407	2552.153342	2574.786842
2020	2596.924231	2618.185468		

Figura 80: Predicción del modelo 2 de red ELM propuesto

3.3. Análisis multivariante: Regresión Dinámica

A continuación, se tratará de hacer predicción con regresión dinámica, utilizando el resto de series como series independientes, y ésta, como serie dependiente. En primer lugar, se deben homogeneizar las series, ya que no todas comenzaban en el mismo año. Dado que la serie más restrictiva es la que comenzaba en 2007, se toman todas desde este momento.

Además, la serie del valor de tasación tenía una frecuencia trimestral, mientras que el resto es mensual. Por ello, se debe emplear el procedimiento *expand* de SAS, que mediante interpolación, convierte los intervalos de frecuencia en lo que se indique. En este caso, se requieren los datos mensuales.

Asimismo, y dado que la predicción del 2018 se tiene que basar en las series que actuarán como independientes, se utilizarán los modelos de ARIMA que se han realizado sobre las series en las secciones anteriores. Por lo tanto, las series incorporarán los valores predichos con dichos modelos.

De esta manera, se tendrá una tabla tal que:

Fecha	Euros_m2	numeroHip	Capital_mile...	numeroViv	numeroCom
JAN2007	3.002.20	18622	4303037	1178	7978
FEB2007	3.000.79	16065	3929815	3303	8214
MAR2007	2.994.39	19344	3837192	3158	8174
APR2007	2.989.10	15919	3257875	4458	6973
MAY2007	2.991.19	17050	3911680	3544	7678
JUN2007	2.998.13	16133	3733810	3906	7504
JUL2007	3.004.80	16479	3874075	3459	7309
AUG2007	3.007.96	13744	3077892	3386	6590
SEP2007	3.008.16	17779	3740588	3109	7528
OCT2007	3.007.40	13448	3433042	4109	6347

Figura 81: Construcción de tabla donde las series se utilizarán como variables

Esta tabla, donde las variables consisten en series observadas en el tiempo, implica que una de las principales hipótesis en los modelos de regresión, sea incumplido: los errores difícilmente estarán incorrelados en un modelo de regresión con estas variables. Por lo tanto, los parámetros de los modelos podrían no ser lo suficientemente eficientes, al no tener en cuenta la posible correlación de los residuos.

Por ello, se propone un modelo de regresión que tenga en cuenta esta posible autocorrelación. Aplicando la teoría, este modelo se puede expresar como ([13]):

$$y_t = \beta_1 x_{1t} + \dots + \beta_p x_{pt} + n_t$$

donde n_t será un proceso estacionario que recoge el efecto de todas las otras variables que pueden tener efecto sobre y_t y que sigue un proceso ARMA tal que $\phi(L)n_t = \theta(L)\varepsilon_t$, siendo ε_t un proceso de ruido blanco.

Para ello, SAS proporciona el procedimiento *autoreg*, que estima y pronostica modelos de regresión lineal para datos de series temporales cuando los errores están autocorrelacionados. Para evaluar esta autocorrelación de los residuos se utiliza el estadístico de Durbin Watson, por el cual es deseable un valor entre 1,5 y 2,5 para suponer la incorrelación. Utiliza un método de stepwise para decidir el número de retardos del modelo. Utilizando este procedimiento, se obtienen los datos de la figura 82; como se ve, el valor de DW se obtiene de 1,96, por lo que se asumen los residuos incorrelados. Además, se observan los retardos que hay que incluir en el modelo como autorregresión para conseguir esta incorrelación, donde los p-valores indican que son significativos. Se observan unos p-valores altos en algunas de las variables, especialmente el capital, por lo que esta variable aportará muy poco en el modelo. Teniendo en cuenta los valores que tomaba esta serie, es normal asumir que el parámetro sea tan bajo.

Procedimiento AUTOREG

Estimaciones de verosimilitud máxima			
SSE	4241.67054	DFE	116
MSE	36.56613	Raiz MSE	6.04699
SBC	906.076691	AIC	865.931209
MAE	4.39924878	AICC	869.583383
MAPE	0.18732877	HQC	882.24366
Verosimilitud log	-418.9656	R-cuadrado regresión transformada	0.0417
Durbin-Watson	1.9698	R-cuadrado total	0.9997
		Observaciones	130

Estimadores de parámetros						
Variable	DF	Estimador	Error estándar	Valor t	Aprox Pr > t	Etiqueta de la variable
Intercept	1	2479	107.7547	23.01	<.0001	
numeroViv	1	0.000232	0.000561	0.41	0.6796	Numero
numeroCom	1	-0.001158	0.000544	-2.13	0.0355	Numero
numeroHip	1	0.000252	0.000264	0.95	0.3425	Numero
Capital_milesEuros	1	-5.87E-8	8.2075E-7	-0.07	0.9431	Capital_milesEuros
AR1	1	-1.4688	0.0309	-47.52	<.0001	
AR3	1	0.5256	0.0474	11.09	<.0001	
AR7	1	-0.4481	0.1366	-3.28	0.0014	
AR8	1	0.8502	0.2515	3.38	0.0010	
AR9	1	-0.4794	0.1829	-2.62	0.0099	
AR11	1	-0.0110	0.0588	-0.19	0.8521	
AR17	1	0.1461	0.0603	2.42	0.0170	
AR19	1	-0.3767	0.1353	-2.78	0.0063	
AR20	1	0.2674	0.0896	2.99	0.0035	

