



Trabajo Fin de Máster

Forecast de ventas

Damián Pallas Carrillo

Máster en Técnicas Estadísticas

Curso 2018-2019

Propuesta de Trabajo Fin de Máster

Título en galego: Predicción de vendas
Título en castellano: Predicción de ventas
English title: Sales forecast
Modalidad: Modalidad B
Autor/a: Damián Pallas Carrillo, Universidade de Santiago de Compostela
Director/a: César Andrés Sánchez Sello, Universidade de Santiago de Compostela
Tutor/a: Jorge López Muñiz, Estrella Galicia
Breve resumen del trabajo: Se hará un estudio de las Series Temporales con la finalidad de aportar predicciones sobre las ventas de Estrella Galicia mediante la utilización de su base de datos. Hablaremos de los modelos Arima y ETS y, luego, de modelos más avanzados como la Regresión Dinámica o las Series Jerárquicas.
Recomendaciones:
Otras observaciones:

Don César Andrés Sánchez Sello, titular de universidad de la Universidad de Santiago de Compostela y don Jorge López Muñiz, director del departamento tLab & Advanced Analytics de Estrella Galicia, informan que el Trabajo Fin de Máster titulado

Forecast de ventas

fue realizado bajo su dirección por don Damián Pallas Carrillo para el Máster en Técnicas Estadísticas. Estimando que el trabajo está terminado, dan su conformidad para su presentación y defensa ante un tribunal.

En Santiago de Compostela, a 04 de julio de 2019.

El director:

El tutor:

Don César Andrés Sánchez Sello

Don Jorge López Muñiz

El autor:

Don Damián Pallas Carrillo

Índice general

Resumen	IX
1. Introducción a las Series de Tiempo	1
1.1. Conceptos básicos	1
1.2. Transformaciones para estabilizar la varianza: Box-Cox.	4
1.3. Criterios de información y selección de modelos	5
2. Familia de Modelos Autorregresivos y de Medias móviles	9
2.1. Modelos estacionarios: AR, MA y ARMA	12
2.2. Modelos no estacionarios: ARIMA	15
2.3. Modelos estacionales: ARIMA estacionales	17
2.4. Estimación de los parámetros de un modelo ARMA	21
2.5. Predicciones con modelos ARIMA	23
2.6. Aplicación de los ARIMA a los datos de Estrella Galicia	26
3. Suavizado Exponencial	33
3.1. Descomposición de series temporales	33
3.2. Clasificación de los métodos de Suavizado Exponencial	34
3.3. Ecuaciones explícitas de los métodos de Suavizado Exponencial	37
3.4. Modelos State Space	40
3.5. Inicialización y estimación de los parámetros de un modelo ETS	43
3.6. Precisión de pronóstico	45
3.7. Aplicación del Suavizado Exponencial a los datos de Estrella Galicia	46
4. Modelos avanzados	55
4.1. Motivación	55
4.2. Regresión Dinámica	57
4.2.1. Relación lineal entre series de tiempo	58
4.2.2. Estimación de parámetros	59
4.2.3. Aplicación de la Regresión Dinámica a Estrella Galicia	59
4.3. Series jerarquizadas	75

4.3.1. Enfoque “Bottom-up”	76
4.3.2. Enfoque “Top-down”	77
4.3.3. Enfoque “Middle-out”	79
4.3.4. Aplicación de las series jerárquicas a Estrella Galicia	79
5. Implementación y conclusiones	83
A. Tablas recopilatorias de los modelos ETS	87
Bibliografía	91

Resumen

Resumen en castellano

Hijos de Rivera es la empresa cervecera más importante de Galicia y una de las más importantes de España. Actualmente, está viviendo un gran crecimiento como entidad ligado con una fuerte expansión de mercado y, por ello, la previsión de ventas se ha convertido en una necesidad para su departamento de Business Intelligence en la gestión de toma de decisiones.

En este proyecto hemos trabajado con las bases de datos de la empresa con la finalidad de analizar las series de tiempo de sus ventas y poder darle así predicciones futuras sobre ellas. Comenzaremos con una breve introducción sobre series temporales para aquellos lectores menos familiarizados con el campo de estudio. Luego, en los dos capítulos siguientes, presentaremos los dos modelos clásicos más importantes para el análisis de series de tiempo: los modelos *ARIMA* y los modelos *ETS*; todo ello ejemplificado con datos reales de nuestra empresa. Debido a ciertas características de los datos, veremos como estos modelos por sí solos no son suficientes y, por ello, necesitaremos estudiar modelos más avanzados como la Regresión Dinámica o, lo que finalmente será el modelo escogido para las predicciones de nuestra cervecera, los modelos de Series Jerárquicas. Por último, comentaremos brevemente la implementación que hemos dejado en la empresa y daremos una serie de conclusiones.

English abstract

Hijos de Rivera is the most important brewery in Galicia and one of the most important in Spain. Currently, it is experiencing a huge growth as an entity linked with a strong market expansion and, therefore, the sales forecast has turned into a necessity for its Business Intelligence department in the management of decision making.

In this project we have worked with data bases of the company with the aim to analyse the Time Series of its sales and be able to obtain future predictions about them. We will begin with a short introduction on time series for those readers who are

less familiar with the field of study. Then, in the next two chapters, we will present the two classic most important models for the analysis of time series: ARIMA models and ETS models; all of this will be exemplified with real data of our company. Due to certain characteristics of the data, we will see how these models are not sufficient by themselves and, because of that, we will need to study more advanced models such as Dynamic Regression or, the one which will finally be chosen for the predictions of our brewery, the Hierarchical Series models. Finally, we will briefly discuss the implementation that we have done for the company and we will give some conclusions.

Capítulo 1

Introducción a las Series de Tiempo

En este capítulo se hará una introducción al campo de las series de tiempo y se presentarán aplicaciones del mismo a los datos de ventas de Estrella Galicia. En la primera sección, presentaremos los conceptos más básicos de series temporales, empezando por la propia definición de serie de tiempo y comentando los puntos más relevantes que hay que llevar a cabo a la hora de trabajar con ellas. Luego, en los capítulos posteriores, presentaremos los distintos métodos y modelos más utilizados para cada uno dichos puntos, siguiendo muy de cerca la estructura presentada en [Aneiros \(2018\)](#). Todo ello lo iremos ilustrando, en la medida de lo posible, con ejemplos gráficos.

1.1. Conceptos básicos

Definición 1.1. *Consideremos el proceso estocástico $\{X_t\}_{t \in \mathbb{Z}}$, donde el subíndice t de cada variable aleatoria representa el instante de tiempo en el que es observada. Una observación de dicho proceso se conoce como una realización o trayectoria del mismo y se denota por $\{x_t\}_{t \in \mathbb{Z}}$. Si consideramos un tramo finito de dicha trayectoria,*

$$x_1, x_2, \dots, x_T,$$

*diremos que estamos ante una **serie de tiempo**. Así, definimos las series temporales como una realización o trayectoria parcial de un proceso estocástico en tiempo discreto.*

Ejemplos de series de tiempo pueden ser encontrados en todo tipo de materias. En meteorología: las temperaturas diarias registradas a lo largo de un año; en economía: el PIB mensual de un Estado; o, como es el caso de nuestro trabajo, en una compañía cervecera: las ventas registradas en un período concreto.

Como bien se explica en [Velasco and del Puerto García \(2009\)](#), lo primero que se debe hacer para estudiar una serie temporal es representar su gráfico secuencial, es decir, representar cada observación x_t frente al instante t en el que fue observada.

Gracias a esta representación seremos capaces de intuir cómo evoluciona la serie a lo largo del tiempo y detectar la presencia de las principales características de una serie temporal:

- **Tendencia:** comportamiento a largo plazo de la serie. Será relevante cuando exista una variación en el valor medio de la serie conforme vamos añadiendo datos a la misma. En nuestra empresa, debido a su crecimiento y expansión de mercado, sus ventas suelen presentar una ligera tendencia creciente.
- **Estacionalidad:** comportamiento periódico de la serie provocado generalmente por factores repetitivos en el tiempo (mensualmente, semanalmente, etc.). En nuestro caso, el efecto estacional es muy significativo en las ventas de cerveza, presentando “cimas” en verano y “valles” en invierno.
- **Heterocedasticidad:** presencia de variabilidad no constante. Si bien es cierto que en los últimos meses podemos empezar a intuir un aumento de varianza, por lo general no nos encontramos con una heterocedasticidad importante.

Con los siguientes gráficos recogidos en la Figura 1.1, ilustraremos los conceptos anteriores.

Cuando una serie de tiempo carece de las características anteriormente mencionadas y, por lo tanto, fluctúa en torno a un valor constante sin un patrón periódico definido, se le denomina **serie de tiempo estacionaria**. Más rigurosamente, podemos definir las como sigue:

Definición 1.2. *Diremos que $\{X_t\}_{t \in \mathbb{T}}$ es un **proceso estrictamente estacionario** si cualesquiera que sean $h > 0$, $n \geq 1$ y t_i con $t_i + h \in \{\dots, -2, -1, 0, 1, 2, \dots\}$ para $i \in \{1, \dots, n\}$, se verifica que los vectores aleatorios $(X_{t_1}, \dots, X_{t_n})$ y $(X_{t_1+h}, \dots, X_{t_n+h})$ tienen la misma distribución conjunta.*

Como se puede comprobar en la bibliografía sobre series temporales, la mayor parte de ésta se centra en el estudio de series estacionarias. Es por ello que, como explica [Hyndman and Athanasopoulos \(2018\)](#), a la hora de afrontarnos a nuestras series de tiempo reales debemos seguir los siguientes pasos en la medida de lo posible:

1. Representar gráficamente la serie e identificar la presencia de tendencia, componente estacional o cualquier otro tipo de variación puntual sobre ella.
2. Eliminar heterocedasticidad en caso de que se presente claramente. Uno de los métodos más extendidos para ello son las Transformaciones de Box-Cox que presentaremos en la sección siguiente.

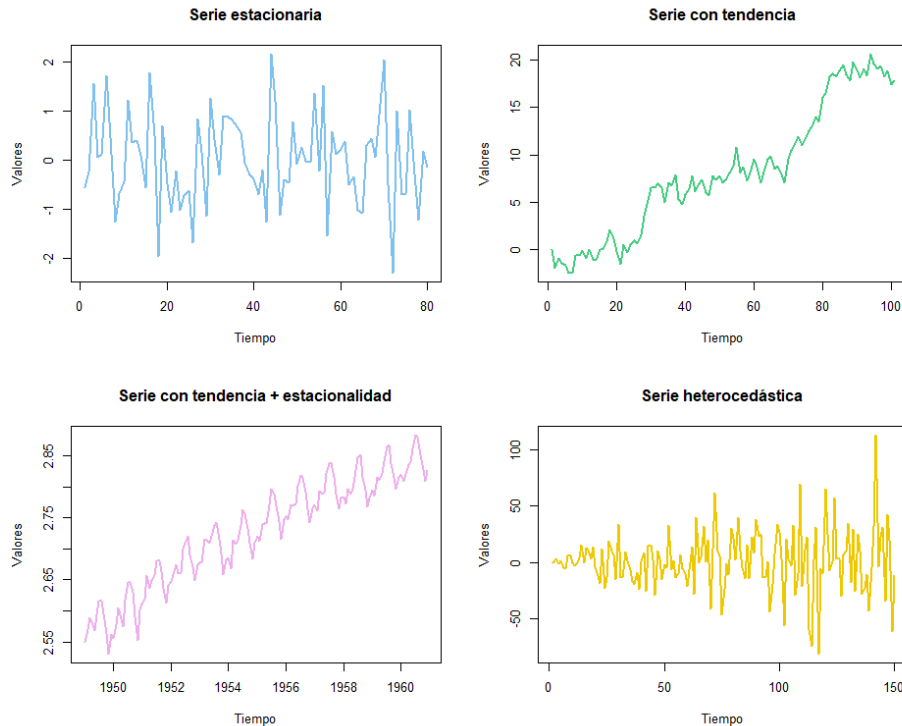


Figura 1.1: Distintos ejemplos de series temporales. En el primer gráfico (arriba a la izquierda) se nos presenta una serie estacionaria, en la que no se observa ni tendencia ni ningún patrón de repetición. En el segundo (arriba a la derecha) podemos observar una serie con tendencia creciente, en la que su valor medio va aumentando con el paso del tiempo. En el tercero (abajo a la izquierda) vemos una serie con tendencia y estacionalidad, pues su valor medio va aumentando con el paso del tiempo y además sigue un patrón de cimas y valles repetitivo. Finalmente, la última serie (abajo a la derecha) no presenta tendencia ni estacionalidad, pero sí heterocedasticidad, porque su varianza va aumentando con el paso del tiempo.

3. Eliminar tendencia y estacionalidad (si existen) para obtener residuos estacionarios. Para ello usaremos el método de la diferenciación que explicaremos más adelante.
4. Elegir un modelo que ajuste la serie de residuos estacionaria resultante de los pasos anteriores. Aquí entramos en una gran variedad de posibilidades que profundizaremos en los siguientes capítulos.
5. Comprobar que una vez ajustado nuestro modelo, se pasan los tests de chequeo de residuos.
6. Por último, predecir. Para ello se predecirán los próximos residuos y se in-

vertirán las transformaciones de los pasos anteriores para conseguir finalmente predicciones de la serie original.

Comentado todo esto, pasaremos ahora a presentar en las siguientes secciones y capítulos cada uno de los métodos más utilizados para cada una de las finalidades de los puntos anteriores.

1.2. Transformaciones para estabilizar la varianza: Box-Cox.

Muchas veces, la no estacionaridad de una serie viene dada por la presencia de heterocedasticidad en los valores observados. En la Figura 1.2(a) representamos una serie temporal en la que se puede observar claramente una varianza no constante a lo largo del tiempo.

Para estabilizar dicha varianza, las transformaciones de Box-Cox (Box and Cox (1964)) son una herramienta muy eficaz y de sencilla definición, que busca darle una forma normal a los datos a los que se le aplica.

Definición 1.3. *Definimos las transformaciones de Box-Cox de un proceso $\{X_t\}_{t \in T}$ como aquellas que transforman, para cada $t = \{1, \dots, T\}$, la observación x_t en*

$$z_t = \begin{cases} \frac{x_t^\lambda - 1}{\lambda}, & \text{si } \lambda \neq 0 \\ \log(x_t), & \text{si } \lambda = 0 \end{cases},$$

donde el parámetro λ se escoge minimizando la desviación estándar de los datos transformados, z_t .

Cabe destacar que la serie debe ser estrictamente positiva. En caso de que no lo fuese, bastaría con trasladarla al área positiva con la suma de una constante, $x_t + M$, para todo t .

La filosofía para obtener una estimación, $\hat{\lambda}$, del óptimo λ es como sigue. Suponiendo que la desviación típica es una función potencial de la media,

$$\sigma_t = k\mu_t^{1-\lambda},$$

aplicando la función logaritmo se obtiene que

$$\log(\sigma_t) = a + b \log(\mu_t), \quad (1.1)$$

donde $a = \log(k)$ y $b = 1 - \lambda$. Por tanto, si disponemos de estimaciones para σ_t y μ_t y ajustamos una regresión lineal al par $(\log(\sigma_t), \log(\mu_t))$ (como en (1.1)), obtendremos una estimación para b y, por lo tanto, también para λ :

$$\hat{\lambda} = 1 - \hat{b}.$$

Ejemplo 1.4. Supongamos la serie de tiempo simulada en R representada en la Figura 1.2(a). Como podemos observar, es claramente heterocedástica. Tras aplicar la transformación de Box-Cox, hemos obtenido un valor estimado óptimo $\hat{\lambda} = -0.2947$, siendo la serie resultante la recogida en la Figura 1.2(b). Efectivamente, ahora es homocedástica, como queríamos conseguir.

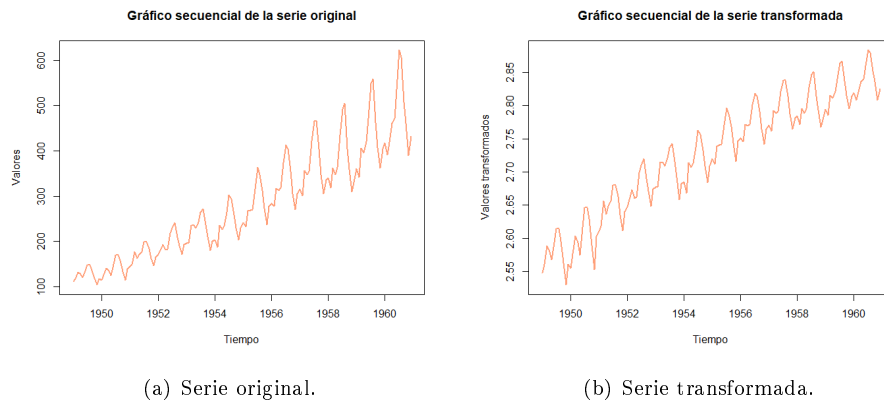


Figura 1.2: Ejemplo de una transformación Box-Cox con $\hat{\lambda} = -0.2947$.

1.3. Criterios de información y selección de modelos

Como veremos en los siguientes capítulos, existen varios modelos para ajustar nuestras series temporales y, cada uno de ellos, constará de una cantidad de parámetros que debemos estimar. Se nos presentará entonces la duda de qué modelo será el más apropiado para nuestra serie y, para poder escoger entre ellos con argumentos estadísticos, utilizaremos distintos criterios de información que presentamos a continuación:

R^2 ajustado

El valor R^2 , conocido ya por todos, puede ser utilizado como un primer dato informativo sobre nuestro ajuste. Sin embargo, no es una buena medida de la bondad de ajuste y capacidad predictiva de un modelo. Imaginemos que un modelo devuelve siempre como predicción el 10 % de los valores reales de la serie. En ese caso, el valor R^2 sería 1 (correlación perfecta), pero los pronósticos no son en absoluto acertados. Además, cuantos más parámetros añadamos a nuestro modelo (por ejemplo, cuanto mayor sea p y q en los modelos *ARIMA* que veremos en el capítulo siguiente), mayor

será el valor R^2 , incluso si dichos parámetros no son relevantes. Es por ello que decimos que R^2 tiende al sobreajuste.

Una idea equivalente es utilizar la suma de errores cuadráticos,

$$SSE = \sum_{t=1}^T \varepsilon_t^2.$$

Sin embargo, minimizar el SSE equivale a maximizar el R^2 , por lo que tenderá también a escoger el modelo con mayor número de variables, sin ser eso ninguna garantía para nosotros.

Criterio de información de Akaike (AIC)

Es uno de los criterios más utilizados. El criterio de información de Akaike se define como sigue:

$$AIC = T \cdot \log\left(\frac{SSE}{T}\right) + 2(k + 2),$$

siendo T el tamaño de la serie y k el número de parámetros del modelo. Su interpretación es muy sencilla: el primer sumando intenta medir (mediante el SSE) tanto la calidad de ajuste como la credibilidad que le da a la serie temporal, pero penalizando con el segundo sumando el número de parámetros a estimar. Nos interesa, por tanto, escoger el modelo con el menor AIC.

Criterio de información de Akaike corregido (AICc)

Para valores pequeños de T (no será el caso en nuestros casos prácticos con la empresa), el criterio AIC tiende a escoger modelos con demasiados predictores. Para arreglar esto se utiliza una versión para corregir el sesgo, el llamado AIC corregido:

$$AICc = AIC + \frac{2(k + 2)(k + 3)}{T - k - 3}.$$

Al igual que en el caso anterior, el AICc debe ser minimizado.

Criterio de información Bayesiano (BIC)

El último criterio que presentaremos antes de abordar los modelos propiamente dichos en los próximos capítulos es el Criterio de Información Bayesiano (BIC), también abreviado en la literatura como SBIC o SC. Es una medida muy relacionada con las anteriores y se define como sigue:

$$BIC = T \cdot \log\left(\frac{SSE}{T}\right) + (k + 2) \cdot \log(T).$$

Nuavemente, la finalidad será escoger el modelo que minimice este valor. Como podemos esperar, muchas veces el BIC devuelve el mismo modelo que el AIC, aunque si

difieren, el criterio BIC suele devolver modelos con menos parámetros. Esto se debe a que el segundo sumando de su definición penaliza más el número de parámetros que el del criterio AIC.

Capítulo 2

Familia de Modelos Autorregresivos y de Medias móviles

En este capítulo, tras habernos ya familiarizado con las series de tiempo en el capítulo anterior, introduciremos los modelos de series temporales univariantes más extendidos y usados en la práctica. Daremos sus definiciones y explicaremos su construcción, al mismo tiempo que daremos pautas para identificar, dada una serie temporal, un modelo de esta familia que verosimilmente la hubiese podido generar. Para ello usaremos la metodología de Box Jenkins que, como se explica en [Velasco and del Puerto García \(2009\)](#), se resume en el esquema que recogemos en la Figura 2.1.

Pero antes de todo, necesitaremos una serie de definiciones previas con las que luego poder presentar los distintos modelos.

Definición 2.1. Se llama *función de medias del proceso* $\{X_t\}_{t \in T}$ a la aplicación $\mu_t = \mathbb{E}[X_t]$, con $t \in T$.

La función de medias es una medida de posición de carácter central de X_t .

Definición 2.2. De forma análoga, llamamos *función de varianzas del proceso* $\{X_t\}_{t \in T}$ a la aplicación $\sigma^2 = \text{Var}[X_t]$, con $t \in T$.

La función de varianzas nos informa sobre el grado de variabilidad de X_t .

Definición 2.3. Llamaremos *función de autocovarianzas del proceso* $\{X_t\}_{t \in T}$ a la aplicación

$$\gamma_{t_1, t_2} = \text{Cov}(X_{t_1}, X_{t_2}) = \mathbb{E}[(X_{t_1} - \mu_{t_1})(X_{t_2} - \mu_{t_2})].$$

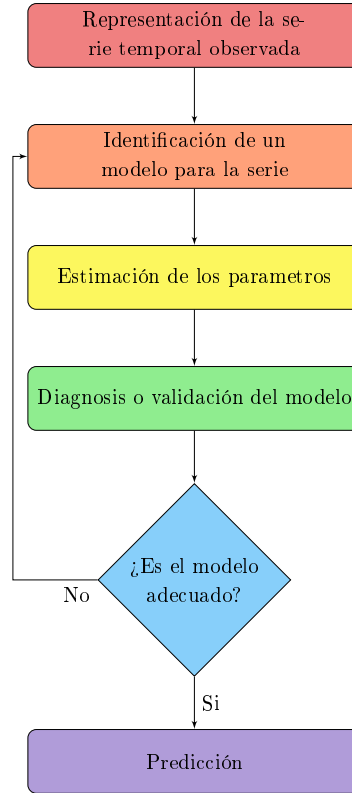


Figura 2.1: Diagrama de la metodología de Box Jenkins.

Su utilidad reside en proporcionar una medida de dependencia lineal existente entre X_{t_1} y X_{t_2} .

Definición 2.4. La *función de autocorrelación (fas)* consiste en una estandarización de la función de autocovarianzas y se define como

$$\rho_{t_1, t_2} = \frac{\gamma_{t_1, t_2}}{\sqrt{\sigma_{t_1}^2 \sigma_{t_2}^2}}.$$

Es una medida del grado de dependencia lineal existente entre X_{t_1} y X_{t_2} . Toma valores entre $[-1, 1]$.

Definición 2.5. La *función de autocorrelaciones parciales (fap)* se define como

$$\alpha_{t_1, t_2} = \frac{\text{Cov}(X_{t_1} - \hat{X}_{t_1}^{(t_1, t_2)}, X_{t_2} - \hat{X}_{t_2}^{(t_1, t_2)})}{\sqrt{\text{Var}(X_{t_1} - \hat{X}_{t_1}^{(t_1, t_2)}, X_{t_2} - \hat{X}_{t_2}^{(t_1, t_2)})}},$$

donde $\hat{X}_{t_1}^{(t_1, t_2)}$ y $\hat{X}_{t_2}^{(t_1, t_2)}$ denotan al mejor predictor lineal de X_{t_1} y X_{t_2} , respectivamente, construidos a partir de las variables medidas en los instantes comprendidos

entre t_1 y t_2 .

Esta función es una medida del grado de dependencia lineal entre X_{t_1} y X_{t_2} , una vez se les ha sustraído el efecto lineal que sobre cada una de ellas ejercen las variables medidas en los instantes comprendidos entre t_1 y t_2 . Al igual que la *fas*, toma valores entre $[-1, 1]$.

Todas las funciones que acabamos de definir nos aportan información sobre el proceso estocástico $\{X_t\}_{t \in \mathbb{Z}}$ que ha generado nuestra serie de tiempo. Sin embargo, nos encontramos con dos limitaciones críticas: por un lado, para calcular dichas funciones debemos conocer de antemano el proceso estocástico (cuando es justamente eso lo que realmente queremos averiguar) por lo que debemos estimarlas y, por otro lado, únicamente conocemos una observación x_t de cada variables X_t , con $t = 1, \dots, T$, lo que dificulta considerablemente esta estimación.

Por todo ello, debemos hacer una serie de suposiciones que nos permitan caracterizar el proceso a partir de nuestra serie de tiempo. Dicha suposición será que estamos ante un proceso estacionario (por eso en el capítulo anterior comentábamos que la gran mayoría de publicaciones en esta materia están basadas en series estacionarias) y, por lo tanto,

- la función de medias será constante: $\mu_t = \mu$, para todo $t \in T$,
- la función de varianzas también lo será: $\sigma_t^2 = \sigma^2$, para todo $t \in T$, y
- la función de autocovarianzas solo dependerá del retardo k : $\gamma_{t,t+k} = \gamma_k$, para todo t, k .

Dicha estacionaridad dota al proceso estocástico de propiedades de estabilidad en la media, varianza y autocovarianzas que nos harán más sencillo estimar las características del proceso a partir simplemente de la serie de tiempo observada, $\{x_t\}_{t=1}^T$.

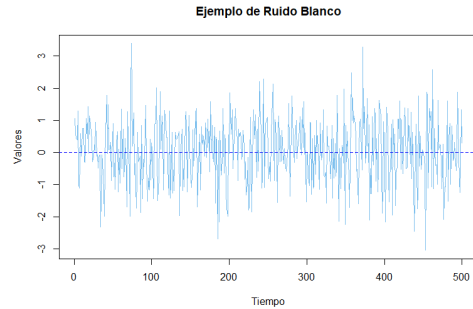
Observación 2.6. Bajo estas premisas, las versiones muestrales de las funciones definidas anteriormente son

- Media muestral: $\bar{x} = \frac{1}{T} \sum_{t=1}^T x_t$,
- Función de autocovarianzas muestrales: $\hat{\gamma}_k = \frac{1}{T} \sum_{t=1}^{T-k} (x_t - \bar{x})(x_{t+k} - \bar{x})$, con $\hat{\gamma}_{-k} = \hat{\gamma}_k$ para todo $k = 0, 1, \dots, T-1$,
- Función de autocorrelaciones simples muestrales: $\hat{\rho}_k = \frac{\hat{\gamma}_k}{\hat{\gamma}_0}$,
- Función de autocorrelaciones parciales muestrales: $\hat{\alpha}_k = \hat{\alpha}_{kk}$, donde $\hat{\alpha}_{kk}$ es el estimador mínimo cuadrático de α_{kk} en la regresión $x_t = \alpha_{k0} + \alpha_{k1}x_{t-1} + \dots + \alpha_{kk}x_{t-k} + \varepsilon$.

Ejemplo 2.7. Pondremos un ejemplo de proceso estacionario muy concreto, que incluso tiene nombre propio y que será de suma importancia en el estudio de las demás secciones: el **ruido blanco**. Llamamos ruido blanco a un proceso estocástico $\{a_t\}_t$ cuyas variables aleatorias a_t son incorreladas, con media cero y varianza finita σ_a^2 .

Así, podemos definir el ruido blanco como un proceso estacionario en el que

- $\mu = 0$,
- $\sigma^2 = \sigma_a^2$, y
- $\gamma_k = \begin{cases} \sigma_a^2 & \text{si } k = 0 \\ 0 & \text{en otro caso} \end{cases}$



También cabe comentar que si el tamaño de la serie, T , es grande, se verifica que $\hat{\rho}_k \approx \mathcal{N}(0, \frac{1}{\sqrt{T}})$ y que $\hat{\alpha}_k \approx \mathcal{N}(0, \frac{1}{\sqrt{T}})$, y por lo tanto debería cumplirse que $\hat{\rho}_k, \hat{\alpha}_k \in (-\frac{1,96}{\sqrt{T}}, \frac{1,96}{\sqrt{T}})$, para $k = 1, 2, \dots, T$.

Una vez llegados a este punto, ya estamos en condiciones de presentar la serie de modelos estocásticos paramétricos que buscamos.

2.1. Modelos estacionarios: AR, MA y ARMA

Definición 2.8. Un proceso estacionario $\{X_t\}_{t \in T}$ que permite la representación

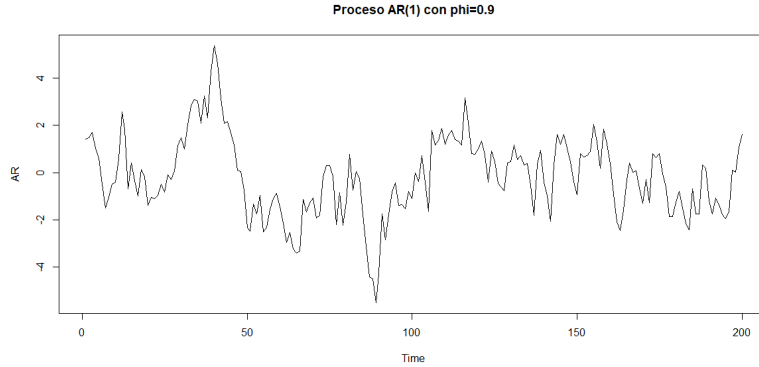
$$X_t = c + \phi_1 X_{t-1} + \dots + \phi_p X_{t-p} + a_t,$$

donde c, ϕ_1, \dots, ϕ_p son constantes y $\{a_t\}_t$ ruido blanco, se conoce como un **proceso autorregresivo de orden p** y se denota por **AR(p)**.

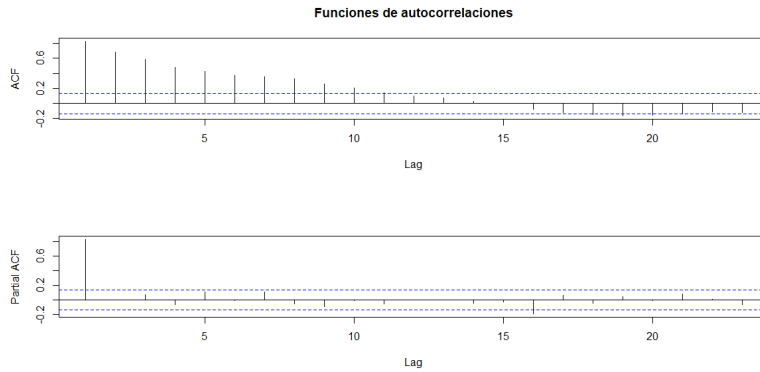
Se verifica que para este modelo, la función de autocorrelaciones parciales se “anula” para todo retardo mayor que p . Concretamente, se tiene que $\hat{\sigma}_k \in (-\frac{1,96}{\sqrt{T}}, \frac{1,96}{\sqrt{T}})$. En la Figura 2.2 se representa el gráfico secuencial de un proceso **AR(1)** junto con sus funciones *fas* y *fap*. Como podemos observar, la última $\hat{\sigma}_k$ en salirse del intervalo $(-\frac{1,96}{\sqrt{T}}, \frac{1,96}{\sqrt{T}})$ es $\hat{\sigma}_1$.

Definición 2.9. Un proceso estacionario $\{X_t\}_{t \in T}$ que permite la representación

$$X_t = c + a_t + \theta_1 a_{t-1} + \dots + \theta_q a_{t-q},$$



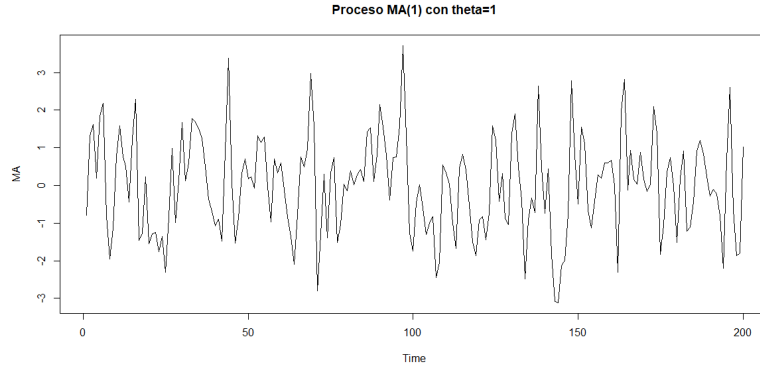
(a) Gráfico secuencial.

(b) Funciones f_{as} y f_{ap} .Figura 2.2: Simulación de un proceso $AR(1)$ con $\phi = 0,9$.

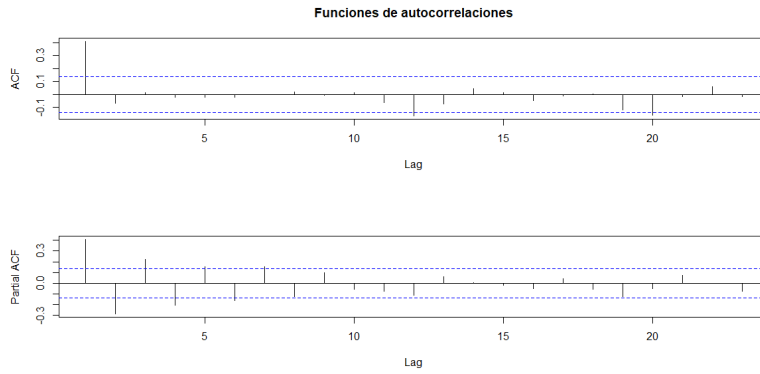
donde $c, \theta_1, \dots, \theta_q$ son constantes y $\{a_t\}_{t \in T}$ ruido blanco, se conoce como **proceso de medias móviles de orden q** y se denota por $MA(q)$.

Esta vez, se verifica para los procesos $MA(q)$ que la f_{as} se “anula” para todo retardo mayor que q . En la Figura 2.3 se representa el gráfico secuencial de un proceso $MA(1)$ junto con sus funciones f_{as} y f_{ap} . Podemos comprobar como la última $\hat{\rho}_k$ en salirse del intervalo $(-\frac{1,96}{\sqrt{T}}, \frac{1,96}{\sqrt{T}})$ es $\hat{\rho}_1$.

Definición 2.10. La fusión en un mismo proceso de una estructura autorregresiva, $AR(p)$, y de medias móviles, $MA(q)$, se conoce como un **proceso mixto autorregresivo-media móvil de orden (p, q)** y se denota por $ARMA(p, q)$. Así, un proceso es-



(a) Gráfico secuencial.

(b) Funciones f_{as} y f_{ap} .Figura 2.3: Simulación de un proceso $MA(1)$ con $\theta = 1$.

estocástico estacionario $\{X_t\}_{t \in T}$ será $ARMA(p, q)$ si permite la representación

$$X_t = c + \phi_1 X_{t-1} + \dots + \phi_p X_{t-p} + a_t + \theta_1 a_{t-1} + \dots + \theta_q a_{t-q}, \quad (2.1)$$

donde $c, \phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q$ son constantes y $\{a_t\}_{t \in T}$ ruido blanco.

Se verifica, como podemos comprobar fácilmente, que

$$[ARMA(p, 0) \Leftrightarrow AR(p)] \text{ y que } [ARMA(0, q) \Leftrightarrow MA(q)].$$

Además, es una clase muy flexible. Si $\{Y_t\}_{t \in T}$ es un proceso estacionario tal que $\gamma_{Y,h} \xrightarrow{h \rightarrow \infty} 0$, entonces, dado un $k > 0$, existe un proceso $ARMA(p, q)$ $\{X_t\}_{t \in T}$ tal que $\gamma_{X,h} = \gamma_{Y,h}$ para todo $h = 0, 1, \dots, k$.

Notación 2.11. La ecuación (2.1) que define un proceso $ARMA(p, q)$ también es muy frecuente encontrarla en la bibliografía de forma compacta escrita como sigue

$$\phi(B)X_t = c + \theta(B)a_t, \quad (2.2)$$

donde $\phi(B) = 1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p$, $\theta(B) = 1 + \theta_1 B + \theta_2 B^2 + \dots + \theta_q B^q$, y B denota el operador retardo definido por $BX_t = X_{t-1}$, $B^2 X_t = X_{t-2}$, etc.

A diferencia de los dos anteriores modelos, la identificación conjunta de los parámetros (p, q) de un modelo $ARMA(p, q)$ es una tarea ardua de hacer a partir de las funciones de autocorrelaciones simples y parciales, ya que en estos casos tienen estructuras bastante complejas con infinitos elementos distintos de cero. Para estos procesos, la mejor táctica a seguir es la simulación numérica, en la que se modela distintos pares (p, q) y se escoge el modelo que mejor ajuste nuestros datos.

2.2. Modelos no estacionarios: ARIMA

Como comentábamos antes, estos modelos solo pueden ser aplicados a procesos estacionarios. Sin embargo, en la realidad suelen presentarse procesos con tendencia e incluso estacionalidad. Es por ello que es necesario ampliar la clase de modelos $ARMA$ para poder estudiar y ajustar todos estos procesos tan frecuentes en el día a día. Comenzaremos por cubrir la no estacionariedad debida a la tendencia, dejando la estacionalidad para el siguiente apartado.

El método más extendido para eliminar tendencia es la **diferenciación**. Dado un proceso estocástico $\{X_t\}_{t \in T}$, aplicar el operador diferencia de orden 1, ∇ , es crear un nuevo proceso $\{Y_t\}_{t \in T}$ definido como

$$Y_t = \nabla X_t = X_t - X_{t-1}.$$

Como podemos comprobar, el operador diferencia es equivalente a $(1 - B)$, siendo B como hemos definido anteriormente en la ecuación (2.2).

En general, definimos el **operador diferencial de orden d** como $\nabla^d = (1 - B)^d := \nabla(\nabla^{d-1}) = (1 - B)(1 - B)^{d-1}$.

Cabe destacar que para hacer la transformación inversa, es decir, integrar el proceso $\{Y_t\}_{t \in T}$, se procede de la siguiente forma:

$$X_t = Y_t + X_{t-1} = Y_t + Y_{t-1} + X_{t-2} = \dots = Y_t + Y_{t-1} + Y_{t-2} + Y_{t-3} + \dots$$

Una vez aplicada la diferenciación a nuestro proceso no estacionario para convertirlo en estacionario, ya estamos en condiciones de intentar modelarlo mediante un $ARMA(p, q)$. Así se deduce la definición de los modelos $ARIMA(p, d, q)$ que presentamos a continuación.

Definición 2.12. Un proceso **ARIMA**(p, d, q) es aquel que, después de aplicar d diferenciaciones regulares, se convierte en un proceso **ARMA**(p, q). Equivalentemente, es un proceso que admite una representación del tipo

$$\phi(B)(1 - B)^d X_t = c + \theta(B)a_t.$$

Es decir, se tiene que

$$\{X_t\}_{t \in T} \text{ es } \text{ARIMA}(p, d, q) \Leftrightarrow (1 - B)^d X_t \text{ es } \text{ARMA}(p, q).$$

En la práctica, propondremos un modelo **ARIMA** como generador de nuestro proceso cuando la no estacionariedad de este esté provocada únicamente por la presencia de tendencia. Esto lo podremos detectar en un primero momento con el gráfico secuencial de la serie y luego con la *fas*, que toma valores positivos cercanos a 1 en los primeros retardos y que luego decrecen lentamente. Ver Figura 2.4.

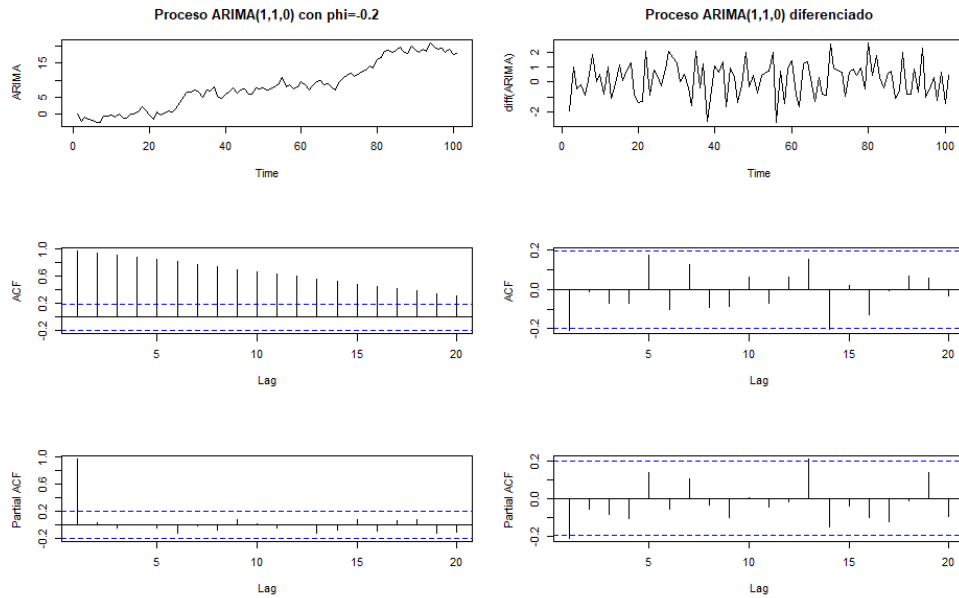


Figura 2.4: En la primera columna se representa una serie temporal generada a partir de un **ARIMA**(1,1,0) con $\phi = -0,2$. Presenta fuerte tendencia y la *fas* tarda mucho en anularse. En la segunda, la misma serie, pero diferenciada. Ya no presenta tendencia y las *fas* y *fap* permiten una identificación sencilla de un modelo **ARMA**(1,0).

2.3. Modelos estacionales: ARIMA estacionales

En este apartado presentaremos la última de las modificaciones de los modelos *ARIMA* que nos servirá para tener en cuenta patrones estacionales de nuestros procesos.

Si recordamos, los procesos *ARMA* modelizan la dependencia regular entre observaciones consecutivas en el pasado inmediato (*AR*(1) : $X_t = c + \phi_1 X_{t-1} + a_t$; *MA*(2) : $X_t = c + \theta_1 a_{t-1} + \theta_2 X_{t-2} + a_t$; etc.). Sin embargo ahora, los *ARMA* estacionales modelizan la dependencia estacional, es decir, la dependencia entre observaciones ocurridas en instantes múltiplos del período estacional s , también llamado frecuencia.

Definición 2.13. *Un proceso estacionario $\{X_t\}_{t \in T}$ será un **ARMA estacional puro**, $\text{ARMA}(P, Q)_s$, si admite la representación*

$$\begin{aligned} X_t = & c + \Phi_1 X_{t-s} + \Phi_2 X_{t-2s} + \dots + \Phi_P X_{t-Ps} + \\ & a_t + \Theta_1 a_{t-s} + \Theta_2 a_{t-2s} + \dots + \Theta_Q a_{t-Qs}, \end{aligned} \quad (2.3)$$

donde $c, \Phi_1, \dots, \Phi_P, \Theta_1, \dots, \Theta_Q$ son constantes y $\{a_t\}_{t \in T}$ ruido blanco.

Nuevamente, esta ecuación (2.3) puede ser escrita de forma compacta como

$$\Phi(B^s)X_t = c + \Theta(B^s)a_t,$$

donde $\Phi(B^s) = (1 - \Phi_1 B^s - \Phi_2 B^{2s} - \dots - \Phi_P B^{Ps})$, $\Theta(B^s) = (1 + \Theta_1 B^s + \Theta_2 B^{2s} + \dots + \Theta_Q B^{Qs})$, y B^s denota el operador retardo estacional definido como $B^s X_t = X_{t-s}$.

Ejemplo 2.14. Veamos un ejemplo de proceso *AR*(1)₃. La Figura 2.5 recoge el gráfico secuencial de un proceso *AR*(1)₃ junto con su *fas* y *fap*. Para detectar el orden estacional del modelo, debemos fijarnos únicamente en los retardos múltiplos de s . En este caso, vemos que la *fas* continúa grande en estos retardos para múltiplos continuados, mientras que la *fap* solo es grande para el primer múltiplo, es por ello que proponemos el ajuste *AR*(1) para la parte estacional.

¿Qué pasa si queremos buscar un modelo que nos ajuste un proceso que presenta tanto dependencia lineal como estacionalidad? Para ello, se presentan los *ARMA* estacionales multiplicativos. Combinando ambos modelos *ARMA* y *ARMA* estacionales, podremos modelizar conjuntamente dependencia regular y estacional.

Definición 2.15. *Un proceso $\{X_t\}_{t \in T}$ será **ARMA estacional multiplicativo**, $\text{ARMA}(p, q) \times (P, Q)_s$, si se puede modelizar a través de la representación*

$$\phi(B)\Phi(B^s)X_t = c + \theta(B)\Theta(B^s)a_t. \quad (2.4)$$

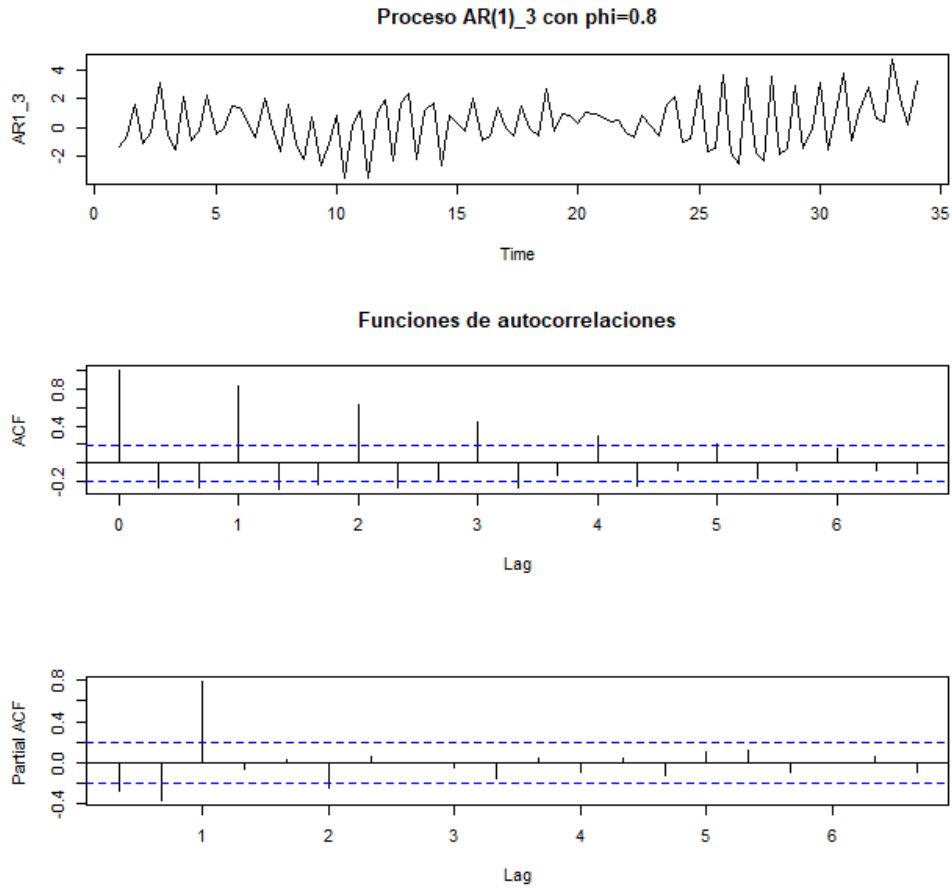


Figura 2.5: Simulación de un proceso $AR(1)_3$ con $\Phi = 0,8$.