Figura 82: Resultados Regresión Dinámica

El modelo resultante será:

$$Y_t = 2479 + 0,000232X_t - 0,001158X_{2t} + 0,000252X_{3t} - 0,000000587X_{4t} + n_t \Rightarrow$$

donde $n_t = 1,4688n_{t-1} - 0,5256n_{t-3} + 0,4481n_{t-7} - 0,8502n_{t-8} + 0,4794n_{t-9} + 0,011n_{t-11} - 0,1461n_{t-17} + 0,3767n_{t-19} - 0,2674n_{t-20} + \varepsilon_t$

Las predicciones de este modelo son los de la figura 83:

Fecha	hat	lcl	ucl
JAN2018	2400.1571525	2362.6836249	2437.6306801
FEB2018	2412.6976685	2360.5474404	2464.8478966
MAR2018	2425.5574834	2358.7235106	2492.3914562
APR2018	2438.4351744	2356.8734052	2519.9969435
MAY2018	2449.435743	2353.5454143	2545.3260717
JUN2018	2458.8084798	2345.1482211	2572.4687384
JUL2018	2469.3836532	2338.9867038	2599.7806026
AUG2018	2479.4548947	2333.5146152	2625.3951742
SEP2018	2487.6089743	2324.2145242	2651.0034244
OCT2018	2496.3617659	2316.389859	2676.3336728
NOV2018	2504.8049777	2308.8821362	2700.7278193
DEC2018	2511.6255569	2299.753207	2723.4979068

Figura 83: Predicciones Regresión Dinámica

3.4. Comparación de modelos

Vistos todos los modelos aplicados sobre la serie *Valor de tasación medio por metro cuadrado en la Comunidad de Madrid*, se va a comprobar con los valores reales cuál es el modelo más acertado. Dado que el último valor real que ha publicado el Ministerio de Fomento es el del primer trimestre de 2018, con un valor de 2413,20 euros, se va a hacer la comparativa con éste y con los valores de 2017 que se han utilizado en la propia serie. Por lo tanto, la predicción real es la de enero de 2018.

FECHA	ene-17	abr-17	jul-17	oct-17	ene-18
REAL	2234,5	2263,6	2305,2	2354,9	2413,2
SUAVIZADO	2227,01761	2261,29076	2291,33932	2338,59025	2394,95
ARIMA (1,1,0) SAS	2231,94228	2258,07887	2285,59183	2336,63849	2392,45993
ARIMA (0,2,1) R	2227,13132	2261,39675	2291,41332	2338,74868	2395,16775
REG DIN	2235,25989	2265,40348	2301,33885	2357,21903	2410,56926
NNAR (1,1)	2220,48745	2250,99172	2279,43865	2320,08787	2368,58345
NNAR(2,5)	2214,20984	2245,05209	2273,42311	2313,39765	2391,90281
NNAR(1,4)	2236,04966	2261,52903	2307,81078	2353,46155	2360,4354
NNAR(6,6)	2234,22588	2270,67703	2292,59218	2343,02626	2418,71773
ELM_40 nodos	2222,43492	2286,78944	2272,21006	2340,49897	2388,91973
ELM_65nodos	2225,93223	2264,85226	2282,75849	2336,50145	2389,88042

Figura 84: Comparación de la predicción

FECHA	ene-17	abr-17	jul-17	oct-17	ene-18	ERROR CUADRADO MEDIO
REAL	2234,5	2263,6	2305,2	2354,9	2413,2	
SUAVIZADO	55,9861242	5,332606	192,118442	266,008105	333,0625	170,5015554
ARIMA (1,1,0) SAS	6,54191932	30,4828654	384,480268	333,482576	430,150404	237,0276065
ARIMA (0,2,1) R	54,2974007	4,85431938	190,072628	260,865105	325,162004	167,0502915
REG DIN	0,57743327	3,25252929	14,9084948	5,37789272	6,92077716	6,207425442
NNAR (1,1)	196,351586	158,968725	663,647257	1211,88446	1990,6368	844,2977667
NNAR(2,5)	411,690796	344,024965	1009,77048	1722,44539	453,570302	788,3003869
NNAR(1,4)	2,40144612	4,28890846	6,81617221	2,06912977	2784,10301	559,9357339
NNAR(6,6)	0,07514397	50,0843253	158,957075	140,985678	30,4453002	76,10950441
ELM_40 nodos	145,566155	537,750313	1088,33594	207,389579	589,531366	513,7146712
ELM_65nodos	73,4066999	1,56815511	503,621236	338,506752	543,803045	292,1811777

Figura 85: Error cuadrado medio de la predicción y los valores reales

Como se puede observar, la mejor predicción la otorga la regresión dinámica con un valor muy cercano a la realidad.