Para entender un poco mejor la ecuación (2.4) desarrollémosla para un caso particular. Por ejemplo, para un $AR(1) \times MA(1)_{12}$ se tendría:

$$(1 - \phi_1 B)X_t = c + (1 + \Theta_1 B^{12})a_t \Leftrightarrow$$

$$X_t = c + \phi_1 X_{t-1} + a_t + \Theta_1 a_{t-12}.$$

O para un $MA(1) \times AR(1)_{12}$, sería:

$$(1 - \Phi_1 B^{12})X_t = c + (1 + \theta_1 B)a_t \Leftrightarrow$$

$$X_t = c + \Phi_1 X_{t-12} + a_t + \theta_1 a_{t-1}.$$

Observación 2.16. Para identificar los valores de p, q, P y Q de estos modelos, debemos fijarnos en las fas y fap atendiendo a que en los retardos bajos ($1, 2, \dots, \frac{s}{2}$), las fas y fap ofrecerán información sobre la parte regular del modelo, mientras que en los retardos múltiplos de la frecuencia ($s, 2s, \dots$) ofrecerán información sobre la parte estacional.

Ejemplo 2.17. Veamos lo expuesto arriba en un ejemplo simulado. La Figura 2.6 ilustra una serie de tiempo generada a partir de un $ARMA(0, 1) \times (1, 0)_{12}$. Como podemos observar, para los primeros retardos, la fap tarda más en anularse que la fas , por lo que proponemos un MA , y además el último retardo en la que no se anula es el primero, por lo que proponemos para la parte regular un $MA(1)$. Por otra parte, en los retardos múltiplos de la frecuencia, 12, vemos que es la fap la primera en anularse, teniendo solo un retardo no nulo. Es por ello que proponemos un $AR(1)_{12}$ para la parte estacional. Resulta así un $MA(1) \times AR(1)_{12}$.

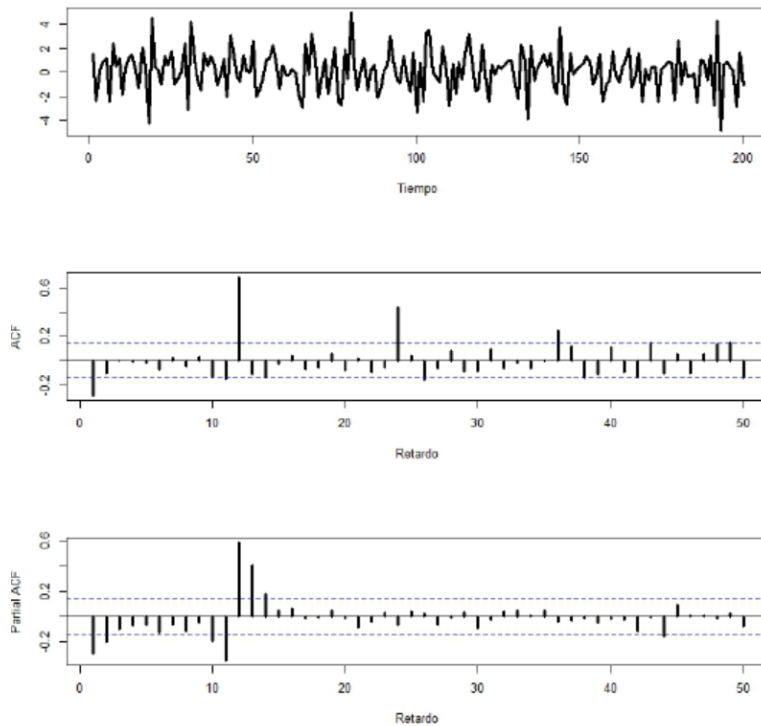


Figura 2.6: Proceso $MA(1) \times AR(1)_{12}$.

Pasamos ahora a hablar de la última clase de modelos de esta familia: los $ARIMA$ estacionales multiplicativos. Del mismo modo que los procesos $ARMA$ juegan un pa-

pel fundamental en la construcción de los procesos *ARIMA* (recordemos que $\{X_t\}_{t \in T}$ es *ARIMA*(p, d, q) si, y sólo si, $(1 - B)^d X_t$ es *ARMA*(p, q)), los *ARMA* estacionales serán decisivos para la construcción de los *ARIMA* estacionales.

Supongamos $\{X_t\}_{t \in T}$ no estacionario debido a una componente estacional S_t , es decir,

$$X_t = S_t + V_t,$$

con $\{V_t\}_{t \in T}$ estacionario. Dicha componente estacional S_t puede ser determinista ($S_t = S_{t-s}$) o estocástica ($S_t = S_{t-s} + a_t$). De cualquier modo, se ve que al aplicar una diferenciación estacional (B^s) a nuestro proceso, este pasa a ser estacionario, ya que

$$X_t - X_{t-s} = V_t - V_{t-s} \quad \text{ó} \quad X_t - X_{t-s} = a_t + V_t - V_{t-s},$$

ambos casos estacionarios como suma de procesos estacionarios. Este hecho nos lleva a la definición los *ARIMA* estacionales multiplicativos.

Definición 2.18. *Un proceso **ARIMA estacional multiplicativo**, denotado por **ARIMA**(p, d, q) \times (P, D, Q) $_s$, es aquel que, después de aplicarle d diferenciaciones regulares y D diferenciaciones estacionales de periodo s , se transforma en un proceso *ARMA*(p, q) \times (P, Q) $_s$. Análogamente a las definiciones anteriores, un proceso *ARIMA*(p, d, q) \times (P, D, Q) $_s$ es aquel que admite una representación del tipo*

$$\phi(B)\Phi(B^s)(1 - B)^d(1 - B^s)^D X_t = c + \theta(B)\Theta(B^s)a_t,$$

donde el polinomio $\phi(z)\Phi(z)$ no tiene raíces de módulo 1.

Para entender mejor esta expresión, desarrollémosla con un caso particular. Por ejemplo, para un *ARIMA*(1, 1, 1) \times (1, 1, 1) $_{12}$, la expresión se reduce a:

$$(1 - \phi_1 B)(1 - \Phi_1 B^{12})(1 - B)(1 - B^{12})X_t = c + (1 + \theta_1 B)(1 + \Theta_1 B^{12})a_t,$$

que equivale a

$$X_t = c + (1 + \phi_1)x_{t-1} - \phi_1 X_{t-2} + (1 + \Phi_1)X_{t-12} - (1 + \phi_1 + \Phi_1 + \phi_1 \Phi_1)$$

$$X_{t-13} + (\phi_1 + \phi_1 \Phi_1)X_{t-14} - \Phi_1 X_{t-24} + (\Phi_1 + \phi_1 \Phi_1)X_{t-25} - \phi_1 \Phi_1$$

$$X_{t-26} + a_t + \theta_1 a_{t-1} + \Theta_1 a_{t-12} + \theta_1 \Theta_1 a_{t-13}.$$

Como podemos observar, los *ARIMA* estacionales multiplicativos generalizan to-

dos los procesos que hemos estudiado en este capítulo:

$$ARIMA(p, 0, 0) \times (0, 0, 0)_1 = AR(p)$$

$$ARIMA(0, 0, q) \times (0, 0, 0)_1 = MA(q)$$

$$ARIMA(p, 0, q) \times (0, 0, 0)_1 = ARMA(p, q)$$

$$ARIMA(p, d, q) \times (0, 0, 0)_1 = ARIMA(p, d, q)$$

$$ARIMA(0, 0, 0) \times (P, 0, 0)_s = AR(P)_s$$

$$ARIMA(0, 0, 0) \times (0, 0, Q)_s = MA(Q)_s$$

$$ARIMA(0, 0, 0) \times (P, 0, Q)_s = ARMA(P, Q)_s$$

$$ARIMA(0, 0, 0) \times (P, D, Q)_s = ARIMA(P, D, Q)_s$$

$$ARIMA(p, 0, q) \times (P, 0, Q)_s = ARMA(p, q) \times (P, Q)_s$$

Estos modelos son capaces de capturar no estacionalidades provocadas por la presencia de tendencia y componente estacional y modelar la dependencia regular y estacional. Dichos modelos son, posiblemente, los más utilizados en la modelización de series de tiempo univariantes.

2.4. Estimación de los parámetros de un modelo AR-MA

Una vez identificado el modelo $ARMA(p, q)$ que puede haber generado nuestra serie de tiempo $\{X_t\}_{t \in T}$, debemos estimar los parámetros desconocidos de dicho modelo: $c, \phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q$ y σ_a^2 . Esta estimación se puede llevar a cabo de forma consistente por el método de Mínimos Cuadrados o por el método de Máxima Verosimilitud.

El método de Mínimos Cuadrados busca minimizar la suma de las innovaciones cuadradas. Dado un vector de parámetros $(\tilde{c}, \tilde{\phi}_1, \dots, \tilde{\phi}_p, \tilde{\theta}_1, \dots, \tilde{\theta}_q)$, se obtienen los residuos

$$\hat{a}_t = X_t - \tilde{c} - \tilde{\phi}_1 X_{t-1} - \dots - \tilde{\phi}_p X_{t-p} - \tilde{\theta}_1 \hat{a}_{t-1} - \dots - \tilde{\theta}_q \hat{a}_{t-q}.$$

Así, para minimizar la suma de cuadrados

$$S(\tilde{c}, \tilde{\phi}_1, \dots, \tilde{\phi}_p, \tilde{\theta}_1, \dots, \tilde{\theta}_q) = \sum_{t=1}^T \hat{a}_t^2,$$

será necesario tener un valor inicial para X_0, X_1, \dots, X_{p-1} y a_0, a_1, \dots, a_{q-1} . Como estos datos no se poseen en un inicio, impondremos las condiciones de que las p primeras observaciones de $\{X_t\}_{t \in T}$ son los valores de nuestra serie observada y que $a_0 = a_1 = \dots = a_{q-1} = 0$. Con estas condiciones, ya seremos capaces de minimizar la suma de cuadrados, obteniendo así la estimación por *Mínimos Cuadrados Condicionados*.

Si preferimos utilizar el método de Máxima Verosimilitud, debemos seguir un razonamiento distinto. Lo que debemos hacer es escribir la función de densidad conjunta y luego maximizarla respecto a los parámetros. Para escribir la densidad conjunta de las T observaciones de $\{X_t\}_{t \in T}$, que denotaremos con el vector $\mathbf{x} = (x_1, \dots, x_T)$, utilizaremos la propiedad de que la distribución conjunta de dos variables cualesquiera x, y puede escribirse como el producto de la distribución marginal de la primera variable y la distribución de la segunda condicionada a los valores de la primera, es decir,

$$f(x, y) = f(x)f(y|x).$$

Esto sigue siendo cierto si las hacemos condicionales a otra variable z , *i.e.*,

$$f(x, y|z) = f(x|z)f(y|x, z).$$

Por lo tanto, si tomamos $x = x_1$ y $y = x_2, \dots, x_T$, podemos escribir

$$f(\mathbf{x}) = f(x_1)f(x_2, \dots, x_T|x_1)$$

y descomponiendo el segundo término tomado ahora $z = x_1$, $x = x_2$ e $y = x_3, \dots, x_T$, obtenemos que

$$f(\mathbf{x}) = f(x_1)f(x_2|x_1)f(x_3, \dots, x_T|x_1, x_2).$$

Repetiendo este procedimiento iterativamente obtendremos al final que

$$f(\mathbf{x}) = f(x_1)f(x_2|x_1)f(x_3|x_2, x_1) \dots f(x_T|x_{T-1}, \dots, x_1).$$

Esta descomposición permite escribir la verosimilitud de un modelo *ARMA* si suponemos normalidad, ya que las distribuciones condicionales serán normales con parámetros $\mathbb{E}(x_t|x_{t-1}, \dots, x_1) = x_{t|t-1}$ y $\text{Var}(x_t|x_{t-1}, \dots, x_1) = \sigma_a^2 v_{t|t-1}$, siendo $\sigma_a^2 = \text{Var}(a_t)$ y $v_{t|t-1}$ las varianzas condicionadas. Con esta notación, la función de densidad conjunta de la muestra puede escribirse como

$$f(\mathbf{x}) = f(x_1) \prod_{t=2}^T \sigma^{-1} v_{t|t-1}^{-1/2} (2\pi)^{-1/2} \exp \left\{ -\frac{1}{2\sigma^2} \sum_{t=1}^T \frac{(x_t - x_{t|t-1})^2}{v_{t|t-1}} \right\},$$

donde hemos separado la densidad de la primera observación por que $x_{1|0}$ y $v_{1|0}$ dependen de las condiciones iniciales.

Vamos a llamar a_t a la diferencia entre los valores observados y sus esperanzas condicionadas, $a_t = x_t - x_{t-1}$. La función de verosimilitud exacta de un proceso *ARMA* es

$$L(c, \phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q, \sigma_a^2) = -\frac{T}{2} \log \sigma_a^2 - \frac{1}{2} \sum_{t=1}^T \log v_{t|t-1} - \frac{1}{2\sigma_a^2} \sum_{t=1}^T \frac{a_t^2}{v_{t|t-1}},$$

donde tanto las varianzas condicionales $v_{t|t-1}$ como los errores a_t dependen de los parámetros.

Por tanto, la estimación de los parámetros $c, \phi_1, \dots, \phi_p, \theta_1, \dots, \theta_q$ y σ_a^2 por medio del método de Máxima Verosimilitud se obtiene a través de los valores $(\tilde{c}, \tilde{\phi}_1, \dots, \tilde{\phi}_p, \tilde{\theta}_1, \dots, \tilde{\theta}_q, \tilde{\sigma}_a^2)$ que maximizan la función de verosimilitud L anterior.

2.5. Predicciones con modelos ARIMA

Al igual que la construcción de los procesos *ARIMA* viene dada por los *ARMA*, sus predicciones también. Es por ello que en esta sección nos centraremos en las predicciones de estos modelos, siendo análogas para todos los demás.

Supongamos que ya hemos ajustado un modelo *ARMA*(p, q) adecuado para la serie observada x_1, \dots, x_T . Lo que queremos hacer ahora es, dada una ventana h , predecir los siguientes h valores de nuestra serie, es decir, predecir los valores de x_{T+1}, \dots, x_{T+h} . Formalmente, a esta predicción se le denomina **predicción con origen T y horizonte h** y se denotará por $\hat{x}_T(h)$.

Para hacer esta lectura más intuitiva, comenzaremos explicando las predicciones en casos sencillos como el *AR*(1) y el *MA*(1), para luego generalizarlos para cualquier *ARMA*(p, q).

Empecemos con un *AR*(1). Como ya sabemos, la ecuación que rige este proceso es $X_t = c + \phi_1 X_{t-1} + a_t$, siendo c y ϕ_1 constantes conocidas y $\{a_t\}_{t \in T}$ ruido blanco. Por lo tanto, si quisiésemos predecir el valor de nuestra serie con origen T y horizonte 1, bastaría con obtener una predicción de la variable aleatoria

$$X_{T+1} = c + \phi_1 X_T + a_{T+1}.$$

Para ello, nos hace falta el valor de X_T que sí conocemos, x_T es un valor observado de nuestra serie, y una estimación del valor futuro a_{T+1} que, al pertenecer a un proceso de ruido blanco y no recoger la serie ninguna información sobre ella, podemos estimar como $\mathbb{E}(a_{T+1}) = 0$. Así, la predicción $\hat{x}_T(1)$ será

$$\hat{x}_T(1) = c + \phi_1 x_T,$$

donde todos los términos son conocidos.

Si quisiésemos estimar ahora una predicción con origen T y horizonte 2, deberíamos estimar la variable aleatoria

$$X_{T+2} = c + \phi_1 X_{T+1} + a_{T+2}.$$

Procediendo como antes, esta estimación, $\hat{x}_T(2)$, la podremos dar una vez conozcamos los valores de X_{T+1} y de a_{T+2} . Para el primer valor usaremos la estimación obtenida en el párrafo anterior, $\hat{x}_T(1) = c + \phi_1 x_T$, y nuevamente $\mathbb{E}(a_{T+2}) = 0$. Así,

$$\hat{x}_T(2) = c + \phi_1(c + \phi_1 x_T) = c + c\phi_1 + \phi_1^2 x_T.$$

Generalizando, podemos concluir que la predicción con origen T y horizonte h de un proceso $AR(1)$ será $\hat{x}_T(h) = c + c\phi_1 + c\phi_1^2 + \dots + c\phi_1^{h-1} + \phi_1^h x_T$.

Procediendo de igual modo, esta estimación puede generalizarse para cualquier orden finito del autorregresivo, $p > 0$. Además, se puede probar que las predicciones a largo plazo de futuros valores de un proceso $AR(p)$ es la media del proceso, es decir, $\hat{x}_T(h) \xrightarrow{h \rightarrow \infty} \mu$.

Analicemos ahora un $MA(1)$. En este caso, la ecuación que representa este proceso sabemos que es $X_t = c + a_t + \theta_1 a_{t-1}$, donde c y θ_1 son constantes conocidas y $\{a_t\}_{t \in T}$ ruido blanco. Por lo tanto, si queremos predecir el valor de nuestra serie con origen T y horizonte 1, tendríamos que predecir la variable aleatoria

$$X_{T+1} = c + a_{T+1} + \theta_1 a_T.$$

Ahora, los valores que necesitamos para nuestra estimación son a_{T+1} y a_T . El valor futuro a_{T+1} no lo tenemos ni podemos obtener información de él a partir de la serie, por lo tanto volvemos a estimarlo como $\mathbb{E}(a_{T+1}) = 0$. En cuanto al valor de a_T , si bien es cierto que no lo tenemos, nuestra serie sí que contiene cierta información sobre él. Podemos ajustarlo y obtener $\hat{a}_T = \hat{a}_T(0)$ siguiendo distintas metodologías:

1. Fijar un valor inicial para a_1 (por lo general, su media 0). Si consideramos la ecuación del proceso $X_t = c + a_t + \theta_1 a_{t-1}$ y despejamos a_t , obtenemos que $a_t = X_t - c - \theta_1 a_{t-1}$. Así, para obtener a_T a partir de la expresión anterior, bastaría con tener los valores a_i con $i = 1, \dots, T-1$. Por lo que si fijamos el valor a_1 podremos recursivamente predecir todos los demás valores. Es importante señalar que, cuando T es grande, la condición inicial ($a_1 = 0$) tiene poco impacto sobre los resultados finales.
2. Predecir a_T a través de “la mejor” combinación lineal de x_1, \dots, x_T . Los coeficientes de la combinación lineal de x_1, \dots, x_T que mejor predice a_T se pueden obtener a partir de la media y de las autocovarianzas de X_1, \dots, X_T y de sus covarianzas entre ellas y a_T .

Una vez obtenida la predicción de ese a_T (recordemos que lo que queríamos era obtener predicciones de $X_{T+1} = c + a_{T+1} + \theta_1 a_T$), basta con sustituir las estimaciones en dicha ecuación. Así, nos quedaría que

$$\hat{x}_T(1) = c + \theta_1 \hat{a}_T(0).$$

Si quisiésemos estimar ahora una predicción con origen T y horizonte 2, deberíamos estimar la variable aleatoria $X_{T+2} = c + a_{T+2} + \theta_1 a_{T+1}$. Procediendo como antes, esta estimación la podremos conseguir una vez tengamos valores para a_{T+2} y a_{T+1} . Sin embargo, ahora ya sí que no tenemos información en la serie para estimar ninguno de estos valores. Por lo tanto, tomamos $\mathbb{E}(a_{T+2}) = \mathbb{E}(a_{T+1}) = 0$. Así,

$$\hat{x}_T(2) = c.$$

Nuevamente, el procedimiento anterior puede generalizarse fácilmente para cualquier horizonte $h > 0$ y cualquier orden finito del proceso $q > 0$. Además, puede demostrarse que la predicción en un $MA(q)$ a horizontes mayores que q es la media del proceso, es decir, $\hat{x}_T(h) = \mu \quad \forall k > q$.

Los procedimientos explicados arriba para predecir valores futuros de procesos $AR(p)$ y $MA(q)$ pueden ser combinados para predecir valores futuros de procesos $ARMA(p, q)$. No vamos a escribir las expresiones explícitas para estos métodos porque en función de sus órdenes cada uno tiene una expresión particular y el texto se volvería demasiado denso de forma innecesaria. Sin embargo, es importante comentar que, como en los casos particulares anteriores, la predicción a largo plazo de futuros valores de un proceso $ARMA(p, q)$ es la media del proceso, es decir, $\hat{x}_T(h) \xrightarrow{h \rightarrow \infty} \mu$.

Como comentábamos antes, las predicciones de los $ARMA$ serán imprescindibles para las de los $ARIMA$. Para predecir un proceso $ARIMA$, se debe diferenciar la serie como ya hemos visto, hasta obtener un proceso estacionario $ARMA$. Luego, predecir los valores de dicho $ARMA$ y, finalmente, deshacer la diferenciación en estas predicciones para obtener predicciones del proceso original.

Por último, comentar que, como es habitual en estadística, a la hora de predecir o estimar un valor, incluso más importante que el valor en sí, puede ser la aportación de un intervalo de confianza que, en nuestro caso, llamaremos intervalo de predicción. Para su construcción, usaremos la distribución muestral del error de predicción e_T :

$$e_T(h) = X_{T+h} - \hat{X}_T(h).$$

Se verifica que, cuando T es suficientemente grande y $\{a_t\}_{t \in T}$ gaussiano, entonces dicha distribución se aproxima a una Normal de media 0 y varianza $\sigma_a^2(1 + \omega_1^2 + \dots + \omega_{h-1}^2)$, siendo ω_i los coeficientes de la representación $X_t = c + a_t + \omega_1 a_{t-1} + \omega_2 a_{t-2} + \dots$. Así, el intervalo de predicción con nivel de significación α será

$$\left(\hat{X}_T(h) - z_{1-\alpha} \sqrt{\sigma_a^2(1 + \omega_1^2 + \dots + \omega_{h-1}^2)}, \hat{X}_T(h) + z_{1-\alpha} \sqrt{\sigma_a^2(1 + \omega_1^2 + \dots + \omega_{h-1}^2)} \right).$$

2.6. Aplicación de los ARIMA a los datos de Estrella Galicia

En esta sección analizaremos la serie de tiempo de los últimos diez años procedente de las ventas mensuales de cerveza en la Zona B¹. Estudiaremos sus características, tales como la tendencia y la componente estacional, procederemos a ajustar un modelo ARIMA adecuado para ella y realizaremos sus predicciones para los meses venideros.

Para ello, seguiremos los pasos que se indican a continuación:

1. Presentación y transformación de las unidades de la serie.
2. Propuesta y ajuste del modelo ARIMA.
3. Comprobación de parámetros ajustados no significativamente distintos de cero e interpretación.
4. Validación del modelo ajustado
5. Conclusiones

Lo primero que haremos será leer los datos, crear el objeto serie y representar su gráfico secuencial (ver Figura 2.7). En él vemos claramente la presencia de estacionalidad e incluso una leve tendencia, pero, antes de eso, debemos fijarnos en que también apreciamos cierta heterocedasticidad en los últimos años. Es por ello que primero intentaremos transformar sus unidades para conseguir homocedasticidad.

Tras aplicar la función logarítmica, vemos cómo mejora nuestro problema. Sin embargo, sospechamos que igual podríamos usar una función transformadora mejor. Para ello, utilizamos el criterio de Box-Cox obteniendo una lambda óptima $\hat{\lambda} = 0,064$. Como podemos ver en la Figura 2.8 apenas hay diferencia entre ambas transformaciones. Por lo tanto, nos decantamos por trabajar con la serie logarítmica porque nos resultará más fácil interpretar los resultados más adelante.

Pasemos ahora a proponer y ajustar el modelo ARIMA. Tanto en su gráfico secuencial como en el gráfico de autocorrelaciones que se muestra en la Figura 2.9, podemos ver cómo claramente nuestra serie no es estacionaria debido, en primer lugar, a la tendencia que presenta. Recordemos que la tendencia se identificaba en el hecho de que la *fas* tardaba muchos retardos en anularse.

Necesitamos una diferenciación regular. Tras aplicarla, obtenemos los gráficos que mostramos en la Figura 2.10. La tendencia claramente ha desaparecido, pero ahora

¹Tanto la zona geográfica como los datos con los que trabajaremos son reales, pero hemos decidido no mencionarla por motivos de confidencialidad. Por este mismo motivo, todas las gráficas que puedan arrojar información sobre el volumen real de ventas de la compañía han sido “silenciadas” borrando su eje *y*.

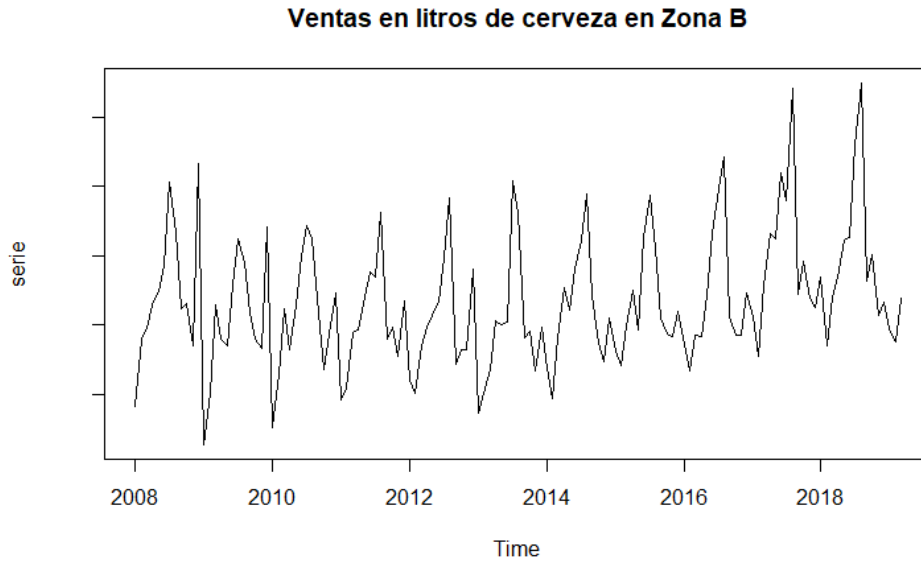


Figura 2.7: Gráfico secuencial de la serie original.

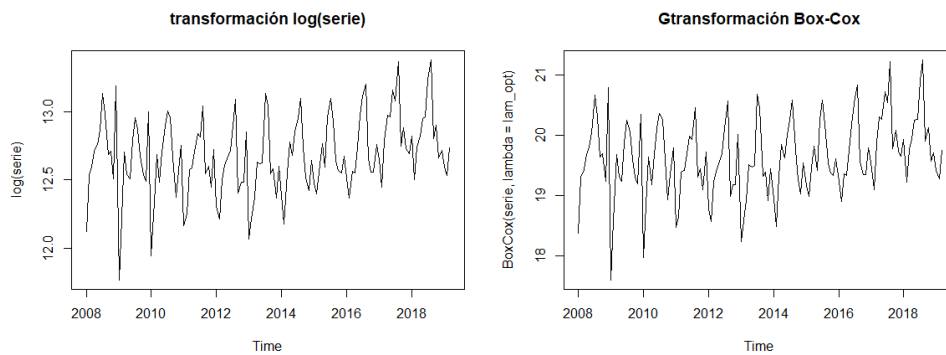


Figura 2.8: Serie transformada por logaritmo y por BoxCox, respectivamente.

podemos apreciar la componente estacional con periodo $s = 12$: los valores de la función de autocorrelaciones simples siguen un patrón repetitivo que se sale de las bandas. Necesitamos, por tanto, una diferenciación estacional.

Aplicando la nueva diferenciación estacional resultan los gráficos recogidos en la Figura 2.11. Si bien es cierto que la f_{as} no tiende a cero excesivamente rápido, podemos afirmar que sí cae dentro de las bandas, cosa que consideramos suficiente.

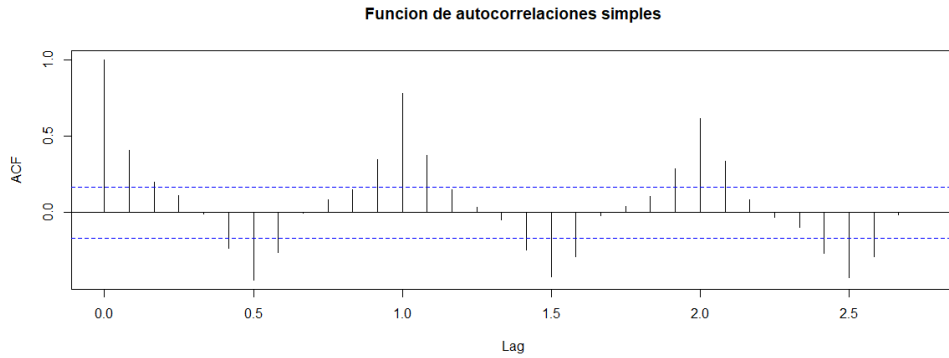
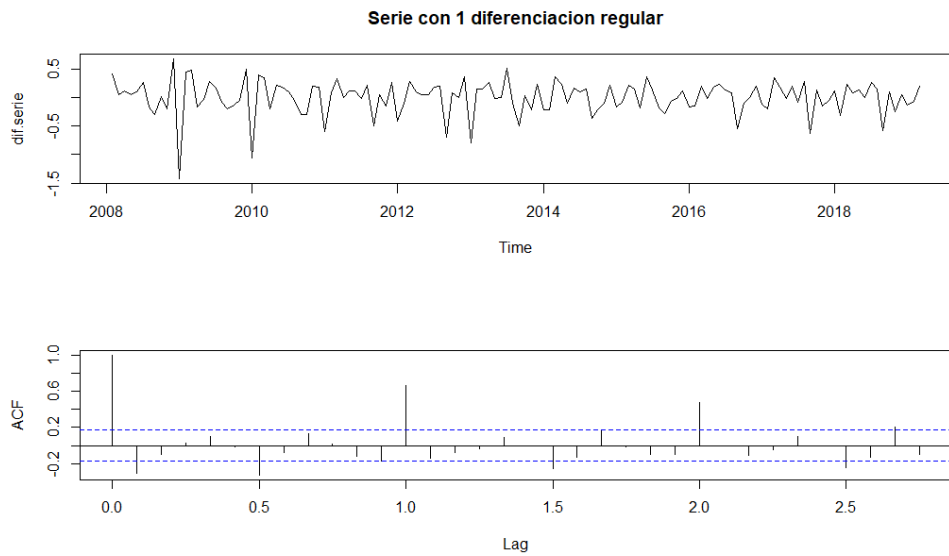


Figura 2.9: Funcion de autocorrelaciones simples.

Figura 2.10: Gráfico secuencial y *fas* de la serie diferenciada.

Concluimos por tanto que nuestra serie diferenciada ya es estacionaria.

Ya tenemos entonces la base de nuestro modelo: $ARIMA(p, 1, q) \times (P, 1, Q)_{12}$. Lo que debemos preguntarnos ahora es ¿cuáles son dichas p, q, P, Q ? Para ello, representaremos las *fas* y *fap* de nuestra serie diferenciada (ver Figura 2.12) e intentaremos proponer un modelo adecuado.

- Para proponer p y q , debemos fijarnos en los $12/2 = 6$ primeros retardos de las *fas* y *fap*:

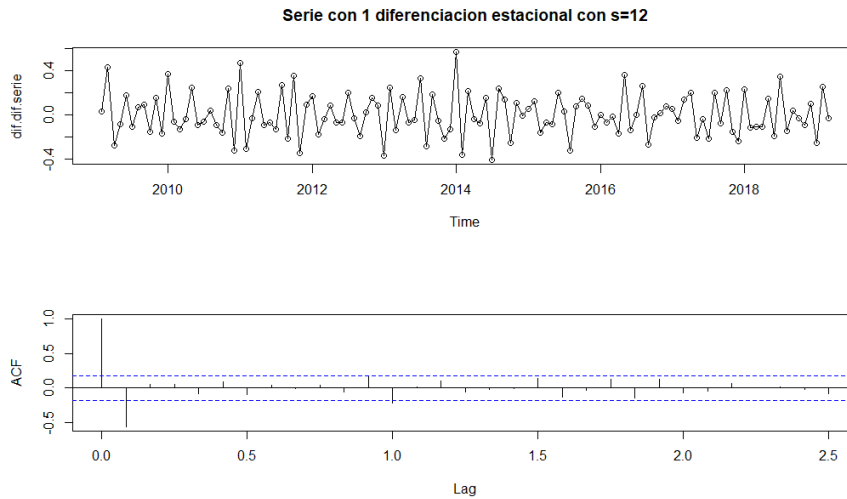


Figura 2.11: Gráfico secuencial y *fas* de la serie diferenciada.

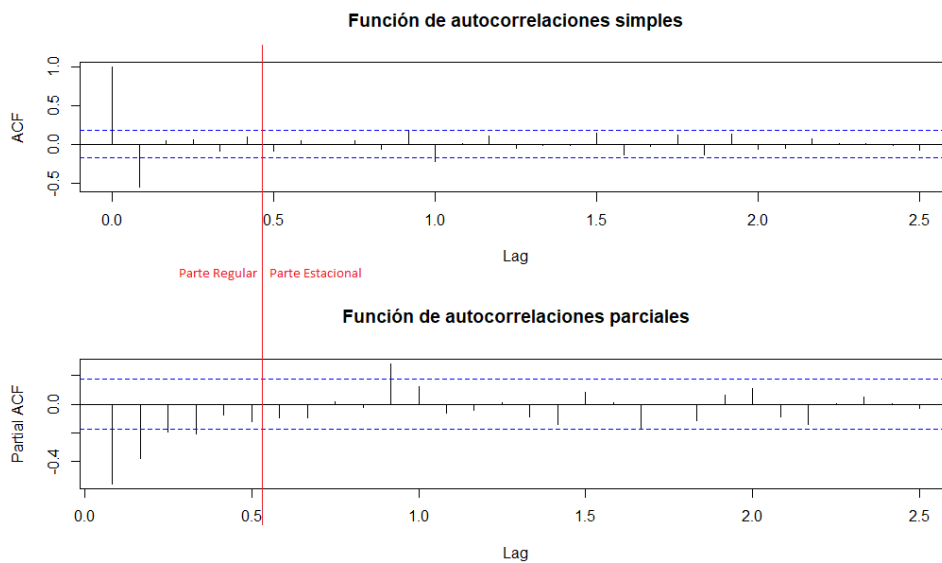


Figura 2.12: *fas* y *fap* de nuestra serie diferenciada.

- FAS: Vemos que la última en salirse es la correspondiente al segundo retardo. Esto lo relacionamos por tanto con un $MA(2)$.
- FAP: En este caso, la última en salir es la cuarta, si bien es cierto que tanto esta como la correspondiente al tercer retardo lo hacen por poco. De todos

modos, esto nos induciría a pensar en un $AR(4)$.

Como siempre pretendemos estimar con un modelo lo más sencillo posible se nos ocurre, por lo tanto, modelar la parte regular del proceso diferenciado mediante un $MA(2)$ o, lo que es lo mismo, mediante un $ARIMA(0, 1, 2)$ para la serie original.

- Para proponer P y Q , debemos fijarnos en las fas y fap en los retardos múltiplos de $s = 12$. Aunque en el gráfico no se aprecia muy bien, la fas está muy próxima a anularse ya en el segundo múltiplo (saliéndose de bandas solo en el primer retardo $r = 12$), mientras que las fap , aún estando entre bandas, vemos que se mantienen más altas a lo largo del tiempo. Proponemos, por tanto, un modelo $ARIMA(0, 1, 1)_{12}$ para la componente estacional.

Así, el modelo que ajustaríamos para la serie logarítmica sería el modelo: $ARIMA(0, 1, 2) \times (0, 1, 1)_{12}$. Usando la función `auto.arima` del paquete `forecast`, se nos sugiere que utilicemos un $ARIMA(0, 1, 2) \times (0, 1, 1)_{12}$. Como es el que nosotros habíamos razonado, ajustaremos definitivamente dicho modelo, resultando las siguientes estimaciones de los parámetros: $c = 0,001$, $\theta_1 = -1,0848$, $\theta_2 = 0,2483$ y $\Theta_1 = -0,4966$. Como podemos ver en la salida de R (ver Figura 2.13), todos son significativamente distintos de cero a excepción de la constante c (era de esperar al tratarse de la serie diferenciada):

```

Coefficients:
      ma1      ma2      sma1  intercept
-1.0848  0.2483 -0.4966    0.0010
s.e.    0.0937  0.0973  0.0860    0.0011
      Parte regular  Parte estac.  Constante
sigma^2 estimated as 0.01552:  log likelihood = 78.51,  aic = -147.01

Training set error measures:
      ME      RMSE      MAE      MPE      MAPE      MASE      ACF1
Training set 0.008695156 0.1245831 0.09874963 83.13016 144.2919 0.3588513 0.03129658

```

Figura 2.13: Valores estimados de los coeficientes del modelo ARIMA y valores de sus criterios de información.

Así:

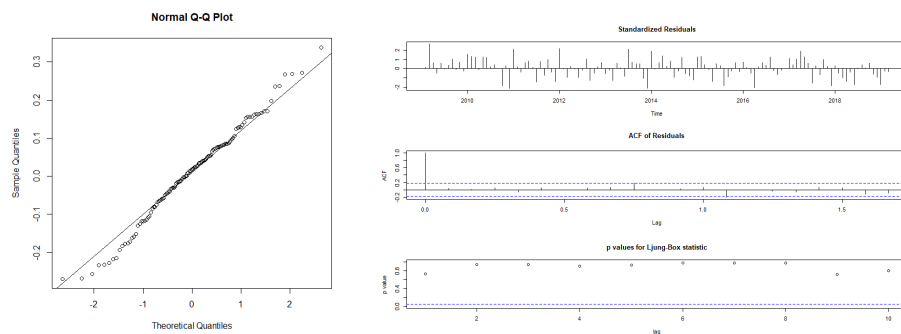
$$Y_t = Y_{t-1} + Y_{t-12} - Y_{t-13} + a_t + \theta_1 a_{t-1} + \theta_2 a_{t-2} + \Theta_1 a_{t-12} + \theta_1 \Theta_1 a_{t-13} + \theta_2 \Theta_1 a_{t-14},$$

que se reduce a

$$Y_t = Y_{t-1} + Y_{t-12} - Y_{t-13} + a_t - 1,0848a_{t-1} + 0,2483a_{t-2} - 0,4966a_{t-12} + 0,5387a_{t-13} - 0,1233a_{t-14}.$$

Ahora, es imprescindible chequear que los residuos $\{a_t\}_{t \in T}$ de este ajuste pasan la diagnosis. Lo más importante es comprobar si nuestras innovaciones pueden ser ruido blanco y luego, en segunda medida, ver si son gaussianas. Cabe destacar que estos contrastes deberían efectuarse sobre las innovaciones, pero como estas no son visibles (o conocidas), se los aplicamos a los residuos (que son sus estimaciones).

Aplicando la función `tsdiag` podemos aceptar la independencia y, por tanto, también la incorrelación. Al mismo tiempo, a la vista del QQ-plot y gracias al test de Shapiro-Wilk, no rechazamos la hipótesis de normalidad. Podemos aceptar, por tanto, que tenemos ruido blanco (ver Figuras 2.14(a) y 2.14(b)).



(a) Gráfico QQ-plot.

(b) Diagnosis `tsdiag`.

Figura 2.14: Diagnosis de residuos.

Por último, una vez poseemos un ajuste válido, debemos hacer las predicciones. Para ello, lo que se hará será calcular una predicción de los residuos futuros (teniendo en cuenta el ARIMA que siguen) y, luego, deshacer las distintas transformaciones que hemos aplicado a nuestra serie; es decir, la diferenciación estacional, luego la regular y, por último, la transformación logarítmica.

Utilizando la función `forecast` del paquete homónimo obtenemos las predicciones recogidas en el Cuadro 2.1, representadas gráficamente en la Figura 2.15, donde también se recogen los intervalos de confianza con nivel de confianza del 80% y 95%.

	Predicción	Inf 80	Sup 80	Inf 95	Sup 95
Abr 2019	341033.56	290923.87	395625.89	267218.40	444820.12
May 2019	364999.60	306957.67	423818.22	282872.23	474877.83
Jun 2019	407928.39	346974.78	472539.65	318947.77	528971.09
Jul 2019	477593.77	404620.22	554397.19	373575.35	620225.03
Ago 2019	557611.48	470799.58	648738.81	433148.35	728945.58
Sep 2019	317869.03	266770.47	372000.41	247458.38	411478.05
Oct 2019	339009.36	286282.40	395195.62	261960.49	438161.20
Nov 2019	285484.07	239407.67	334812.43	218981.42	370067.07
Dic 2019	304517.90	256179.31	355803.72	233705.04	396838.87
Ene 2020	283693.74	239887.53	333659.16	218106.22	367975.73
Feb 2020	244930.91	204738.38	290143.81	185618.76	320719.76
Mar 2020	308655.07	258962.82	362841.56	235651.27	403939.32

Cuadro 2.1: Predicciones de litros de cerveza que se venderán en la Zona B. Estos valores han sido transformados por un parámetro α para respetar la confidencialidad de los datos reales.

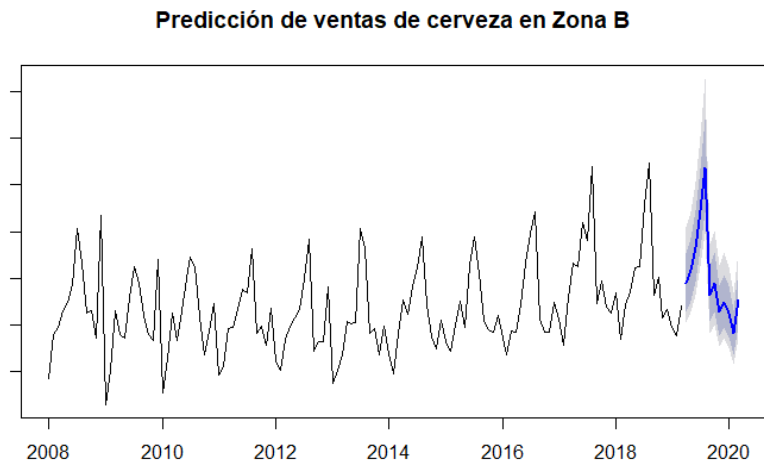


Figura 2.15: Gráfico secuencial de la serie “litros de cerveza vendidos en la Zona B” junto con sus predicciones para los 12 meses siguientes.

Capítulo 3

Suavizado Exponencial

En este nuevo capítulo abordaremos el otro gran enfoque clásico en el estudio de las series temporales: el Suavizado Exponencial. Como veremos, dicho enfoque se basa en la idea de descomposición de las series de tiempo en sus 3 componentes principales: tendencia (T), estacionalidad (S) y errores (E). En función de cómo estas se definan y combinen, podremos dar una clasificación de los distintos métodos resultantes y estudiar las propiedades de cada uno de ellos para obtener las predicciones pertinentes. En este capítulo, seguiremos muy de cerca la referencia [Hyndman et al. \(2008\)](#).

Antes de empezar, es importante distinguir entre *método de predicción* y *modelo estadístico*. Por una banda, un método es un algoritmo que nos proporciona una predicción puntual, es decir, un valor concreto que predice la evolución futura de nuestra serie. Sin embargo, un modelo estadístico nos define un proceso estocástico que podremos utilizar para definir toda una distribución de probabilidad para valores futuros de nuestra serie. La ventaja de los métodos es que nos permiten obtener una predicción de forma más rápida y sencilla que los modelos, mientras que los segundos, aunque son más costosos, también nos proveerán de intervalos de confianza para un nivel de significación dado. De todos modos, como veremos más adelante, en el Suavizado Exponencial ambos conceptos están íntimamente ligados.

En resumen, el Suavizado Exponencial, en su idea más primitiva, es un método de predicción. Se caracteriza por calcular las nuevas predicciones como combinaciones ponderadas de las observaciones pasadas, donde las observaciones más recientes reciben más peso que las antiguas. Esta disminución de peso muchas veces se realiza de forma exponencial, de ahí el nombre del método.

3.1. Descomposición de series temporales

Como ya sabremos en este punto del trabajo, es muy común pensar en las series temporales como una combinación de una tendencia (T), una estacionalidad (S) y un

error “aleatorio” (E), suponiendo que otras características como la heterocedasticidad ya no están presentes (bastaría con arreglar previamente la serie como hemos discutido en la sección 1.2).

Estas tres componentes principales se pueden combinar de diferentes formas para ajustar nuestra serie x : puramente aditiva como

$$x = T + S + E,$$

puramente multiplicativa como

$$x = T \times S \times E,$$

o cualquier otra combinación más compleja entre estas dos, como por ejemplo

$$x = (T + S) \times E.$$

Así, podemos definir, por ejemplo, la **serie ajustada estacionalmente**, y , como la resultante de extraer la componente estacional a la serie original, dejando únicamente las componentes tendencia y error. Para modelos aditivos, bastará con calcular $y = x - S$, mientras que para los multiplicativos será $y = x/S$.

3.2. Clasificación de los métodos de Suavizado Exponencial

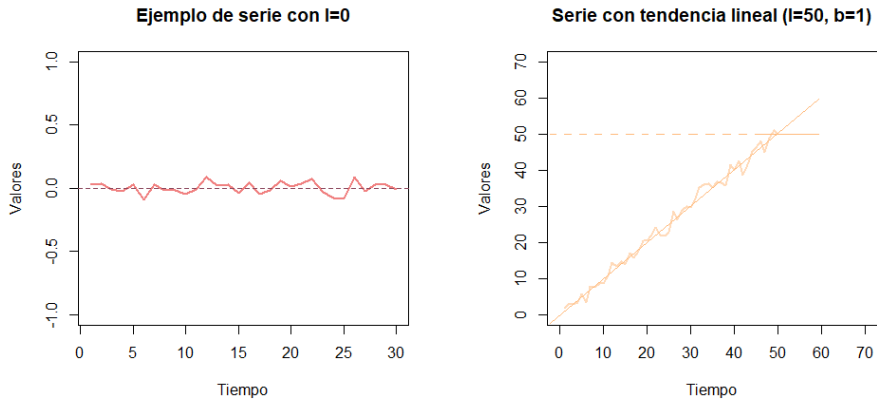
La clasificación de nuestros modelos se basará en cómo definamos cada componente de la descomposición. Comenzaremos por la tendencia y luego abordaremos la estacionalidad.

La tendencia, T , será una combinación de lo que llamamos *nivel* (l) y *crecimiento* o *pendiente* (b), parámetros que luego deberemos estimar. Su significado es justo lo que todos intuimos por su nombre, pero se aclara en la Figura 3.1. En función de como combinemos ambos parámetros, podremos obtener cinco tipos de pendientes futuras. Sea T_h la pendiente predicha para h tiempos futuros, y ϕ el parámetro de suavización, con $0 < \phi < 1$. Así, podemos definir los siguientes tipos de tendencia:

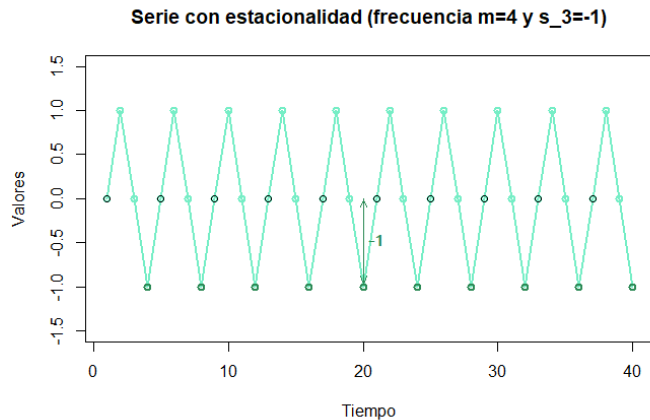
- Nula (N): $T_h = l$. No hay tendencia futura esperada. La evolución de la serie se espera que sea una oscilación en torno al nivel horizontal l .
- Aditiva (A): $T_h = l + bh$. Ahora sí hay un crecimiento esperado, con pendiente b , de forma lineal.
- Aditiva suavizada (A_d): $T_h = l + (\phi + \phi^2 + \dots + \phi^h)b$. Es una versión ajustada de la anterior. En este caso, como $0 < \phi < 1$, nos indica que la pendiente de la tendencia cada vez será más pequeña, tendiendo finalmente a una horizontal.

3.2. CLASIFICACIÓN DE LOS MÉTODOS DE SUAVIZADO EXPONENCIAL 35

- Multiplicativa (M): $T_h = lb^h$.
- Multiplicativa suavizada (M_d): $T_h = lb^{(\phi + \phi^2 + \dots + \phi^h)}$.