4. CONCLUSIONES

Recordando el objetivo principal de este trabajo, consistente en tratar de prever los valores que tomarán las distintas series que influyen en el precio de venta de la vivienda, y en base a las figuras de la sección anterior donde se establecían los gráficos comparativos y los errores medios cuadrados (figuras 84 y 85) para la serie del Valor tasado medio en la Comunidad de Madrid, se observa lo siguiente:

- La serie Valor Tasado medio por metro cuadrado en Madrid está creciendo a un ritmo muy alto. Teniendo en cuenta que el estudio se ha realizado partiendo de los datos desde el año 2000, es curioso que de los 10 métodos empleados, sólo uno de ellos (NNAR (6,6)) haya dado una predicción sobrepasada. El resto de modelos, (con la salvedad de la regresión, que ha acertado completamente) se han quedado cortas. La principal conclusión que se puede extraer de ello es el ritmo vertiginoso al que está creciendo el nivel de la serie, inexplicable desde un punto de vista objetivo basado en los propios datos.

-Considerar la serie como un modelo de regresión dinámica dependiente de otras variables que influyen en el mercado inmobiliario, como las financiaciones que dan los bancos, las viviendas construidas, etc., hace mejorar el modelo. La principal conclusión que se extrae de ello es que es más acertado considerar las series relacionadas con la vivienda, como series con influencias entre ellas, y no como series independientes, pues el mercado de la construcción y la vivienda son precisamente temas que pueden provocar crisis potentes en el panorama nacional si no se les da relevancia suficiente y no se toman medidas para evitar situaciones vividas y sufridas.

-El resto de series analizadas mantienen también una tendencia positiva, lo que refleja el fin de la crisis.

El trabajo a futuro que cabría continuar, sería establecer modelos de regresión donde se incorporara la relación con los precios de alquiler, pues creemos que las subidas de los arrendamientos está provocando también que la gente opte por comprar en lugar de alquilar, y por tanto, se produzcan subidas en las transmisiones de compraventas, y en las hipotecas solicitadas y concedidas. También sería útil hacer el mismo estudio con dichos precios de alquiler, para analizar a fondo cómo se está desarrollando el mercado inmobiliario de los alquileres, y establecer paralelismos con el de la venta.

BIBLIOGRAFÍA

- [1] BOX,G, JENKINS,G *Time series analysis: forecasting and control*, Holden Day, San Francisco, 1976.
- [2] MAURICIO, J.A. *Introducción al análisis de series temporales*, Universidad Complutense de Madrid, 2007.
- [3] PEÑA,D. *Análisis de series temporales*, Alianza Editorial, Madrid, 2005.
- [4] PEÑA,D. *Análisis de datos multivariantes*, McGraw-Hill, Madrid, 2002.
- [5] PULIDO, A., LÓPEZ A.M., RODRÍGUEZ J. *Curso Combinado de Predicción y Simulación*, <http://www.uam.es/predysim>. Universidad Autónoma de Madrid, 2004, cap. 2,4,5.
- [6] URIEL, E. *Análisis de series temporales: Modelos ARIMA*, Paraninfo, Valencia, 1985.
- [7] WEI, W.S.I. *Time series analysis: univariate and multivariate methods*, Adison - Wesley, California, 1990.
- [8] SINGH, R. , BALASUNDARAM S. Application of Extreme Learning Machine Method for Time Series Analysis *International Journal of Intelligent Technology*, Vol.2(Núm.4).
- [9] HYNDMAN, ROB J.,ATHANASOPOULOS, G. *Forecasting: Principles and Practice*, <http://OTexts.com>. OTexts, Monash University, 2018.
- [10] ZHANG, P.G. Time series forecasting using a hybrid ARIMA and neural network model *Neurocomputing* 50, Enero 2003 p.159-175.
- [11] PORTELA, J. Introduccion. Redes Neuronales *Material de la asignatura Redes Neuronales*, Universidad Complutense de Madrid,2018
- [12] ALONSO REVENGA, J.M. Análisis Descriptivo de una serie temporal. *Material de la asignatura Complementos de Formación en Técnicas de Minería de Datos*, Universidad Complutense de Madrid,2016

- [13] ALONSO REVENGA, J.M. Técnicas avanzadas de predicción *Material de la asignatura Técnicas avanzadas de predicción del grado en Estadística Aplicada*, Universidad Complutense de Madrid,2016
- [14] PARRA, M *GitHub Development Platform* <https://github.com/manuparra/seriestemporales>
- [15] AUTORES VARIOS *Banco de Datos de Vivienda del Instituto de Estadística* http://www.madrid.org/baco_web/html