(a) Serie sin tendencia con nivel constante $l = 0$. Todos los valores de esta serie fluctúan entorno al valor cero. (b) Serie con tendencia aditiva $T = l + bh$, donde podemos observar que en la última observación tenemos un nivel $l = 50$ y una pendiente constante $b = 1$.



(c) Serie periódica con frecuencia $m = 4$ en la que podemos observar perfectamente la componente estacional para cada "estación": $s_0 = 0$, $s_1 = 1$, $s_2 = 0$, $s_3 = -1$.

Figura 3.1: Gráficos ilustrativos sobre los elementos no observables de una serie.

Una vez escogido el tipo de tendencia (T) de nuestro método, debemos añadirle ahora la componente estacional (S) de forma aditiva ($T + S$) o multiplicativa ($T \times S$),

como hemos explicado en la sección anterior. Dicha estacionalidad estará completamente definida, a diferencia de la tendencia, por un único parámetro: el elemento estacional s .

Finalmente, deberíamos hacer lo mismo con los errores, aunque la forma de añadirlos (aditiva o multiplicativamente) no crea ninguna diferencia en las predicciones puntuales obtenidas, por lo que habitualmente son ignorados en la definición de los métodos (pero no en la definición de los modelos estadísticos que veremos luego). De este modo, se define el Cuadro 3.1, donde se recogen las quince combinaciones/métodos posibles.

Tendencia \ Estacionalidad	N (ninguna)	A (aditiva)	M (multiplic.)
N (ninguna)	N,N	N,A	N,M
A (aditiva)	A,N	A,A	A,M
A_d (aditiva suavizada)	A_d,N	A_d,A	A_d,M
M (multiplicativa)	M,N	M,A	M,M
M_d (Multiplicativa suavizada)	M_d,N	M_d,A	M_d,M

Cuadro 3.1: Clasificación de los métodos de suavizado exponencial en referencia al tipo de tendencia y estacionalidad.

Algunas de estas combinaciones son conocidas con nombre propio. Por ejemplo, el método (N,N) es el llamado *método de suavizado exponencial simple*, *SES*. El método (A,N) se conoce como *método lineal de Holt*, mientras que las combinaciones (A,A) y (A,M) son los famosos métodos de *Holt-Winter adictivo* y *Holt-Winter multiplicativo*, respectivamente.

Para cada uno de estos quince métodos, existen dos modelos state space posibles que resultan al añadir la componente de los errores (E) de forma aditiva o multiplicativa. Como comentábamos hace un momento, si se utilizan las mismas estimaciones de los parámetros, ambos modelos devolverán la misma predicción puntual, pero distintos intervalos de predicción. De aquí nace la gran importancia de distinguir entre ellos.

3.3. Ecuaciones explícitas de los métodos de Suavizado Exponencial

Ahora que ya conocemos la estructura de los distintos métodos, debemos estudiar las ecuaciones explícitas en cada uno de ellos para que, en primer lugar, nos definan sus elementos no observables y, en segundo lugar, nos proporcionen una ecuación “global” que nos sirva para calcular nuevas predicciones. Llamamos **vector de elementos no observables** en el instante t , \mathbf{v}_t , al vector formado por los elementos: nivel (l), pendiente (b) y estacionalidad (s) en el instante t , *i.e.*, $\mathbf{v}_t = (l_t, b_t, s_t)$. Así, cada método tendrá una ecuación que defina cada elemento no observable que intervenga en él y una última que las englobe siguiendo el esquema del método en cuestión. Por ejemplo, para el método (A,A), una vez dadas las ecuaciones que nos definan l , b y s , tendremos una cuarta del estilo $(l + bh) + s$; o, por ejemplo, para el método (A,M) será del estilo $(l + bh)s$.

Como podremos ver más adelante, las ecuaciones siguen el mismo patrón en todos los métodos, por lo que empezaremos desarrollándolas y explicándolas para alguno de los métodos más famosos y así luego será inmediato entenderlas para todos los demás.

Suavizado Exponencial Simple (N,N)

Comencemos con el método más simple de todos, el (N,N) o Suavizado Exponencial Simple. Como ya sabemos, las siglas de este método nos indican que tiene una tendencia tipo “nula” y que carece de componente estacional ($x = T$, siendo T de tipo “nula”).

Si recordamos, la tendencia tipo “nula” era aquella en la que no había un cambio de nivel esperado a lo largo del tiempo, y se expresaba como $T = l$. Por lo tanto, en este caso solo interviene el elemento no observable nivel (l). Como además no existe componente estacional, nuestro método se definirá con tan solo dos ecuaciones: una que nos definirá el nivel y otra que nos definirá las predicciones (la que denominábamos antes como “global”).

El nivel de nuestra serie en el instante t , l_t , lo calcularemos como una combinación convexa entre el nivel del instante anterior, l_{t-1} , y la observación real que se acabó dando en el instante t , para todo $t = 1, \dots, T$:

$$l_t = \alpha x_t + (1 - \alpha)l_{t-1},$$

siendo α un parámetro entre 0 y 1 que luego habrá que estimar. Por decirlo de algún modo, lo que hacemos es tomar el nivel que tenía la serie hasta el instante $t - 1$ y lo corregimos levemente con la nueva observación añadida.

Ahora, como ya tenemos definida la tendencia (T) y no existe componente estacional, es lógica la definición de las predicciones futuras, dada por

$$\hat{x}_T(h) = l_T.$$

Cabe destacar que para el cálculo de l_T debemos proceder iterativamente del siguiente modo:

$$\begin{aligned}
 \hat{x}_T(h) = l_T &= \alpha x_T + (1 - \alpha)l_{T-1} \\
 &= \alpha x_T + (1 - \alpha)[\alpha x_{T-1} + (1 - \alpha)l_{T-2}] \\
 &= \alpha x_T + \alpha(1 - \alpha)x_{T-1} + (1 - \alpha)^2 l_{T-2} \\
 &= \dots \\
 &= \alpha x_T + \alpha(1 - \alpha)x_{T-1} + \alpha(1 - \alpha)^2 x_{T-2} + \dots + \\
 &\quad + \alpha(1 - \alpha)^{T-1} x_1 + (1 - \alpha)^T l_0.
 \end{aligned}$$

Este desarrollo nos manifiesta dos puntos: el primero es que la predicción de valores futuros es calculado como una media ponderada de todas las observaciones pasadas con sus pesos decreciendo exponencialmente (de ahí el nombre de Suavizado Exponencial) y, el segundo punto, es que para poder llevar a cabo estos cálculos necesitamos un valor inicial l_0 , así como una estimación del parámetro α . Este problema que afecta a todos los métodos se conoce como *problema de inicialización* y lo discutiremos en la Sección 3.5.

Método lineal de Holt (A,N)

Pasemos ahora a explicar un método en el que tengamos un elemento no observable más. Por ejemplo, la pendiente o crecimiento b . Para ello, analizaremos el método (A,N) o Método lineal de Holt. Como sus siglas explican, este método tiene una tendencia de tipo “aditivo” y carece de estacionalidad.

Como recordaremos, la tendencia tipo “aditiva” era aquella que aumentaba linealmente con pendiente b , siguiendo la expresión $T_h = l + hb$. Por lo tanto, en este caso, a diferencia del anterior, intervienen dos elementos no observables: el nivel (l) y la pendiente (b). Como no disponemos de componente estacional, el Método lineal de Holt quedará definido por 3 ecuaciones: una que nos definirá el nivel, otra que nos definirá la pendiente y una última que las englobe para obtener nuevas predicciones.

Nuevamente, el nivel de nuestra serie en el instante t , l_t , lo calcularemos como una combinación convexa entre la observación en el instante t (x_t) y el nivel que teníamos en el instante anterior (l_{t-1}), pero ahora añadiendo el efecto de la pendiente b_{t-1} , resultando

$$l_t = \alpha x_t + (1 - \alpha)(l_{t-1} + b_{t-1}),$$

para todo $t = 1, \dots, T$ y con $\alpha \in [0, 1]$. Esto lo podemos interpretar como que, para el instante t , el nivel que vamos a tomar es el dado por la etapa anterior ($l_{t-1} + 1 \cdot b_{t-1}$), pero levemente corregido por la observación real que se acabó registrando en el instante t .

En cuanto a la pendiente de la tendencia en el instante t , b_t , se calculará como una combinación convexa entre la pendiente en el instante anterior y la diferencia de niveles que finalmente se produjo entre dichos instantes, es decir,

$$b_t = \beta(l_t - l_{t-1}) + (1 + \beta)b_{t-1},$$

siendo β un parámetro entre 0 y 1.

Así, teniendo ya definidos nuestros elementos no observables, concluimos con la última ecuación para el cálculo de nuevas predicciones:

$$\hat{x}_T(h) = l_T + hb_T.$$

Método de Holt-Winter aditivo (A,A)

Por último, consideremos un método en el que intervengan todos los elementos no observables posibles: el nivel (l), la pendiente (b) y la estacionalidad (s). Para ello, consideremos el método (A,A) o Método de Holt-Winter aditivo, en el que la tendencia es de tipo “aditivo”, al igual que la estacionalidad.

En este método, el nivel de la serie en el instante t se calcula como en el caso anterior, pero ahora, como existe estacionalidad, las observaciones serán ajustadas estacionalmente. Recordemos que, como en este caso la estacionalidad es añadida de modo aditivo, la serie ajustada estacionalmente es el resultado de la transformación $x - S$. La importancia de esta transformación se puede ver claramente con el ejemplo anterior de la Figura 3.1(c), ya que en ese caso, si no ajustásemos el valor de las observaciones, el nivel de la serie estaría variando continuamente al son del seno/coseno, cuando en realidad entendemos que esa fluctuación es debida a la estacionalidad y que el nivel de la serie es constante a cero. Por ello, en el caso de métodos con estacionalidad, todas las ecuaciones presentan las observaciones ajustadas estacionalmente (ya sea sustrayendo el elemento no observable, s , para estacionalidades aditivas o dividiéndolo para estacionalidades multiplicativas). Así,

$$l_t = \alpha(x_t - s_{t-m}) + (1 - \alpha)(l_{t-1} + b_{t-1}),$$

siendo m la frecuencia de observación de la serie y α un parámetro entre 0 y 1.

En cuanto a la pendiente de la tendencia, como seguimos con tendencia “aditiva”, la expresión que la define sigue siendo la misma que en el caso anterior y su interpretación también es igual:

$$b_t = \beta(l_t - l_{t-1}) + (1 + \beta)b_{t-1},$$

siendo β un parámetro entre 0 y 1.

Nos falta, por tanto, la ecuación referente a la componente no observable de la estacionalidad, s . El cálculo de la componente estacional en el instante t se calculará como una combinación convexa entre la componente estacional de ese instante en el “año” anterior, s_{t-m} , y la diferencia finalmente observada entre el valor de la serie y el valor esperado dado por la tendencia (que recordemos no tiene en cuenta la estacionalidad), $x_t - l_{t-1} - b_{t-1}$. Es decir,

$$s_t = \gamma(x_t - l_{t-1} - b_{t-1}) + (1 - \gamma)s_{t-m},$$

siendo γ un parámetro entre 0 y 1.

Finalmente, una vez definidos todos los elementos no observables que intervienen en el método, la ecuación “global” quedaría como sigue:

$$\hat{x}_T(h) = l_T + hb_T + s_{t-m+h_m^+},$$

siendo $h_m^+ = [(h-1) \bmod m] + 1$, que podemos interpretar como el número de “años” enteros que han pasado durante el tiempo h .

Descripción de todos los métodos del Suavizado Exponencial

Ahora que ya hemos vistos estos casos, seremos capaces de interpretar las ecuaciones de los demás métodos. Solo hay que tener en cuenta qué tipo de tendencia utiliza nuestro método y, en caso de existir componente estacional, recordar que debemos ajustar nuestra serie estacionalmente. En el Cuadro A.1 del Apéndice A se recogen las ecuaciones explícitas de todos los métodos de Suavizado Exponencial.

3.4. Modelos State Space

Como comentábamos al principio del capítulo, de cada *método* que hemos visto en la sección anterior podemos crear dos *modelos* distintos: uno añadiendo errores de forma aditiva y otro de forma multiplicativa. En la notación estándar, para distinguir entre estos dos modelos, se añadirá una nueva inicial a las siglas, de modo que los modelos con errores aditivos incluirán la inicial A , mientras que los multiplicativos tendrán la inicial M . Así, los State Space Models se definirán como un triplete (E,T,S) correspondiendo con la naturaleza de cada una de las tres componentes de la serie. Por ejemplo, el modelo ETS(A,M,N) será el modelo con errores aditivos, tendencia multiplicativa y sin estacionalidad; o el modelo ETS(M,A_d,M) será el modelo con errores multiplicativos, tendencia aditiva suavizada y estacionalidad multiplicativa.

Una vez tengamos definido un modelo, podremos estudiar la distribución de probabilidad para valores futuros de nuestra serie y calcular, por ejemplo, la media condicionada de valores futuros dados los últimos elementos no observables de los que disponemos, v_t . Esta media la denotaremos como $\mu_t(h) = \mathbb{E}(x_{t+h}|v_t)$, donde v_t es

el vector de elementos no observables l_t, b_t y s_t . Para muchos modelos, esta media condicional coincidirá con la predicción puntual que proporcionaba el método del que subyacen, es decir, $\mu_t(h) = \hat{x}_t(h)$. Sin embargo, para los modelos con tendencia o estacionalidad multiplicativa, la media condicional y la predicción puntual diferirán levemente para $h \geq 2$. Por ello, en esta sección, estudiaremos las ecuaciones que definen todos estos modelos y repasaremos sus propiedades, procediendo como en la sección anterior: comenzaremos con unos casos sencillos para luego presentar todos los demás.

State Space Models para el Método lineal de Holt (A,N)

Empecemos con el Método lineal de Holt o (A,N). Como ya sabemos, de él surgen dos modelos distintos: uno con errores aditivos y otro con errores multiplicativos.

Modelo con errores aditivos: ETS(A,A,N) Sea $\mu_t = \hat{x}_t = l_{t-1} + b_{t-1}$ la predicción a un paso de x_t , suponiendo que conocemos los valores de todos los parámetros que necesitamos. Sea también $\varepsilon_t = x_t - \mu_t$ el error de predicción a un paso en el instante t . Si recordamos de la sección anterior, las ecuaciones del Método lineal de Holt son:

$$l_t = \alpha x_t + (1 - \alpha)(l_{t-1} + b_{t-1}),$$

$$b_t = \beta(l_t - l_{t-1}) + (1 + \beta)b_{t-1},$$

$$\hat{x}_T(h) = l_T + hb_T.$$

Por lo tanto, siguiendo la tercera ecuación se tiene que

$$x_t = l_{t-1} + b_{t-1} + \varepsilon_t,$$

y usando las dos primeras podemos escribir

$$l_t = l_{t-1} + b_{t-1} + \alpha\varepsilon_t, \text{ y}$$

$$b_t = b_{t-1} + \beta(l_t - l_{t-1} - b_{t-1}) = b_{t-1} + \alpha\beta\varepsilon_t.$$

De aquí en adelante escribiremos $\bar{\beta} = \alpha\beta$. Las tres ecuaciones anteriores construyen uno de los State Space Models subyacentes del Método lineal de Holt.

Matricialmente, podemos reescribirlas considerando $v_t = (l_t, b_t)'$ como sigue:

$$\begin{aligned} x_t &= \begin{pmatrix} 1 & 1 \end{pmatrix} v_{t-1} + \varepsilon_t, \\ v_t &= \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} v_{t-1} + \begin{pmatrix} \alpha \\ \beta \end{pmatrix} \varepsilon_t. \end{aligned}$$

El modelo quedará completamente especificado cuando demos una distribución para los errores, ε_t . Normalmente, se asume que son independientes e idénticamente distribuidos, siguiendo una distribución Normal de media cero y varianza σ^2 .

Modelo con errores multiplicativos: ETS(M,A,N) El modelo con errores multiplicativos sigue la misma filosofía. Solo hay que tener en cuenta en este caso que ε_t es un error relativo definido por $\varepsilon_t = (x_t - \mu_t)/\mu_t$. Así, siguiendo el mismo procedimiento que antes, obtenemos las siguientes ecuaciones:

$$\begin{aligned} x_t &= (l_{t-1} + b_{t-1})(1 + \varepsilon_t), \\ l_t &= (l_{t-1} + b_{t-1})(1 + \alpha\varepsilon_t), \\ b_t &= b_{t-1} + \bar{\beta}(l_{t-1} + b_{t-1})\varepsilon_t. \end{aligned}$$

Equivalentemente,

$$\begin{aligned} x_t &= \begin{pmatrix} 1 & 1 \end{pmatrix} v_{t-1}(1 + \varepsilon_t), \\ v_t &= \begin{pmatrix} 1 & 1 \\ 0 & 1 \end{pmatrix} v_{t-1} + \begin{pmatrix} 1 & 1 \end{pmatrix} v_{t-1} \begin{pmatrix} \alpha \\ \beta \end{pmatrix} \varepsilon_t, \end{aligned}$$

asumiendo nuevamente que $\varepsilon_t \sim \mathcal{N}(0, \sigma^2)$.

State Space Models para todos los métodos

Siguiendo el mismo procedimiento que el explicado para los modelos subyacentes del Método lineal de Holt, podemos acabar obteniendo las ecuaciones de cada uno de los 30 modelos totales que podemos crear. Las ecuaciones se recogen en los Cuadros [A.2](#) y [A.3](#) del Apéndice A.

3.5. INICIALIZACIÓN Y ESTIMACIÓN DE LOS PARÁMETROS DE UN MODELO ETS43

De forma general, podemos presentar la formulación siguiente:

$$\begin{aligned}x_t &= \mu_t + r(v_{t-1})\varepsilon_t, \\v_t &= f(v_{t-1}) + g(v_{t-1})\varepsilon_t,\end{aligned}$$

donde $v_t = (l_t, b_t, s_t, s_{t-1}, \dots, s_{t-m+1})'$, r , f y g son funciones sobre los elementos de v_t , y $\{\varepsilon_t\}$ es ruido blanco gaussiano.

Cabe comentar que los modelos con errores multiplicativos son muy efectivos cuando la serie está compuesta por datos estrictamente positivos, pero no son numéricamente estables cuando aparece algún dato nulo o negativo. Es por ello que cuando nuestra serie contiene algún dato así solo podremos aplicar los modelos con errores aditivos.

Las predicciones puntuales que nos proporcionan estos modelos se obtienen fácilmente iterando sus ecuaciones explícitas para $t = T + 1, \dots, T + h$ y fijando $\varepsilon_{T+i} = 0$ para todo $i = 1, \dots, h$. Como comentábamos antes, en la mayoría de casos estas predicciones coincidirán con la media condicional $\mu_t(h)$, a excepción de aquellos con tendencia o estacionalidad multiplicativa.

Los modelos también nos permiten obtener intervalos de predicción. Para los modelos lineales, la distribución de las predicciones es claramente gaussiana, no siendo así para aquellos no lineales. Sin embargo, como se explica en diversos textos (ver [Taylor \(2003\)](#) y [Hyndman et al. \(2005\)](#)), para los no lineales también acaba siendo una buena aproximación. Así, podemos calcular la varianza condicional $var_t(h) = \mathbb{V}ar(x_{t+h}|v_t)$ y obtener intervalos de predicción acorde a ella.

Un procedimiento más directo y que sí funciona con todos los modelos consiste en simular multitud de desarrollos futuros de nuestra serie condicionados a los últimos elementos no observables estimados, v_T . Así, los intervalos de predicción se obtendrán simplemente como los percentiles de esa simulación. De igual modo, la predicción puntual puede ser calculada como la media de todos los valores simulados para cada instante de tiempo futuro.

3.5. Inicialización y estimación de los parámetros de un modelo ETS

Una vez estudiados los distintos modelos y tras seleccionar el adecuado para nuestro caso, para poder aplicarlos debemos estimar los parámetros que intervienen en él (α, β, γ y ϕ) y también los valores iniciales necesarios para la construcción de los métodos (v_0). Empezaremos por la inicialización y luego abordaremos la estimación.

Inicialización

La selección de valores iniciales (v_0), como ya hemos comentado anteriormente, aunque es indispensable por la estructura del modelo, normalmente no suele ser crítica. Existen múltiples enfoques, generalmente basados en una selección “ad hoc”, pero también hay heurísticas elaboradas con esta finalidad. Una de ellas es la que se recoge en Hyndman et al. (2002), que es una de las más usadas por sus buenos resultados:

Inicialización de s_0 : Supongamos que tenemos una serie estacional con frecuencia m . Lo primero que haremos será calcular un $2m$ -MA¹ para los primeros datos de la serie. Denotemos estos valores como f_t , con $t = m/2 + 1, m/2 + 2, \dots$. Si la estacionalidad es aditiva, ajustaremos la serie con la transformación $x_t - f_t$, mientras que si es multiplicativa, la transformaremos mediante x_t/f_t . Una vez hecho esto, calcularemos los valores iniciales s_{-m+1}, \dots, s_0 promediando los datos transformados para cada “estación”.

Inicialización de l_0 : Para inicializar el nivel, realizaremos una regresión lineal de los primeros diez valores frente a la variable temporal $t = 1, \dots, 10$. Para las series estacionales, se realizará tomando la serie ajustada estacionalmente como explicábamos arriba y, para series no estacionales, los datos originales. El valor l_0 será el intercepto de dicha regresión.

Inicialización de b_0 : Aprovechando la regresión anterior, si la tendencia es aditiva, fijaremos b_0 como la pendiente de dicha regresión; mientras que si la tendencia es multiplicativa, fijaremos $b_0 = 1 + \beta_0/\beta_1$, siendo β_0 el intercepto y β_1 la pendiente de la regresión.

Esta heurística nos permite calcular fácilmente una inicialización de las componentes no observables de un modo muy intuitivo. De todos modos, estos valores iniciales serán luego refinados junto con la estimación de los demás parámetros del modelo, como veremos a continuación.

Estimación

La estimación de los parámetros de nuestros modelos se hará mediante el criterio de máxima verosimilitud. Utilizando la formulación general de las ecuaciones vistas en la sección 3.4 se puede llegar a la expresión de la función de verosimilitud siguiente:

$$\mathcal{L}^*(\theta, v_0) = n \log \left(\sum_{t=1}^T \varepsilon_t^2 \right) + 2 \sum_{t=1}^T \log |r(v_{t-1})|.$$

¹Definimos la **media móvil simple** de orden m (m -MA) como la expresión

$$\frac{1}{m} \sum_{j=-k}^k x_{t+j},$$

donde $m = 2k + 1$.

Así, los parámetros del modelo (recogidos en el vector θ) y los valores iniciales v_0 pueden ser estimados minimizando la función \mathcal{L}^* .

Cabe destacar que otro método de estimación podría ser minimizando el error cuadrático medio (MSE) o minimizando la varianza de los errores σ^2 , aunque no son los procedimientos más habituales.

3.6. Precisión de pronóstico

Tras estimar los parámetros de nuestros modelos y seleccionar aquel que nos resulte más adecuado bajo distintos criterios (como el AIC o el BIC), es muy interesante poder cuantificar la precisión de las predicciones. Este concepto ha sido muy estudiado y a continuación presentaremos un breve resumen, pero puede ampliarse en [Diebold and Mariano \(2002\)](#).

Errores absolutos

Definimos el error absoluto de la predicción en un paso a $e_{t+1} = x_{t+1} - \hat{x}_t(1)$. De forma análoga, definimos el error de predicción con horizonte h a $e_{t+h} = x_{t+h} - \hat{x}_t(h)$. Cabe destacar que estos errores están en la misma escala que los datos, por lo que la medida de ajuste que hagamos con ellos también serán dependientes de ella. Las 2 medidas más conocidas de este tipo son:

- Error absoluto medio (MAE) = $media(|e_t|)$, y
- Error cuadrático medio (MSE) = $media(e_t^2)$.

Errores relativos

Ahora, el error relativo vendrá dado por $p_t = e_t/x_t$. Estos errores tienen la ventaja de ser independientes a la escala de la serie y puede ser, por lo tanto, usado para comparar distintas series. El método más utilizado es

- Error relativo absoluto medio (MAPE) = $media(|p_t|)$.