Parte I

Anexo

A. Código SAS

```
/* Total hipotecas constituidas en Comunidad de Madrid */
```

```
libname libtfm 'C:\Users\Tere\Desktop\tfm\libtfm';  
proc import datafile =  
"C:\Users\Tere\Desktop\tfm\TFM datos SAS\total_hipotecas_constituidas_Madrid.xls"  
out=libtfm.hipotecasConst dbms=xls replace;  
run;
```

```
data libtfm.hipotecas2;  
set libtfm.hipotecasConst;  
format Fecha mmyyd.;  
run;
```

```
proc sgplot data =libtfm.hipotecas2;  
series x=fecha y=numero;  
RUN;
```

```
proc sgplot data =libtfm.hipotecas2;  
where fecha between '01jan2003'd and '01jan2005'd;  
series x=fecha y=numero;  
RUN;
```

```
ods graphics on; /*necesario para los plots siguientes*/  
proc timeseries data=libtfm.hipotecas2 PLOT=(DECOMP)  
PRINT=(SEASONS DECOMP);  
id fecha interval=month;  
var numero;  
run;
```

```

proc esm data=libtfm.hipotecas2 out=salida lead=12 print=(FORECASTS ESTIMATES)
PLOTS =(FORECASTS MODELS SEASONS);
id fecha interval=month;
forecast numero / model=Addwinters;
run;
/*MULTIPLICATIVO*/

```

```

proc esm data=libtfm.hipotecas2 out=salida lead=12 print=(FORECASTS ESTIMATES)
PLOTS =(FORECASTS MODELS SEASONS);
id fecha interval=month;
forecast numero / model=multwinters;
run;
/*LINEAL*/

```

```

proc esm data=libtfm.hipotecas2 out=salida lead=12 print=(FORECASTS ESTIMATES)
PLOTS =(FORECASTS MODELS SEASONS);
id fecha interval=month;
forecast numero / model=lineal;
run;

```

```

/*proc arima data=libtfm.hipotecas2;*/
/*identify var= numero nlag=24;*/
/*run;*/
/**/
/*proc arima data=libtfm.hipotecas2;*/
/*identify var= numero (1 12) nlag=24;*/
/*run;*/

```

```

proc arima data=libtfm.hipotecas2;
identify var= numero (1 12) nlag=24;
estimate p=(1)(12) q=1 noconstant ;
run;
proc arima data=libtfm.hipotecas2;
identify var= numero (1 12) nlag=24;
estimate p=(1)(12) q=1 noconstant ;
forecast lead=12 id=fecha interval=month out=Predicciones;

```

```

run;

/*data libtfm.transacciones3;*/
/*set libtfm.transacciones2;*/
/*leuros=log(euros);*/
/*run;*/
/*proc arima data = libtfm.transacciones3;*/
/*identify var=Leuros(1);*/
/*estimate p=1 noconstant;*/
/*run;*/
/**/

/* Capital total hipotecado en Comunidad de Madrid */

libname libtfm 'C:\Users\Tere\Desktop\tfm\libtfm';
proc import datafile
='C:\Users\Tere\Desktop\tfm\TFM datos SAS\
Capital_total hipotecado en Madrid.xls'
out=libtfm.capitalhip dbms=xls replace;
run;

data libtfm.capital2;
set libtfm.capitalhip;
format Fecha mmyyd.;
run;

proc sgplot data =libtfm.capital2;
series x=fecha y=Capital_milesEuros;
RUN;

proc sgplot data =libtfm.capital2;
where fecha between '01jan2003'd and '01jan2005'd;
series x=fecha y=Capital_milesEuros;
RUN;

ods graphics on; /*necesario para los plots siguientes*/
proc timeseries data=libtfm.capital2 PLOT=(DECOMP)

```

```

PRINT=(SEASONS DECOMP);
id fecha interval=month;
var Capital_milesEuros;
run;

proc esm data=libtfm.capital2 out=salida lead=12
print=(FORECASTS ESTIMATES)
PLOTS =(FORECASTS MODELS SEASONS);
id fecha interval=month;
forecast Capital_milesEuros / model=addwinters;
run;

proc esm data=libtfm.capital2 out=salida lead=12
print=(FORECASTS ESTIMATES)
PLOTS =(FORECASTS MODELS SEASONS);
id fecha interval=month;
forecast Capital_milesEuros / model=multwinters;
run;

proc esm data=libtfm.capital2 out=salida lead=12
print=(FORECASTS ESTIMATES)
PLOTS =(FORECASTS MODELS SEASONS);
id fecha interval=month;
forecast Capital_milesEuros / model=lineal;
run;

proc arima data=libtfm.capital2;
identify var= Capital_milesEuros nlag=24;
run;
data libtfm.capital3;
set libtfm.capital2;
LCapital_milesEuros=log(Capital_milesEuros);
run;
proc arima data=libtfm.capital3;