La desventaja de este método es que no se define para datos nulos de la serie y que, además, tiene una distribución muy asimétrica para datos próximos a cero.

Errores escalados

Este método se llama *Error escalado absoluto medio* (MASE) y fue presentado en [Hyndman and Koehler \(2006\)](#). Se define como

$$\text{MASE} = media(|q_t|),$$

donde

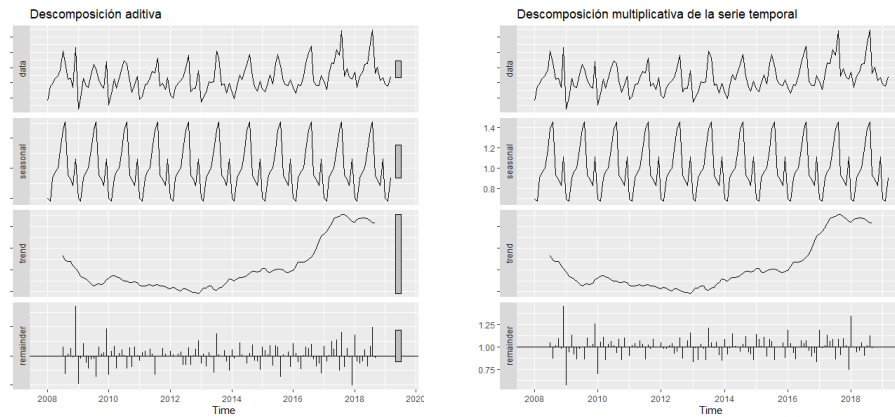
$$q_t = \frac{e_t}{\frac{1}{n-1} \sum_{i=2}^n |x_i - x_{i-1}|}.$$

Es independiente de la escala. Puede ser utilizado para comparar métodos de predicción sobre la misma serie o para comparar la precisión del pronóstico sobre series distintas.

3.7. Aplicación del Suavizado Exponencial a los datos de Estrella Galicia

En esta sección haremos una aplicación ilustrativa de los modelos ETS a los datos de Estrella Galicia. Para hacerlo más cómodo para el lector, usaremos la misma serie que en el caso de los modelos ARIMA: la venta mensual de cerveza en la Zona B.

Si recordamos, hemos empezado este capítulo hablando de la descomposición de una serie. Presentamos nuestra serie de estudio en la Figura 3.2 junto con su descomposición en sus componentes principales E , T y S , en su versión aditiva y multiplicativa. Como podemos observar, la serie presenta una tendencia no lineal y una clara componente estacional, por lo que tendremos que escoger un modelo con estacionalidad aditiva o multiplicativa.



(a) Descomposición aditiva: $x = T + S + E$. (b) Descomposición multiplicativa: $x = T \times S \times E$.

Figura 3.2: Descomposición de la serie de ventas mensuales de cerveza en la Zona B.

El siguiente paso es escoger un modelo adecuado. Por lo que hemos visto en la descomposición, podemos ahorrarnos los modelos sin estacionalidad. Tras ajustarlos uno a uno en R y recoger su AIC, los resultados que obtenemos son los mostrados en el

Método	AIC	Método	AIC
A,N,A	3576.221	A,A_d,M	3558.035
M,N,A	3570.915	M,A_d,M	3549.692
A,N,M	3553.889	A,M,A	3579.248
M,N,M	3545.335	M,M,A	3570.878
A,A,A	3579.001	A,M,M	3556.188
M,A,a	3563.242	M,M,M	3548.442
A,M,M	3556.188	A,M_d,A	3579.804
M,A,M	3549.373	M,M_d,A	3571.038
A,A_d,A	3579.828	A,M_d,M	3557.841
M,A_d,A	3571.461	M,M_d,M	3549.298

Cuadro 3.2: Criterio AIC de los distintos modelos ajustados sobre nuestra serie de estudio.

Cuadro 3.2. Como podemos ver en el mismo, el menor AIC se obtiene para el modelo ETS(M,N,M), es decir, para el modelo sin tendencia, estacionalidad multiplicativa y errores también multiplicativos:

$$x = T \times S \times E, \text{ con } T \text{ de tipo "nula" .}$$

Además, viendo la salida propia de ese ajuste, podemos ver que las estimaciones de los parámetros son: $\alpha = 0,1244$ y $\gamma = 0,4946$. En cuanto a los valores iniciales del vector de elementos no observables, v_0 , se tiene que $l_0 = 328554$ y que $(s_{-m+1}, \dots, s_0) = (1,44, 0,82, 0,92, 0,95, 1,26, 1,4, 1,12, 0,98, 0,92, 0,92, 0,75, 0,53)$. Finalmente, se tiene que la varianza de los errores es $\sigma^2 = 0,1279$ y los criterios de información son $AIC=3545.335$, $AICc=3549.368$ y $BIC=3588.914$. Una representación gráfica de los valores que va tomando el vector v_t a lo largo del tiempo se puede ver en la Figura 3.3.

Por lo tanto, las ecuaciones explícitas de nuestro modelo que nos permitirán obtener tanto las predicciones puntuales como los intervalos de predicción son las siguientes:

$$l_t = l_{t-1}(1 + 0,1244\varepsilon_t),$$

$$s_t = s_{t-12}(1 + 0,4946\varepsilon_t),$$

$$\mu_t = l_{t-1}s_{t-12}.$$

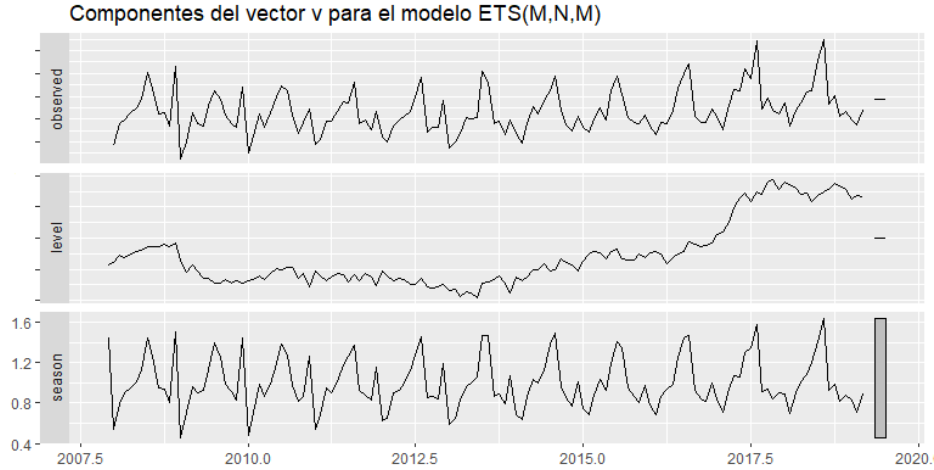


Figura 3.3: Representación gráfica de los valores de v_t a lo largo de la serie.

Ya estamos, al fin, en condiciones de calcular las predicciones puntuales y sus intervalos de confianza. Si recordamos lo explicado en la sección 3.4, eso se puede llevar a cabo asumiendo gaussianidad en los residuos o mediante un procedimiento de simulación. El primer enfoque nos devuelve las predicciones recogidas en el Cuadro 3.3 y representadas en la Figura 3.4. En cambio, siguiendo el segundo enfoque, obtenemos las predicciones del Cuadro 3.4 y la Figura 3.5.

Con estos resultados hemos acabado la predicción de nuestra serie temporal. Sin embargo, sospechamos que podríamos afinarlas un poco más. En la representación gráfica de la descomposición de nuestra serie de la Figura anterior 3.2, vemos que esta tiene una tendencia cambiante a lo largo del tiempo de estudio. En 2008 empieza un decrecimiento en la tendencia general de las ventas, muy probablemente debido a la crisis económica en la que se envuelta la zona de estudio. No es hasta 2013 cuando se empieza a apreciar una recuperación al alza en las ventas, empezando en este año una tendencia creciente continua hasta el último año. Por todo esto, sospechamos que si consideramos la serie temporal iniciada en 2013, la componente tendencia T debería ser mucho más clara para nuestros ajustes pudiendo finalmente mejorar las predicciones futuras.

Si consideramos por tanto esta “nueva” serie, obtenemos la descomposición mostrada en la Figura 3.6.

Como vemos, podemos incluso considerar que en este caso la tendencia es li-

	Predic. puntual	Inf 80	Sup 80	Inf 95	Sup 95
Abr 2019	337109.61	281870.63	392348.60	252628.88	421590.35
May 2019	360665.79	301103.85	420227.73	269573.66	451757.91
Jun 2019	402612.52	335610.30	469614.74	300141.46	505083.58
Jul 2019	471338.92	392303.16	550374.67	350464.17	592213.67
Ago 2019	546848.23	454464.14	639232.33	405558.95	688137.51
Sep 2019	308879.27	256312.36	361446.18	228485.12	389273.42
Oct 2019	328492.29	272180.85	384803.73	242371.37	414613.21
Nov 2019	273904.84	226614.25	321195.43	201580.12	346229.56
Dic 2019	294457.38	243258.78	345655.98	216155.87	372758.89
Ene 2020	276873.16	228396.20	325350.12	202734.05	351012.27
Feb 2020	237627.56	195735.71	279519.41	173559.51	301695.61
Mar 2020	299227.38	246117.83	352336.94	218003.32	380451.44

Cuadro 3.3: Predicciones del modelo ETS(M,N,M) siguiendo un enfoque paramétrico (gausiano). La segunda columna recoge las predicciones puntuales, mientras que la tercera y cuarta su intervalo de confianza a un nivel del 80 %, y la quinta y sexta el intervalo de confianza con un nivel del 95 %. Estos valores han sido transformados por un parámetro α para respetar la confidencialidad de los datos reales.

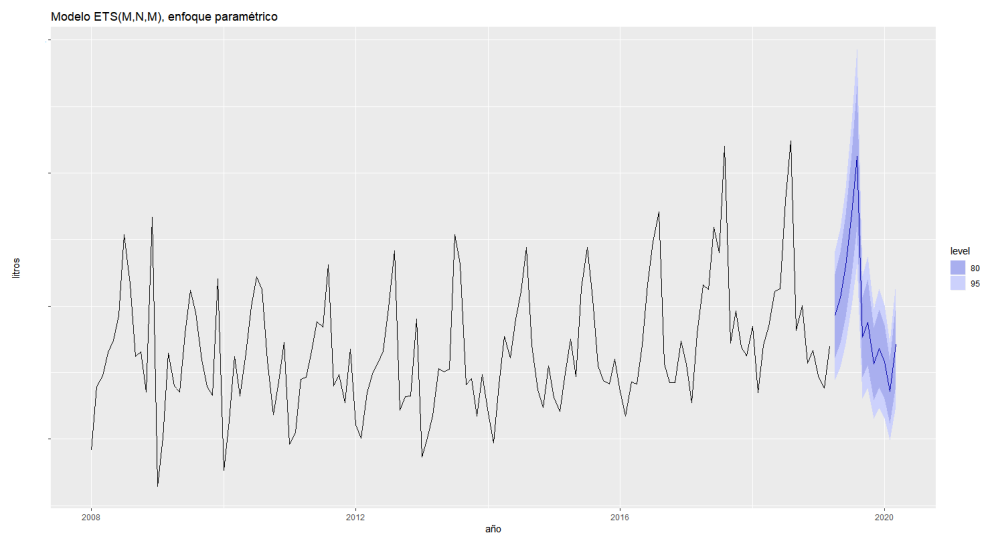


Figura 3.4: Representación gráfica de las predicciones bajo un enfoque paramétrico.

	Predic. puntual	Inf 80	Sup 80	Inf 95	Sup 95
Abr 2019	337109.61	282260.25	385007.97	263826.39	435440.63
May 2019	360665.79	304345.51	414105.26	281103.15	462101.56
Jun 2019	402612.52	338173.81	461734.35	313274.10	514672.69
Jul 2019	471338.92	395680.95	539594.90	361390.13	604195.17
Ago 2019	546848.23	460073.30	629328.85	420245.46	698077.57
Sep 2019	308879.27	260089.19	357369.33	237094.23	395181.03
Oct 2019	328492.29	275727.03	379916.31	250762.87	420518.23
Nov 2019	273904.84	231309.51	316383.14	210506.47	346355.18
Dic 2019	294457.38	247587.40	342890.40	225106.79	375273.34
Ene 2020	276873.16	231648.26	320034.52	211671.93	352564.10
Feb 2020	237627.56	198727.29	276053.94	180690.45	301794.01
Mar 2020	299227.38	250215.99	349170.57	228507.92	385460.38

Cuadro 3.4: Predicciones del modelo ETS(M,N,M) siguiendo el enfoque de simulación y bootstrap. La segunda columna recoge las predicciones puntuales, mientras que la tercera y cuarta su intervalo de confianza a un nivel del 80%, y la quinta y sexta el intervalo de confianza con nivel 95%. Estos valores han sido transformados por un parámetro α para respetar la confidencialidad de los datos reales.

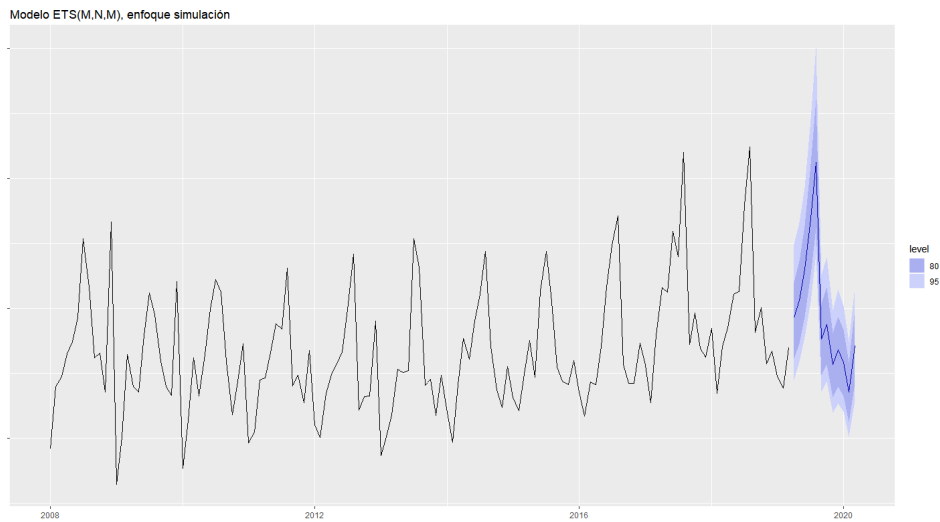
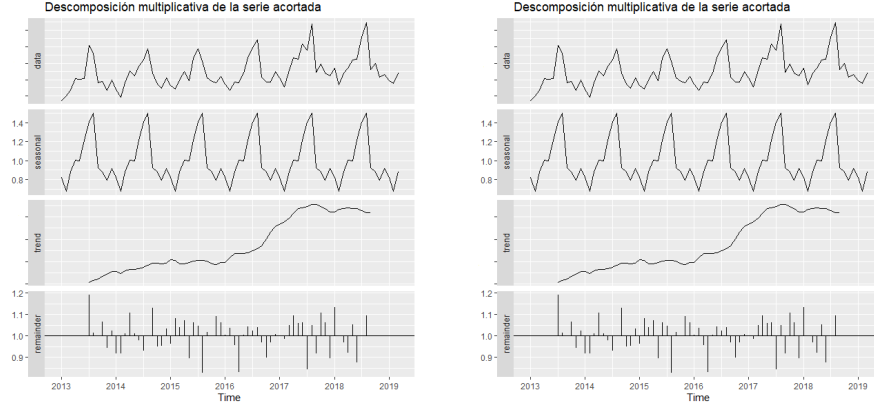


Figura 3.5: Representación gráfica de las predicciones bajo un enfoque de simulación y bootstrap.



(a) Descomposición aditiva: $x = T + S + E$. (b) Descomposición multiplicativa: $x = T \times S \times E$.

Figura 3.6: Descomposición de la serie de ventas mensuales de cerveza en la Zona B a partir del año 2013.

neal. Aplicando una selección automática en R, obtenemos que el modelo más adecuado para esta serie es el modelo ETS(M,M,M), es decir, $x = T \times S \times E$, con $T_h = lb^h$. Los parámetros ajustados son $\alpha = 0,2109$ y $\beta = \gamma = 10^{-4}$. En cuanto a los valores iniciales de v_0 se tiene que $l_0 = 288373,3674$, $b_0 = 1,0053$ y que $(s_{-m+1}, \dots, s_0) = (0,90, 0,79, 0,90, 0,93, 1,52, 1,41, 1,19, 1,02, 1,02, 0,87, 0,67, 0,78)$. Finalmente, se tiene que la varianza de los errores es $\sigma^2 = 0,1016$ (cuando antes era $\sigma^2 = 0,1279$) y sus criterios de información son $AIC=1906.299$, $AICc=1917.036$ y $BIC=1945.696$. Una representación gráfica de los valores que va tomando el vector v_t a lo largo del tiempo se puede ver en la Figura 3.7.

Por lo tanto, las ecuaciones explícitas de nuestro modelo que nos permitirán obtener tanto las predicciones puntuales como los intervalos de predicción son las siguientes:

$$l_t = l_{t-1}b_{t-1}(1 + 0,2109\varepsilon_t),$$

$$b_t = b_{t-1}(1 + 10^{-4}\varepsilon_t),$$

$$s_t = s_{t-12}(1 + 10^{-4}\varepsilon_t),$$

$$\mu_t = l_{t-1}b_{t-1}s_{t-12}.$$

Para este nuevo caso, las predicciones (con el enfoque simulación) son las mos-

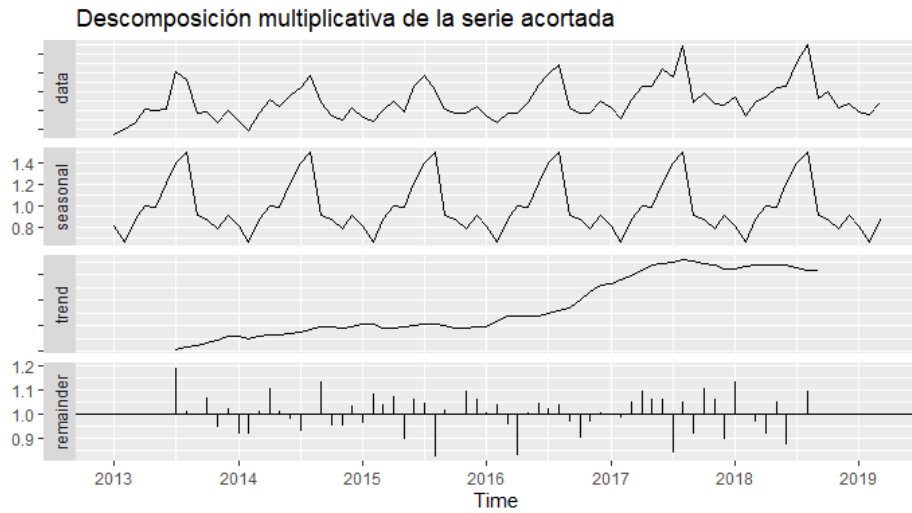


Figura 3.7: Representación gráfica de los valores de v_t a lo largo de la serie acortada.

tradas en el Cuadro 3.5 y representadas en la Figura 3.8. Como podemos ver, las predicciones parecen a simple vista más realistas (los picos parecen seguir mejor el patrón al alza de los últimos años) y, sobretodo, los intervalos resultan ser más estrechos, lo que es de mucha más utilidad para la empresa.

	Predic. puntual	Inf 80	Sup 80	Inf 95	Sup 95
Abr 2019	361456.25	317115.06	397791.71	284688.43	420234.98
May 2019	362724.10	318476.90	400224.42	289619.03	421675.31
Jun 2019	429334.19	375041.50	475203.88	341408.52	500241.46
Jul 2019	509063.05	444490.34	564019.07	403711.60	596130.41
Ago 2019	549901.16	479491.06	614031.66	436500.70	651303.20
Sep 2019	337562.10	293490.83	378356.41	267792.69	402474.13
Oct 2019	330127.84	286870.39	369892.61	260585.28	394927.22
Nov 2019	294253.47	253902.11	331527.37	231783.12	351761.71
Dic 2019	335322.46	288754.99	377230.07	264154.71	403551.74
Ene 2020	290128.39	250308.09	328026.03	227864.02	349425.70
Feb 2020	251881.22	217013.30	285182.68	197056.67	304944.06
Mar 2020	326818.91	279784.22	371741.57	251041.75	399434.08

Cuadro 3.5: Predicciones del modelo ETS(M,M,M) siguiendo el enfoque de simulación y bootstrap. Se recoge las predicciones puntuales junto con su intervalo de confianza a un nivel del 80 % y 95 %. Estos valores han sido transformados por un parámetro α para respetar la confidencialidad de los datos reales.

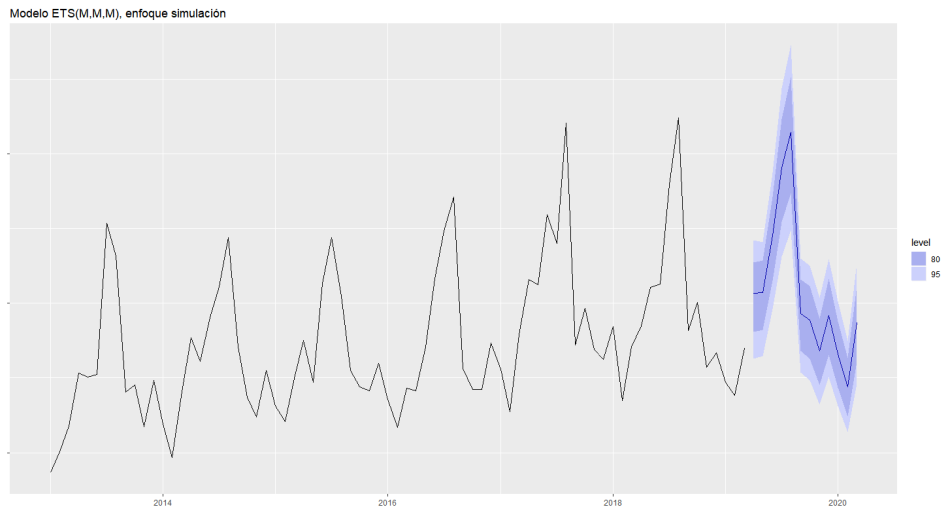


Figura 3.8: Representación gráfica de las predicciones bajo un enfoque de simulación y bootstrap.

Capítulo 4

Modelos avanzados

4.1. Motivación

Si bien es cierto que los modelos que hemos estudiado en los capítulos anteriores son la base de trabajo con series temporales, hemos observado que en el día a día de nuestra cervecera pueden no ser del todo útiles. Esto se debe a dos grandes motivos.

El primero es que, por motivos de logística, la empresa está especialmente interesada en la predicción a corto plazo: le sería más útil poder predecir las ventas a nivel semanal y no tanto a nivel mensual. El problema lo encontramos en que, por la naturaleza de los datos, las series de ventas semanales presentan una alta volatilidad y, aunque las predicciones puntuales pueden ser aceptables, los intervalos de confianza que se obtienen son demasiado amplios como para ayudar en la toma de decisiones de la empresa. Un ejemplo de serie semanal junto con los intervalos de predicción obtenidos para ella se ilustra en la Figura 4.1.

Para solucionar esta problemática, hemos pensado en la incorporación de covariables externas como la temperatura o la presencia de días festivos, para ayudar a explicar al menos parte de esta volatilidad. Así, sabiendo en qué condiciones se presenta la semana futura, podremos ser un poco más precisos en su predicción. Para ello, estudiaremos en la siguiente sección la **regresión dinámica**.