```

```

identify var= lCapital_milesEuros nlag=24;
run;

proc arima data=libtfm.capital3;
identify var= lCapital_milesEuros (1 12) nlag=24;
run;

proc arima data=libtfm.capital3;
identify var= lCapital_milesEuros (1 12) nlag=24;
estimate q=(1)(12) noconstant;
run;

/*data libtfm.transacciones3;*/
/*set libtfm.transacciones2;*/
/*leuros=log(euros);*/
/*run;*/
/*proc arima data = libtfm.transacciones3;*/
/*identify var=Leuros(1);*/
/*estimate p=1 noconstant;*/
/*run;*/
/**/

/*OUTLIERS*/
proc arima data=libtfm.capital3;
identify var= lCapital_milesEuros (1 12) nlag=24;
estimate q=(1)(12) noconstant;
outlier maxnum=3 id=fecha;
run;
/*data libtfm.capital4;*/
/*set libtfm.capital3;*/
/*AO102008=(fecha='1oct2008'd);*/
/*LS042011=(fecha>='1apr2011'd);*/
/*AO042014=(fecha='1apr2014'd);*/
/*AO032014=(fecha='1mar2014'd);*/
/*AO122008=(fecha='1dec2008'd);*/

```

```

/*run;*/
/**/
/*proc arima data=libtfm.capital4;*/
/*identify var= lCapital_milesEuros (1 12)
crosscorr=(AO102008 LS042011 AO042014 AO032014 AO122008) nlag=24;*/
/*estimate q=(1)(12)
input= (AO102008 LS042011 AO042014 AO032014 AO122008) noconstant;*/
/*outlier maxnum=5 alpha=0.01 id=fecha;*/
/*run;*/
/**/
/*proc arima data=libtfm.capital4;*/
/*identify var= lCapital_milesEuros
crosscorr=(AO102008 LS042011 AO042014 AO032014 AO122008) ;*/
/*estimate q=(1)(12)
input= AO102008 LS042011 AO042014 AO032014 AO122008
method=ml noconstant;*/
/*outlier maxnum=5 alpha=0.01 id=fecha;*/
/*forecast printall lead=20 id=fecha interval=month out=Predicciones;*/
/*run;*/
data libtfm.capital5;
set libtfm.capital3;
AO102008=(fecha='1oct2008'd);
LS042011=(fecha>='1apr2011'd);
AO042014=(fecha='1apr2014'd);
/*AO032014=(fecha='1mar2014'd);*/
/*AO122008=(fecha='1dec2008'd);*/
run;
/*AO032014 (1 12) AO122008 (1 12)*/

proc arima data=libtfm.capital5;
identify var= lCapital_milesEuros (1 12)
crosscorr=(AO102008(1 12) LS042011 (1 12) AO042014(1 12) )
nlag=24;
estimate q=(1)(12) input= (AO102008 LS042011 AO042014 ) noconstant;
/*outlier maxnum=5 alpha=0.01 id=fecha;*/
run;

```

```

data an(drop=i);
do i=1 to 12;
LS042011=1;AO102008=0;AO042014=0;
fecha=intnx('month', '01jan2018'd, i-1);output;
end;
run;
proc append base=libtfm.capital5 data=an force;run;

proc arima data=libtfm.capital5;
identify var= lCapital_milesEuros (1 12)
crosscorr=(AO102008(1 12) LS042011 (1 12) AO042014(1 12) )
nlag=24;
estimate q=(1)(12) input= (AO102008 LS042011 AO042014 ) noconstant;
/*outlier maxnum=5 alpha=0.01 id=fecha;*/
forecast printall lead=12 id=fecha interval=month out=Predicciones;
run;

data pr;
set predicciones;
FORECAST_T=exp(forecast);
if _n_>180;
run;

/* Viviendas libres terminadas en Comunidad de Madrid */

libname libtfm 'C:\Users\Tere\Desktop\tfm\libtfm';
proc import datafile =
"C:\Users\Tere\Desktop\tfm\TFM datos SAS\Viviendas libres terminadas en Madrid.xls"
out=libtfm.viviendas dbms=xls replace;
run;

data libtfm.viviendas2;
set libtfm.viviendas;
format Fecha mmyyd.;

```

```

run;

proc sgplot data =libtfm.viviendas2;
series x=fecha y=Numero;
RUN;

proc sgplot data =libtfm.viviendas2;
where fecha between '01jan2003'd and '01jan2005'd;
series x=fecha y=numero;
RUN;

ods graphics on; /*necesario para los plots siguientes*/
proc timeseries data=libtfm.viviendas2 PLOT=(DECOMP)
PRINT=(SEASONS DECOMP);
id fecha interval=month;
var numero;
run;

proc esm data=libtfm.viviendas2 out=salida lead=12 print=(FORECASTS ESTIMATES)
PLOTS =(FORECASTS MODELS SEASONS);
id fecha interval=month;
forecast numero / model=Addwinters;
run;

proc esm data=libtfm.viviendas2 out=salida lead=12 print=(FORECASTS ESTIMATES)
PLOTS =(FORECASTS MODELS SEASONS);
id fecha interval=month;
forecast numero / model=multwinters;
run;

proc esm data=libtfm.viviendas2 out=salida lead=12 print=(FORECASTS ESTIMATES)
PLOTS =(FORECASTS MODELS SEASONS);
id fecha interval=month;
forecast numero / model=lineal;
run;

proc arima data=libtfm.viviendas2;
identify var= numero nlag=24;
run;