El segundo motivo por lo que debemos aportar algún modelo más avanzado que los estudiados en los primeros capítulos es que, como podemos imaginar, Hijos de Rivera tiene una amplia gama de productos a la venta y, cada uno de esos productos, en formatos y dimensiones distintas. Por ejemplo, nuestra empresa vende cerveza Estrella Galicia en botellas de vidrio de 33 cl, pero también en barriles de 5 litros; o por ejemplo, agua Cabreiroá en botellas de plástico de 1.5 litros y en botellas de vidrio de 50 cl. Esto hace que no sea del todo útil predecir simplemente los litros totales de un producto que se venderán en un tiempo t , ya que la fábrica podría embotellarlos en un

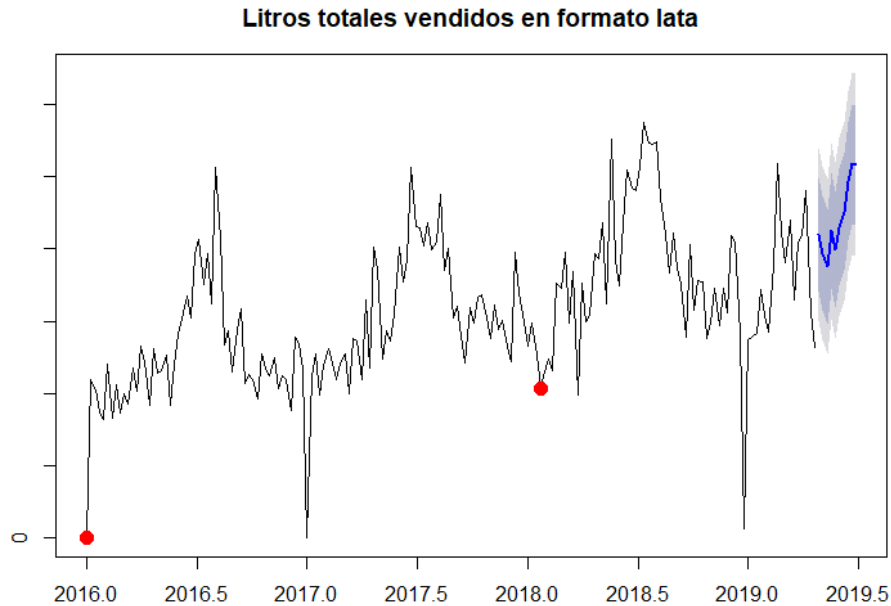


Figura 4.1: Serie con datos semanales de los litros totales de cerveza vendidos en formato lata y sus predicciones para las próximas 10 semanas. Se puede observar que las predicciones puntuales son “creíbles” y aceptables, pero los intervalos de confianza más anchos de lo que nos gustaría. En rojo resaltamos las ventas de la primera semana de 2016 y la primera de 2018.

formato que igual no es el más demandado finalmente (produciendo así sobrantes). Lo ideal para la empresa sería poder desagregar los datos de ventas en función de materiales y formatos y predecir conjuntamente dichas series.

Toda este abanico de productos hace que nuestros datos acaben presentándose en estructuras de series temporales jerarquizadas o agregadas. En un primer momento podemos pensar en desagregar todo lo posible los datos, considerar esas series temporales, hacer predicciones sobre ellas y finalmente agregar las predicciones. Sin embargo, como cabría esperar, este enfoque acaba produciendo unas previsiones en las series más agregadas muy elevadas, claramente irreales. Es por ello que necesitamos un modelo que sepa trabajar con estas estructuras y dar predicciones en su conjunto. Para ello, estudiaremos en la última sección de este capítulo las **series jerarquizadas o agregadas**.

4.2. Regresión Dinámica

Como comentábamos hace un momento, los modelos que hemos estudiado anteriormente nos permiten explicar los valores futuros de la serie a partir exclusivamente de los observados en el pasado, pero no dan lugar a integrar información externa que pueda ser relevante. La finalidad de esta sección será extender los modelos ARIMA para permitir que tengan en cuenta otras covariables.

La idea de la regresión dinámica es muy sencilla: simplemente consiste en permitirles a los errores de una regresión que estén autocorrelados y ajustar dicha autocorrelación mediante un modelo ARIMA. Así, si consideráramos en una regresión clásica que la variable y venía determinada por las variables explicativas x_1, \dots, x_k mediante la expresión

$$y = \beta_0 + \beta_1 x_1 + \dots + \beta_k x_k + \varepsilon,$$

ahora consideraremos que

$$y_t = \beta_0 + \beta_1 x_{1,t} + \dots + \beta_k x_{k,t} + \eta_t,$$

$$\eta_t \sim ARIMA(p, d, q) \times (P, D, Q)_s.$$

Por ejemplo, si tras hacer la regresión pertinente estudiamos sus errores y observamos que pueden ser explicados mediante un $ARMA(1, 1)$, el modelo que resultaría sería

$$y_t = \beta_0 + \beta_1 x_{1,t} + \dots + \beta_k x_{k,t} + \eta_t,$$

$$(1 - \phi_1 B)\eta_t = c + (1 + \theta_1 B)\varepsilon_t,$$

con $\beta_0, \beta_1, \dots, \beta_k, \phi_1$ y θ_1 los parámetros a estimar.

Notación 4.1. Cabe resaltar que ahora denotaremos por η_t los errores de la regresión para resaltar la idea de que dichos errores se consideran autocorrelados, mientras que los errores del ARIMA (que sí denotaremos con ε_t) son los únicos que asumimos como ruido blanco.

Además, como se da el hecho de que la variable explicativa también viene expresada como una serie de tiempo, puede darse el caso de que el efecto de esta se manifieste en la variable dependiente con cierto retardo, r . Supongamos que solo tenemos una variable explicativa, en ese caso, también podríamos considerar la regresión

$$y_t = \beta_0 + \gamma_0 x_t + \gamma_1 x_{t-1} + \dots + \gamma_r x_{t-r} + \eta_t,$$

$$\eta_t \sim ARIMA(p, d, q) \times (P, D, Q)_s.$$

Se nos presentan, por lo tanto, tres puntos clave ausentes en la regresión clásica: la posibilidad de que los errores de la regresión η_t estén correlados, el desconocimiento del retardo r para las variables explicativas y, por último, la posibilidad de que los procesos (tanto la serie explicativas como la serie respuesta) no sean estacionarios. En las siguientes subsecciones estudiaremos estos puntos.

4.2.1. Relación lineal entre series de tiempo

Para estudiar el grado de existencia, o inexistencia, entre dos series de tiempo, así como el valor del posible retardo entre ellas, empezaremos considerando unas definiciones previas.

Definición 4.2. *Llamamos **función de covarianzas cruzadas** entre las series X e Y a la aplicación*

$$\gamma_{s,t}(X, Y) = Cov(X_s, Y_t).$$

Definición 4.3. *Llamamos **función de correlaciones cruzadas** entre las series X e Y a la aplicación*

$$\rho_{s,t}(X, Y) = \frac{\gamma_{s,t}(X, Y)}{\sigma_{X_s} \sigma_{Y_t}}.$$

Cuando las correlaciones cruzadas solo dependen del retardo entre las variables y no del instante inicial considerado, es decir,

$$\rho_{t,t-k}(X, Y) = \rho_{s,s-k}(X, Y), \forall t, s, k$$

denotaremos esta función como $\rho_k(X, Y) = \rho_{t,t-k}(X, Y)$.

Ambas funciones miden el grado de dependencia lineal entre X_s e Y_t . Cabe destacar que la función de correlaciones cruzadas carece de unidades y toma valores entre -1 y 1.

Supongamos ahora dos procesos $\{X_t\}$ y $\{e_t\}$, ambos ruido blanco e independientes entre si, y consideremos el proceso $\{Y_t\} = \beta_0 + \beta_1 X_{t-r} + e_t$. En este caso, se tendría que el único valor no nulo de $\rho_k(X, Y)$ se alcanzaría en $k = -r$. Por lo tanto, podemos concluir que los valores $\rho_k(X, Y)$ nos asesoran acerca de valor del retardo r .

Como ocurría con las funciones *fas* y *lap*, la función de correlaciones cruzadas también es desconocida y debemos trabajar con su versión muestral:

$$\hat{\rho}_k(X, Y) = \frac{\sum_{t=1}^T (x_t - \bar{x})(y_{t-k} - \bar{y})}{\sqrt{\sum_{t=1}^T (x_t - \bar{x})^2} \sqrt{\sum_{t=1}^T (y_t - \bar{y})^2}}.$$

Si $\{X_t\}$ e $\{Y_t\}$ son ruido blanco (salvo constantes) e independientes entre si, se tiene que

$$\hat{\rho}_k(X, Y) \approx \mathcal{N}\left(0, \frac{1}{\sqrt{T}}\right).$$

Por lo tanto, diremos que son significativamente distintas de cero si se salen de las bandas

$$\left(\frac{-1,96}{\sqrt{T}}, \frac{1,96}{\sqrt{T}} \right).$$

Cabe destacar que si las series consideradas no son ruido blanco (o al menos una de ellas y la otra estacionaria), se pueden deducir relaciones lineales de forma errónea. Esto se debe a que, en tal caso, la variabilidad de los estimadores de las correlaciones cruzadas puede ser muy superior a $\frac{1}{\sqrt{T}}$. Es por ello que debemos transformar el proceso no estacionario en ruido blanco. A dicha transformación la llamaremos **preblanqueado**.

En resumen, el proceso habitual en el estudio de la relación lineal entre series temporales es el siguiente:

1. Transformar el proceso $\{X_t\}$ en un proceso de ruido blanco $\{\tilde{X}_t\} = \pi(B)X_t$.
2. Aplicar la misma transformación al proceso $\{Y_t\}$ para obtener $\{\tilde{Y}_t\} = \pi(B)Y_t$.
3. Proponer un valor de retardo r , si para algún k se verifica que $|\hat{\rho}_k(\tilde{X}, \tilde{Y})| \geq 1,96/\sqrt{T}$. En ese caso, propondríamos la regresión

$$Y_t = \beta_0 + \beta_1 X_{t-r} + \eta_t.$$

4.2.2. Estimación de parámetros

El enfoque habitual que seguíamos en una regresión para estimar sus coeficientes era el de mínimos cuadrados, que consistía en minimizar la suma de los errores al cuadrado. Ahora, si minimizamos los errores η_t ignorando sus autocorrelaciones, nos surgirán varios problemas. El primero es que las estimaciones $\hat{\beta}_0, \dots, \hat{\beta}_k$ dejarán de ser verdaderamente las mejores estimaciones por haber ignorado cierta información a la hora de su cálculo y, además, los p -valores asociados a estos coeficientes casi siempre serán muy pequeños, lo que les dará una importancia que probablemente no tengan realmente (es lo conocido como *regresión espuria*). Otro problema que encontraremos es que cualquier contraste estadístico que realicemos sobre el modelo será incorrecto y, además, los valores AIC o BIC dejarán de ser una guía fiable. Para evitar todo esto, lo que se deberá hacer es minimizar la suma de cuadrados de los errores ε_t del *ARIMA*, o bien utilizar el enfoque de máxima verosimilitud que, en la mayoría de los casos, devolverá estimaciones parecidas.

4.2.3. Aplicación de la Regresión Dinámica a Estrella Galicia

Como comentábamos al principio del capítulo, el objetivo de este modelo era intentar ayudar en las predicciones de ventas semanales de nuestra compañía. Tras barajar varias posibilidades, hemos concluido que las variables exógenas que más

sentido tiene que estudiemos son las temperaturas y la existencia de festivos en cada semana.

Regresión Dinámica con los días festivos

Consideremos ahora el número de días laborables que tiene la semana como variable factorial de nuestra regresión. Este planteamiento surgió al comprobar que las semanas con grandes festividades como Semana Santa o el puente de la Constitución suponían picos llamativos en las series de ventas. Sin embargo, debido a la propia naturaleza de algunas de estas festividades que hace variar su localización en el calendario de año en año, nos resulta muy complicado recoger dichos sucesos mediante la estacionalidad de la serie. Además, el hecho de que la propia enumeración de las semanas varíe levemente, hace que en algunos años las ventas de la primera o última semana sean prácticamente nulas y en otros no. Por ejemplo, el día 1 de enero de 2016 fue viernes y festivo, por lo que la primera semana de 2016 no tuvo ningún día laborable y, por lo tanto, como se puede comprobar en la Figura 4.1 que presentábamos al principio de este capítulo, las ventas de esa semana fueron nulas. Por el contrario, el 1 de enero de 2018 fue lunes, por lo que la primera semana de 2018 tuvo 4 días laborables y el registro de sus ventas, aunque se sitúe en un valle, está lejos del cero (ver Figura 4.1).

Esta problemática es propia de las series con frecuencia semanal y diaria y, por ello, se nos ha ocurrido utilizar como variable exógena el número de días laborables que tiene cada semana. Para ilustrar este modelo, usaremos nuevamente la serie de ventas semanales de litros de cerveza embotellados en lata.

Lo primero que haremos será comprobar que realmente hay motivos para sospechar de la relación entre las ventas semanales y el número de días laborables. Para ello, recogemos el box-plot de la Figura 4.2, en el que podemos ver claramente la diferencia en el volumen de ventas dependiendo de la cantidad de días trabajados. Además, también podemos ajustar una regresión lineal de las ventas frente al factor días laborables, resultando la siguiente salida:

Call:

```
lm(formula = ventas ~ ndias)
```

Residuals:

Min	1Q	Median	3Q	Max
-328150	-136386	-30079	129539	498839

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	2658	128203	0.021	0.98348

```

ndias1      23507      222055    0.106  0.91582
ndias3      466968     145369    3.212  0.00157 **
ndias4      590703     132149    4.470  1.42e-05 ***
ndias5      653324     129157    5.058  1.08e-06 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

```

```

Residual standard error: 181300 on 171 degrees of freedom
Multiple R-squared:  0.2065, Adjusted R-squared:  0.1879
F-statistic: 11.12 on 4 and 171 DF,  p-value: 4.822e-08

```

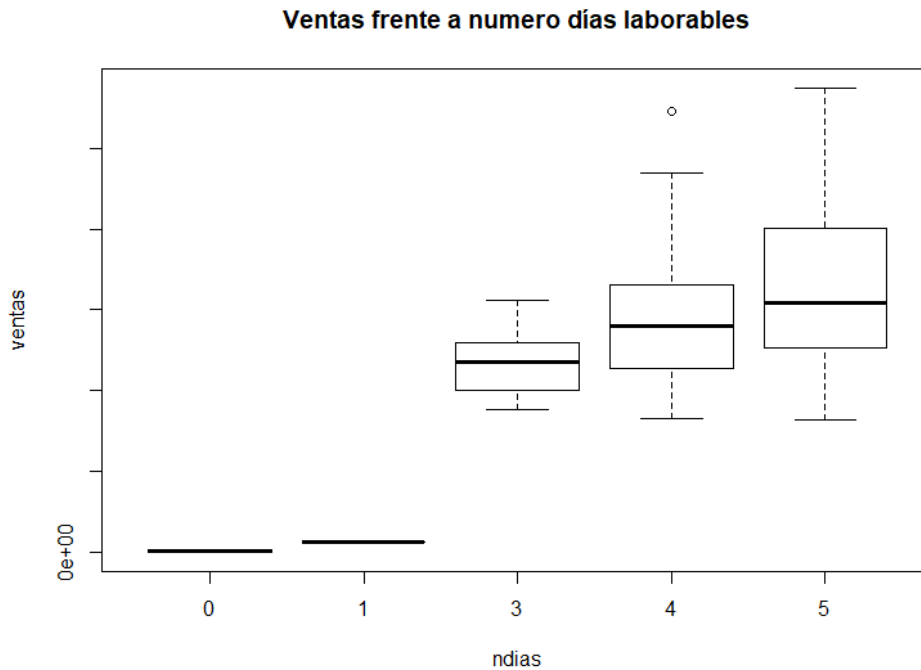


Figura 4.2: Box-Plot de las ventas frente al número de días laborables de esa semana.

Hay que tener en cuenta que las observaciones utilizadas para la anterior regresión no son independientes, por lo que el nivel de significación de los coeficientes debe tomarse con cierta precaución. De todos modos, parece evidente que el número de días laborables en esa semana sí afecta en las ventas totales.

Por ello, considerando los datos como series temporales, podemos calcular las correlaciones cruzadas teniendo en cuenta los posibles retardos. Como hemos comentado

en la sección anterior, debemos cerciorarnos de que las series, en primer lugar, son estacionarias y luego preblanquearlas. En nuestro caso, se tiene que la serie de días laborables sí es estacionaria, pero sin embargo la de ventas no (ver Figura 4.3). Es por ello que debemos diferenciarla y, para mantener la coherencia entre ambas series, debemos hacer las mismas transformaciones en ambas, resultando las funciones fas y fap que se recogen en la Figura 4.4.

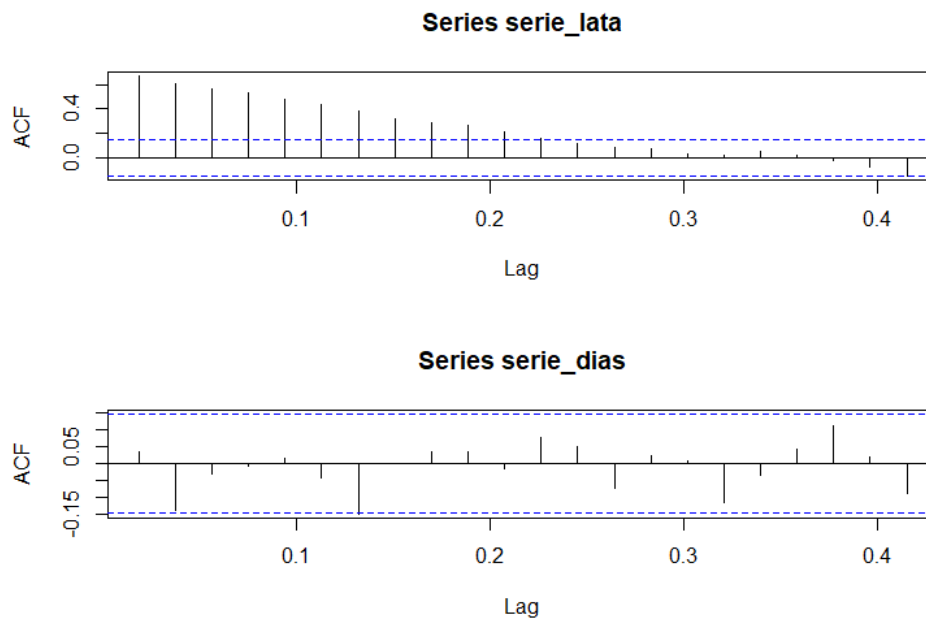


Figura 4.3: Función fas de la serie de ventas y la serie de días laborables. En la primera, tarda mucho en anularse, por lo que no la podemos considerar estacionaria.

Ahora ya sí estamos en condiciones de aplicar el preblanqueado, obteniendo la función de correlaciones cruzadas que se muestra en la Figura 4.5. Como podemos observar, al salirse una de ellas de las bandas, podemos concluir que sí existe correlación lineal entre las series y que además se produce instantáneamente, es decir, con retardo $r = 0$. Por lo tanto, el modelo de regresión que proponemos es $\nabla(y_t) = \beta_0 + \beta_1 \cdot \nabla(x_t) + \eta_t$, siendo y_t las ventas en el instante t , x_t el número de días laborables de la semana t y η_t el error procedente de un modelo *ARIMA* que debemos ajustar.

Ahora tenemos que estimar los parámetros necesarios. Para ello, calcularemos primero la regresión lineal entre la serie diferenciada de las ventas frente a la serie

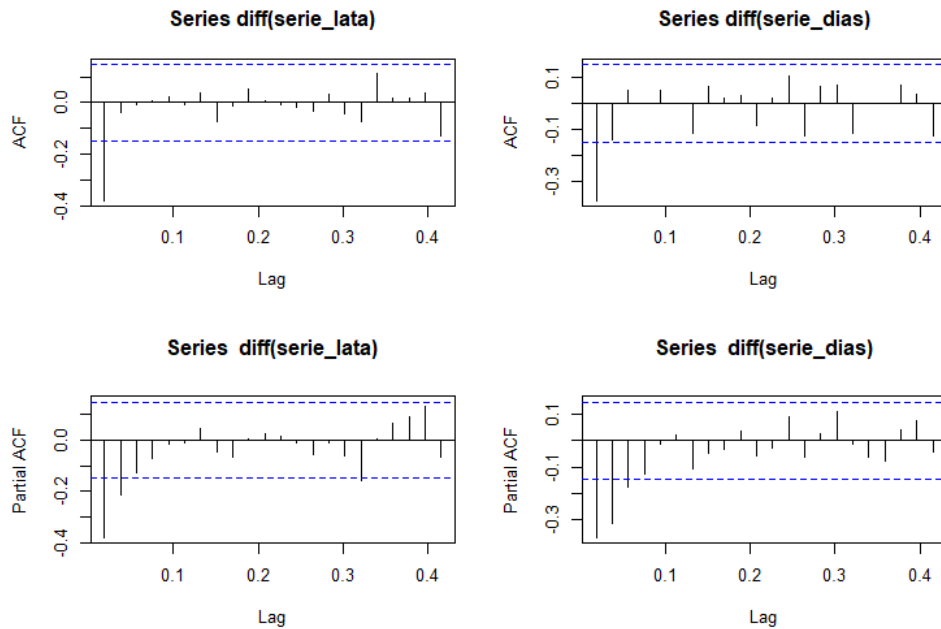


Figura 4.4: Funciones fas y fap de la serie diferenciada de ventas y la serie diferenciada de días laborables. En ellas se aprecia que ambas series son estacionarias..

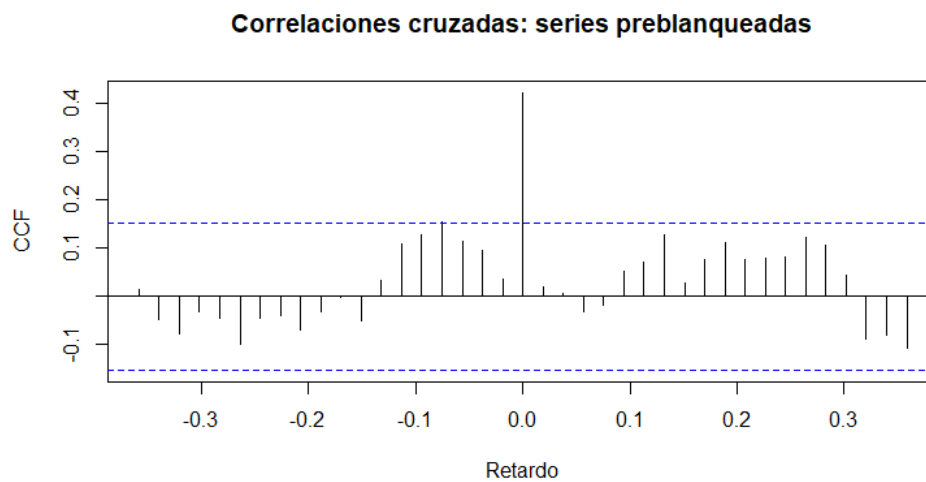


Figura 4.5: Función de correlaciones cruzadas para las series preblanqueadas. Vemos que para el retardo $r = 0$ la correlación es significativamente distinta de cero.

diferenciada de los días laborables. Luego tomaremos sus residuos y, si estos son estacionarios y tras ajustar un modelo *ARIMA* para ellos el ajuste de la regresión final supera el análisis de residuos, habremos terminado.

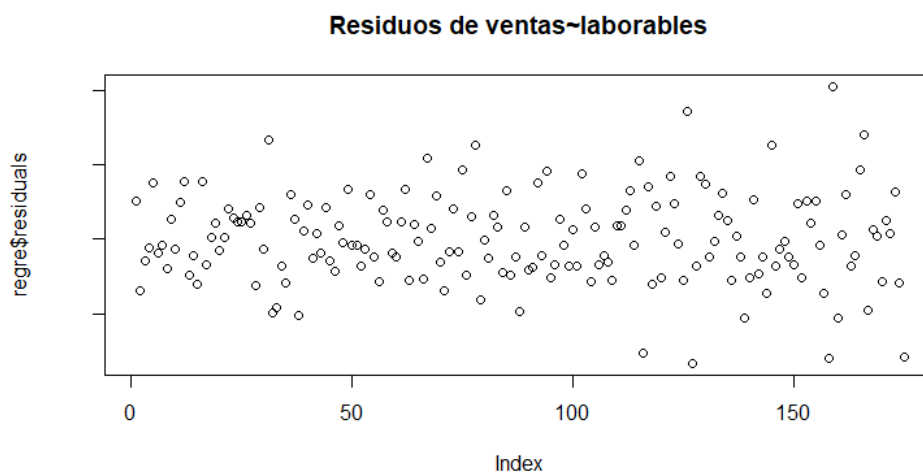


Figura 4.6: Representación gráfica de los residuos de la regresión lineal ventas frente días laborables.