```

```

data libtfm.viviendas3;
set libtfm.viviendas2;
lnumero=log(numero);
run;

proc arima data=libtfm.viviendas3;
identify var= lnumero (1 12) nlag=24;
run;

proc arima data=libtfm.viviendas3;
identify var= lnumero (1 12) nlag=24;
estimate q=(1)(12) noconstant ;
forecast printall lead=12 id=fecha interval=month out=Predicciones;

run;

/* Transmisiones de propiedad de viviendas con
compraventa en Comunidad de Madrid. */

libname libtfm 'C:\Users\Tere\Desktop\tfm\libtfm';
proc import datafile =
"C:\Users\Tere\Desktop\tfm\
TFM datos SAS\Transmisiones propiedad compraventa en Madrid.xls"
out=libtfm.compraventa dbms=xls replace;
run;

data libtfm.comprav2;
set libtfm.compraventa;
format Fecha mmyyd.;
run;

proc sgplot data =libtfm.comprav2;
series x=Fecha y=Numero;
RUN;

proc sgplot data =libtfm.comprav2;

```

```

series x=Fecha y=Numero;
where Fecha > '01jan2009'd and fecha<'01jan2011'd;

RUN;
/*
ods graphics on; /*necesario para los plots siguientes*/
proc timeseries data=libtfm.comprav2 PLOT=(DECOMP)
PRINT=(SEASONS DECOMP);
id fecha interval=month;
var numero;
run;

proc esm data=libtfm.comprav2 out=salida lead=12 print=(FORECASTS ESTIMATES)
PLOTS =(FORECASTS MODELS SEASONS);
id fecha interval=month;
forecast numero / model=addwinters;
run;
proc esm data=libtfm.comprav2 out=salida lead=12 print=(FORECASTS ESTIMATES)
PLOTS =(FORECASTS MODELS SEASONS);
id fecha interval=month;
forecast numero / model=multwinters;
run;

proc arima data=libtfm.comprav2;
identify var= numero nlag=24;
run;
/*data libtfm.viviendas3;*/
/*/*set libtfm.viviendas2;*/*/
/*/*lnúmero=log(numero);*/*/
/*/*run;*/;

proc arima data=libtfm.comprav2;
identify var= numero (1 12) nlag=24;
run;

proc arima data=libtfm.comprav2;

```

```

identify var= numero (1 12) nlag=24;
estimate q=(1)(12) noconstant ;
forecast printall lead=12 id=fecha interval=month
out=Predicciones;

```

```
run;
```

```
/* Valor tasado medio de la vivienda libre en Comunidad de Madrid */
```

```
libname libtfm 'C:\Users\Tere\Desktop\tfm\libtfm';
```

```

data libtfm.tasacion2;
set libtfm.tasacion;
format fecha ddmmyy10.;
informat fecha ddmmyy10.;
fecha2=put(fecha , ddmmyy10.);
run;

```

```
DATA libtfm.tasacion;
```

```
LENGTH
```

```

Fecha          8
Euros _m2      8 ;

```

```
FORMAT
```

```

Fecha          YYQ9.
Euros _m2      COMMA8.2 ;

```

```
INFORMAT
```

```

Fecha          DATE9.
Euros _m2      BEST7. ;

```

```
INFILE
```

```
'C:\Users\Tere\AppData\Local\Temp\SEG260\valortasadomedioComunidadMadrid.txt'
```

```
LRECL=14
```

```
ENCODING="WLATIN1"
```

```
TERMSTR=CRLF
```

```
DLM= '7F'x
```

```
MISSOVER
```

```
DSD ;
```

INPUT

```
Fecha          : ?? ANYDTDTE6.  
Euros _m2      : ?? COMMAX7. ;
```

RUN;

```
proc sgplot data =libtfm.tasacion;  
series x=fecha y=Euros_m2;
```

RUN;

```
proc sgplot data =libtfm.tasacion;  
where fecha between '01jan2004'd and '01jan2008'd;  
series x=fecha y=Euros_m2;
```

RUN;

```
ods graphics on; /*necesario para los plots siguientes*/  
proc timeseries data=libtfm.tasacion PLOT=(DECOMP)  
PRINT=(SEASONS DECOMP);  
id fecha interval=qtr;  
var Euros_m2;  
run;
```

```
/**/  
/*proc esm data=libtfm.transacciones2  
out=salida lead=4 print=(FORECASTS ESTIMATES)*/  
/*PLOTS =(FORECASTS MODELS SEASONS);*/  
/*id fecha interval=qtr;*/  
/*forecast euros / model=Addwinters;*/  
/*run;*/
```

```
proc esm data=libtfm.tasacion out=salida lead=4 print=(FORECASTS ESTIMATES)  
PLOTS =(FORECASTS MODELS SEASONS);  
id fecha interval=qtr;  
forecast Euros_m2 / model=multwinters;  
run;
```

```

proc esm data=libtfm.tasacion out=salida lead=4 print=(FORECASTS ESTIMATES)
PLOTS =(FORECASTS MODELS SEASONS);
id fecha interval=qtr;
forecast Euros_m2 / model=lineal;
run;

proc arima data=libtfm.tasacion;
identify var= Euros_m2 nlag=12;
run;

proc arima data=libtfm.tasacion;
identify var= Euros_m2(1) nlag=12;
run;

/*proc arima data=libtfm.tasacion;*/
/*identify var= Euros_m2 (1) nlag=12;*/
/*estimate p=1 q=1 noconstant;*/
/*run;*/
/*proc arima data=libtfm.tasacion;*/
/*identify var= Euros_m2 (1) nlag=12;*/
/*estimate p=1 noconstant;*/
/*run;*/
/*proc arima data=libtfm.tasacion;*/
/*identify var= Euros_m2 (1) nlag=12;*/
/*estimate p=1 q=1 noconstant;*/
/*forecast lead=12 id=fecha interval=qtr
out=Predicciones;*/
/*run;*/
proc arima data=libtfm.tasacion;
identify var= Euros_m2 (1) nlag=12;
estimate p=1 noconstant;
forecast lead=8 id=fecha interval=qtr
out=Predicciones;
run;
data pr;
set predicciones;
/*FORECAST_T=exp(forecast);*/

```

```

if _n.>72;
run;

libname libtfm 'C:\Users\Tere\Desktop\tfm\libtfm';

proc expand data=libtfm.tasacion out=libtfm.tasacionRG from=qtr to=month
plot= all;
id fecha;

run;
data libtfm.tasacionRG;
set libtfm.tasacionRG;
if fecha<'01jan2007'd then delete;
run;
data libtfm.capitalRG;
set libtfm.capital2;
if fecha<'01jan2007'd /*or fecha>'01oct2017'd */then delete;
format fecha monyy7.;
run;
proc sql;
create table capitalpredRG as
select fecha, exp(forecast) as Capital_milesEuros
from predicciones where fecha>'31dec2017'd;
quit;
proc append base=libtfm.capitalRG data=capitalpredRG force;run;

data libtfm.hipotecasRG;
set libtfm.hipotecas2;
if fecha<'01jan2007'd
/*or fecha>'01oct2017'd*/
then delete;
format fecha monyy7.;
rename numero=numeroHip;
run;
proc sql;
create table hippredRG as

```

```

select fecha , forecast as numeroHip
from predicciones where fecha>'31dec2017'd;
quit;
proc append base=libtfm.hipotecasRG data=hippredRG force;run;

data libtfm.viviendasRG;
set libtfm.viviendas2;
if fecha<'01jan2007'd
/*or fecha>'01oct2017'd */
then delete;
format fecha monyy7.;
rename numero=numeroViv;
run;
proc sql;
create table vivipredRG as
select fecha , exp(forecast) as numeroViv
from predicciones where fecha>'31dec2017'd;
quit;
proc append base=libtfm.viviendasRG data=vivipredRG force;run;
data libtfm.compravRG;
set libtfm.comprav2;
if fecha<'01jan2007'd
/*or fecha>'01oct2017'd */
then delete;
format fecha monyy7.;
rename numero=numeroCom;
run;
proc sql;
create table CVpredRG as
select fecha , forecast as numeroCom
from predicciones where fecha>'31dec2017'd;
quit;
proc append base=libtfm.compravRG data=CVpredRG force;run;
/*JUNTAMOS TODAS*/
data libtfm.todas;
merge libtfm.tasacionRG
libtfm.hipotecasRG libtfm.capitalRG libtfm.viviendasRG libtfm.compravRG;

```

```

by fecha ;
run ;

proc autoreg data=libtfm.todas ;
model euros_m2=numeroViv numeroCom numeroHip capital_mileseuros /
nlag=20 backstep method=ml ;
output out=p p=hat pm=ptrend lcl=lcl ucl=ucl ;
run ;
/*data rg2;*/
/*set libtfm.todas;*/
/*fecha2=intnx( 'month', fecha, _n_-1 );*/
/*EurosD=dif(Euros_m2);*/
/*numeroHipD=dif(numeroHip);*/
/*capiD=dif(Capital_milesEuros);*/
/*numeroVivD=dif(numeroViv);*/
/*numeroComD=dif(numeroCom);*/
/*run;*/
/**/
/**/
/*proc autoreg data=rg2;*/
/*model EurosD=numeroHipD capiD numeroComD numeroVivD / nlag=5 backstep method=ml;*/
/*output out=p p=hat pm=ptrend lcl=lcl ucl=ucl;*/
/*run;*/
/**/
/*data pred;*/
/*euros_m2=.;*/
/*do fecha='01nov2017'd to '01dec2018'd ;*/
/*output;*/
/*fecha=intnx( 'month', fecha, 1);*/
/*end;*/
/*format fecha monyy7.;*/
/*run;*/
/*data b;*/
/*merge libtfm.todas pred;*/
/*by fecha;*/
/*run;*/
/**/

```

```

/*proc autoreg data=b;*/
/*model euros_m2=numeroViv numeroCom numeroHip capital_mileseuros /nlag=2;*/
/*output out=p p=hat pm=ptrend lcl=lcl ucl=ucl;*/
/*run;*/

```

B. Código R

```

#Carga librerias
library(boot)
library(sas7bdat)
library(xlsx)
library(readxl)
library(lubridate)
library(TSA)
library(tseries)
library(forecast)
library(nnfor)

#Lectura de datos
#viv <- read_excel('C:/Users/Tere/Desktop/tfm/R/tasacion.xlsx')

viv2 <- read_excel('C:/Users/Tere/Desktop/tfm/R/tasacion_st.xlsx')
#viv$Fecha=as.Date(viv$Fecha, "%Y-%m-%d")

class(viv$Fecha)
class(viv$Euros_m2)
plot(viv)
viv.ts <- ts(data = viv2, start = c(2000,1), frequency = 4)
frequency(viv.ts)
plot(viv.ts)
print(viv.ts)
end(viv.ts)
cycle(viv.ts)
boxplot(viv.ts ~ cycle(viv.ts))
viv.des <- decompose(viv.ts)
plot(viv.des, xlab='Anyo')

vividif <- diff(viv.ts)

```

```

vividif2 <- diff(vividif)
f <- ndiffs(viv.ts)
s <- nsdiffs(viv.ts)
plot(vividif2)
adf <- adf.test(vividif)
adf$p.value
adf <- adf.test(vividif2)
adf$p.value
autoplot(acf(vividif2, plot = FALSE))
autoplot(pacf(vividif2, plot = FALSE))
#vemos que arima nos calcula R

autoarima <- auto.arima(viv.ts)

autoarima

#se obtienen varios modelos para ver cual tiene menor AIC, y se comparan con los
#modelos de autoarima

arima1<- arima(viv.ts, c(1,1,0)) #modelo original con una dif y ar1
arima2<- arima(vividif, c(1,0,1)) #modelo difer 1 vez y ar1 ma1
arima3<- arima(vividif, c(1,0,0)) #modelo difer 1 vez y ar1 (modelo SAS)
arima4<- arima(vividif, c(0,0,1)) #modelo difer 1 vez y ma1
arima5<- arima(vividif2, c(0,0,1)) #modelo difer 2 veces y ma1 (modelo R)
arima6<- arima(vividif2, c(1,0,1)) #modelo difer 2 veces y ar 1 ma1

AIC(arima1, arima2, arima3, arima4, arima5, arima6, autoarima)
autoplot(acf(autoarima$residuals, plot = FALSE))
autoplot(pacf(autoarima$residuals, plot = FALSE))
bp <- Box.test(autoarima$residuals)
bp$p.value
lb <- Box.test(autoarima$residuals, type="Ljung-Box")
lb$p.value
forec<-forecast(autoarima, h=10)
forec$fitted

```

```

viv.ts
plot(forec)
#View(forec)
accuracy(forec)
matriz.forec <-data.frame(forec$mean,forec$lower,forec$upper)

#REDES NEURONALES
#nnetar
set.seed(123)
rn <- nnetar(viv.ts, P=0, repeats=100, act.fct='logistic')
rn

rn2 <- nnetar(viv.ts, P=0, size=5,p=2, repeats=100, act.fct='logistic')
rn2
rn3 <- nnetar(viv.ts, P=0, size=4,p=1, repeats=100, act.fct='tanh')
rn3
rn4 <- nnetar(viv.ts, P=0, size=6,p=6, repeats=100, act.fct='logistic')
rn4
f <- forecast(rn, PI=TRUE,h=10)
f2 <- forecast(rn2, PI=TRUE,h=10)
f3 <- forecast(rn3, PI=TRUE,h=10)
f4 <- forecast(rn4, PI=TRUE,h=10)
autoplot(f,h=10)
f
f2
f3
f4
f$fitted
f2$fitted
summary(rn3)
plot(rn)

#nnfor
set.seed(123)
rnelm <- elm(viv.ts, rep=100, type="step")
rnelm2 <- elm(viv.ts, rep=100, type="lasso")

```

```
rnelm
rnelm2
summary(rnelm)
plot (rnelm)
felm <- forecast(rnelm, h=10)
felm2 <- forecast(rnelm2, h=10)
felm2<- forecast(rnelm)
felm$fitted
felm
felm2
autoplot(felm, ,h=10)
rnelm
plot(rnelm, felm)
```