Tras ajustar la regresión lineal entre ambas series, recogemos en la Figura 4.6 sus residuos, donde podemos intuir que, efectivamente, son estacionarios. De todos modos, en la Figura 4.7 representamos sus funciones *fas* y *fap* y concluimos que sí son estacionarios. Además, a la vista de esas funciones, podemos proponer un modelo *MA(1)* para modelizar su proceso estocástico.

Así, ahora que ya sabemos el retardo ($r = 0$) y un modelo generador de nuestros residuos (*MA(1)*), ya estamos en condiciones de ajustar los coeficientes de nuestra regresión en conjunto. Para ello, hemos utilizado la función `arima()` del paquete `forecast` obteniendo la salida siguiente:

Coefficients:

	ma1	intercept	xreg
	-0.4011	1262.908	102637.31
s.e.	0.0820	5166.414	10769.95

sigma^2 estimated as 1.29e+10: log likelihood=-2285.43, aic=4576.86

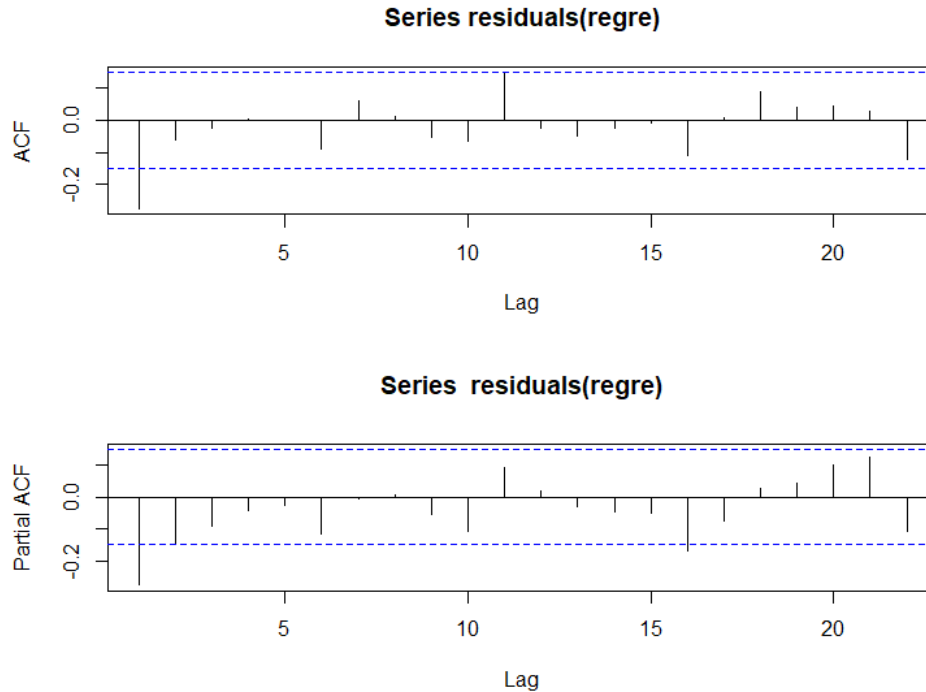


Figura 4.7: Funciones f_{as} y f_{ap} de los residuos de la regresión lineal ventas frente días laborables.

Así, el modelo final que habíamos obtenido

$$y_t - y_{t-1} = \beta_0 + \beta_1(x_t - x_{t-1}) + \eta_t,$$

$$\eta_t \sim MA(1).$$

se traduce en

$$y_t - y_{t-1} = 1262,908 + 102637,31(x_t - x_{t-1}) + (1 - 0,4011)\varepsilon_t.$$

Por último, no hay que olvidar checkear los test de residuos (ver Figura 4.8). En este caso, el t-test no rechaza que la media sea nula (con un p-valor = 0.987), el test de Jarque-Bera no rechaza asimetría (con un p-valor = 0.1575) y el test de Shapiro no rechaza normalidad (p-valor = 0.1197). Concluimos que los test fueron superados y que por lo tanto el modelo es válido.

Estamos ya en condiciones de calcular las predicciones futuras. Para ello, nosotros daremos los nuevos valores para la variable explicativa y, por otro lado, los nuevos

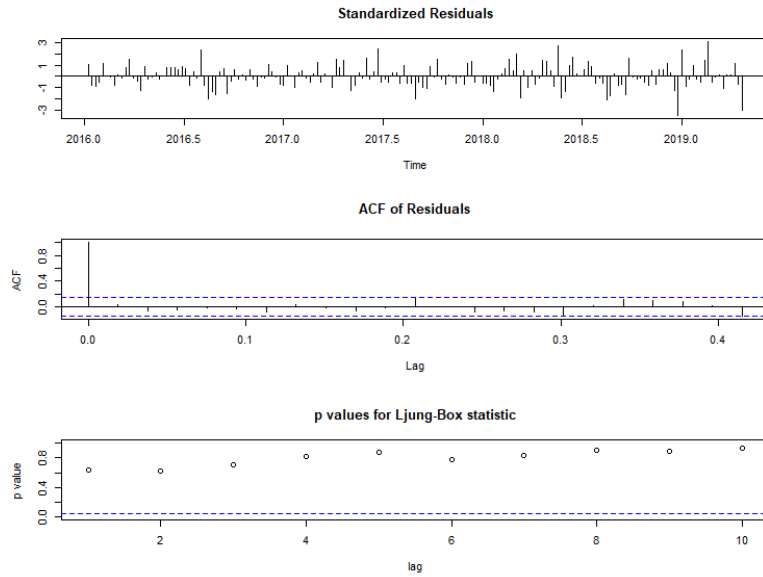


Figura 4.8: Representaciones gráficas de los tests sobre los residuos del ajuste final.

residuos se calcularán siguiendo el modelo $MA(1)$ que los define siguiendo el procedimiento que hemos estudiado en el Capítulo 2. Tras calcular las predicciones en R con un horizonte $h = 5$ y suponiendo que las próximas 5 semanas tienen 5 días laborables, hemos obtenido los valores recogidos en el Cuadro 4.1 y Figura 4.9.

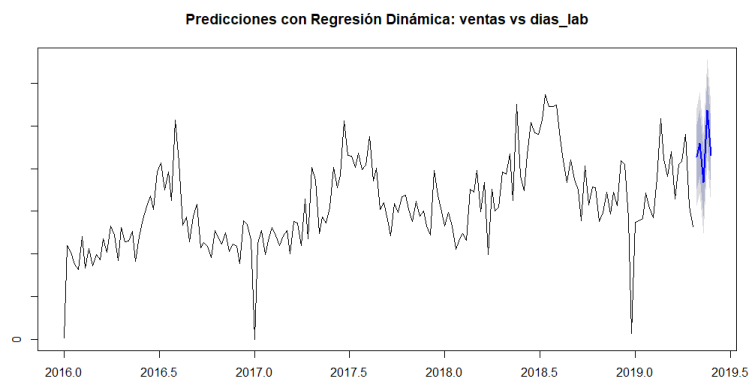


Figura 4.9: Representación gráfica de las predicciones con Regresión Dinámica de las ventas frente al número de días laborables.

	Predic. puntual	Inf 80	Sup 80	Inf 95	Sup 95
29 Abr - 5 May	748117.96	616738.26	879497.67	547190.04	949045.88
6 May - 12 May	804146.41	668129.37	940163.44	596126.30	1012166.51
13 May - 19 May	642116.12	506099.08	778133.15	434096.01	850136.22
20 May - 26 May	940915.56	804898.52	1076932.59	732895.46	1148935.66
27 May - 2 Jun	753224.51	617207.48	889241.55	545204.41	961244.61

Cuadro 4.1: Predicciones para la serie de ventas con Regresión Dinámica frente a los días laborables. Estos valores han sido transformados por un parámetro α para respetar la confidencialidad de los datos reales.

Regresión Dinámica con las Temperaturas

No hace falta explicar por qué hemos considerado la temperatura como un factor relevante en la venta de cerveza. Para estudiar su relación estadísticamente, consideraremos la venta (en litros) de cerveza Estrella Galicia Especial en la delegación de Pontevedra. Por otra parte, hemos obtenido un registro diario de la temperatura máxima en la ciudad de Pontevedra desde la Base de Datos pública de la AEMET y luego las hemos promediado para obtener la temperatura media semanal entre los años 2016 y 2018. Así, obtenemos la serie de ventas y la serie de temperaturas que mostramos en la Figura 4.10.

En los gráficos secuenciales podemos comprobar que, a simple vista, existe cierto paralelismo entre las temperaturas y las ventas. Además, si hacemos una regresión simple enfrentando ambos conjuntos de datos, obtenemos la representación gráfica de la Figura 4.11 y la siguiente salida:

```
Call:
lm(formula = ventas ~ temps)

Coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept)   -8601         8088  -1.063   0.289
temps           5231          403  12.982 <2e-16 ***
---
Residual standard error: 28200 on 154 degrees of freedom
Multiple R-squared:  0.5225, Adjusted R-squared:  0.5194
F-statistic: 168.5 on 1 and 154 DF, p-value: < 2.2e-16
```

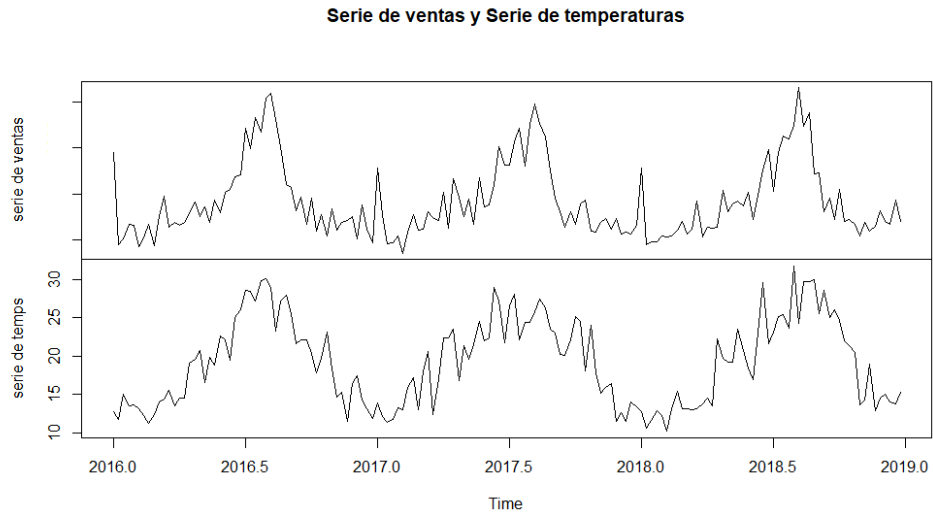


Figura 4.10: Gráficos secuenciales de las series de ventas y de temperaturas en Pontevedra.

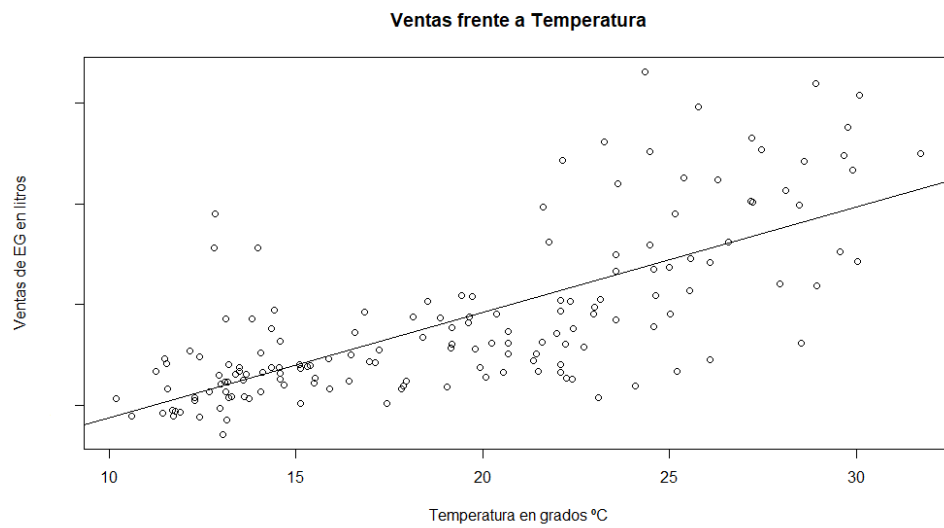


Figura 4.11: Nube de puntos y recta de regresión lineal obtenida al enfrentar las ventas semanales y sus temperaturas medias.

Hay que destacar que las observaciones utilizadas para esta regresión no son independientes, por lo que el nivel de significación final de los distintos coeficientes debe ser tomado con cierta prudencia. De todos modos, parece evidente que la temperatura

en esa semana sí afecta en las ventas totales.

Pasemos, por lo tanto, a estudiar la posible existencia de correlación entre las series. Como hemos comentado en la sección anterior, debemos trabajar con series estacionarias y además, luego, preblanquearlas. Como podemos intuir en los gráficos secuenciales de la Figura 4.10, nuestras series no son estacionarias. Además, sus funciones *fas* recogidas en la Figura 4.12 nos lo confirman. Es por ello que debemos diferenciarlas, teniendo siempre en cuenta hacer las mismas transformaciones en ambas para no perder la interpretabilidad de los resultados. Tras hacer una diferenciación regular, obtenemos las *fas* y *fap* de la Figura 4.13, concluyendo así que ya son estacionarias.

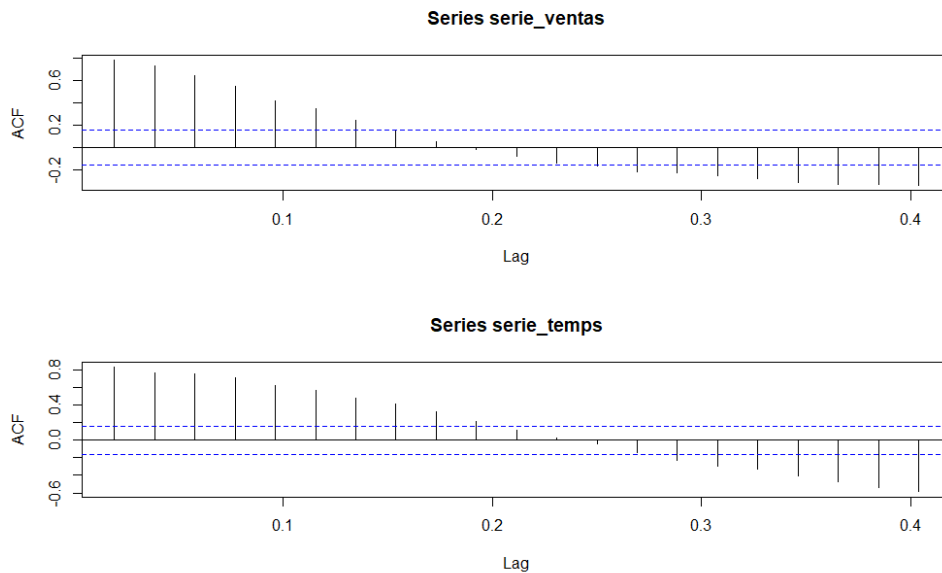


Figura 4.12: Función *fas* para las series de ventas y de temperaturas. Ambas tardan mucho en anularse, por lo que concluimos que ninguna de las series es estacionaria.

Estamos en condiciones ya de efectuar el preblanqueo y calcular las correlaciones cruzadas. Obtenemos el resultado mostrado en la Figura 4.14. Como podemos observar, consideramos significativa la correlación con retardo $r = -1$. Por lo tanto, el modelo de regresión que obtenemos es $\nabla(y_t) = \beta_0 + \beta_1 \cdot \nabla(x_{t-1}) + \eta_t$, siendo y_t las ventas en la semana t , x_t la temperatura media de esa semana y η_t el error de la regresión explicado mediante un modelo *ARIMA* que tendremos que ajustar a continuación.

Lo que debemos hacer ahora es encontrar el modelo *ARIMA* que se ajusta a

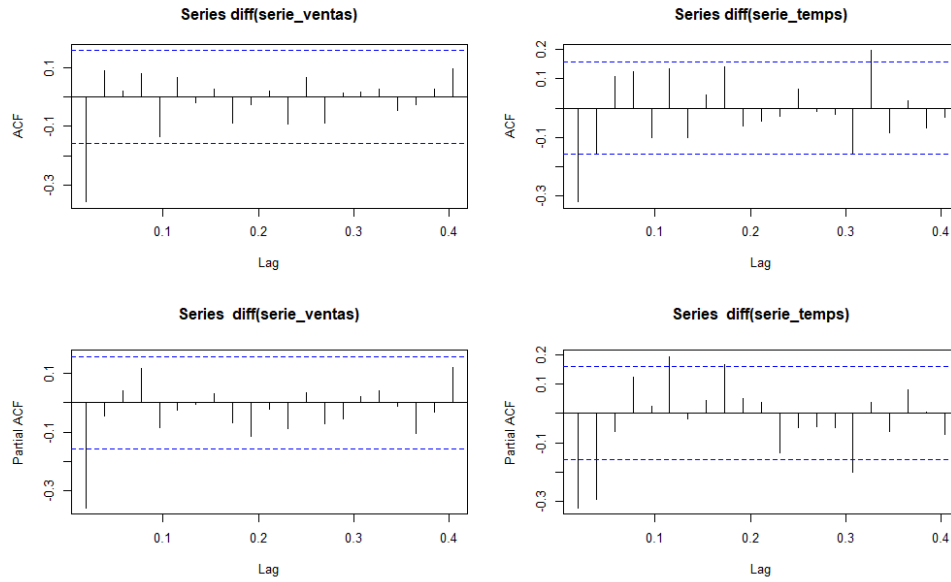


Figura 4.13: Funciones f_{as} y f_{ap} para las series diferenciadas de ventas y temperaturas. Ambas son estacionarias.

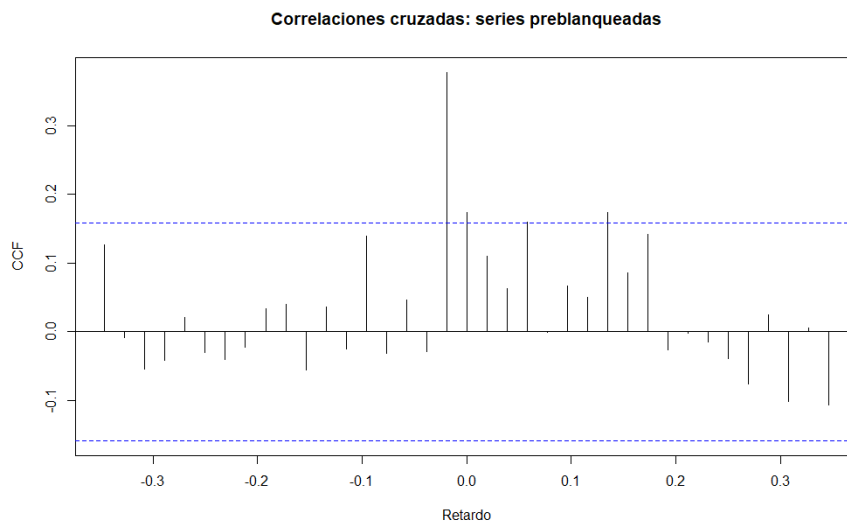


Figura 4.14: Función de correlación para las series diferenciadas de ventas y temperaturas. Existe correlación lineal con retardo $r = -1$.

nuestra serie de residuos $\{\eta_t\}$. Como podemos intuir en la Figura 4.15, dichos residuos son estacionarios. De todos modos, la Figura 4.16 recoge sus funciones f_{as} y f_{ap} con las que confirmamos que efectivamente son estacionarios y que, además, nos sirven para proponer el modelo $AR(1)$.

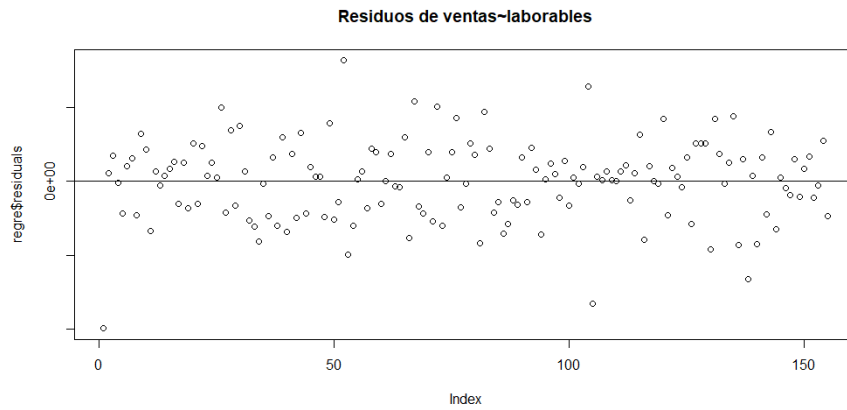


Figura 4.15: Representación gráfica de los residuos η_t de la regresión ventas frente temperaturas.

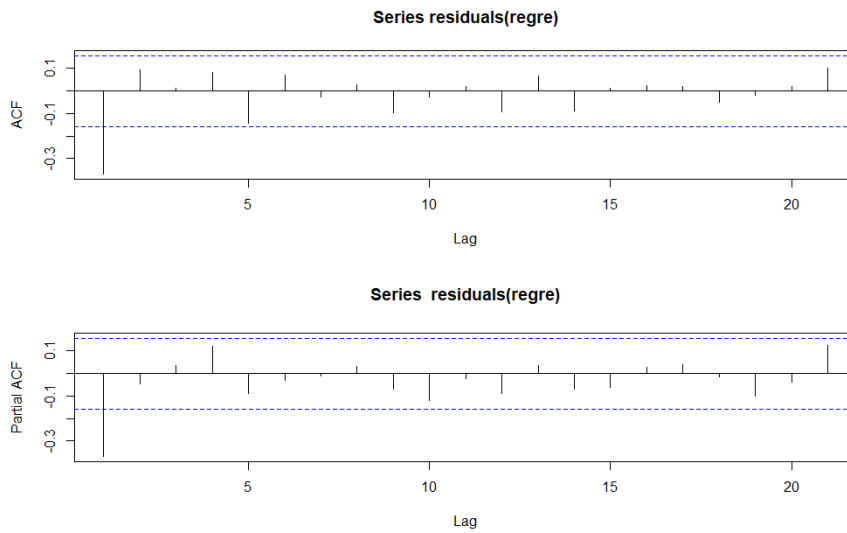


Figura 4.16: Funciones f_{as} y f_{ap} de los residuos η_t de la regresión ventas frente temperaturas. Vemos que son estacionarios y que podemos proponer el modelo $AR(1)$ como generador de ese proceso.

Así, una vez conocido el retardo ($r = -1$) y propuesto un modelo $AR(1)$ para los residuos, debemos ajustar en conjunto los coeficientes de nuestra regresión. Para ello, utilizamos la función `arima()` del paquete `forecast`, obteniendo la salida siguiente:

Call:

```
arima(x = diff(serie_ventas), order = c(1, 0, 0), xreg = diff(serie_temps))
```

Coefficients:

	ar1	intercept	xreg
	-0.4381	-266.891	1422.4466
s.e.	0.0786	1369.128	669.8395

sigma^2 estimated as 598210896: log likelihood = -1786.27, aic = 3578.55

Así, el modelo final que habíamos propuesto

$$\nabla(y_t) = \beta_0 + \beta_1 \cdot \nabla(x_{t-1}) + \eta_t,$$

$$\eta_t \sim AR(1).$$

se traduce en

$$y_t - y_{t-1} = -266,891 + 1422,4466(x_{t-1} - x_{t-2}) + \eta_t,$$

$$(1 - \phi_1 B)\eta_t = \varepsilon_t.$$

Por último, no hay que olvidar de checkear que los residuos finales $\{\varepsilon_t\}$ efectivamente son ruido blanco (ver Figura 4.17). El t-test no rechaza que la media sea nula (con un p-valor = 0.9136), el test de Jarque-Bera no rechaza asimetría (con un p-valor = 0.1575) y el test de Shapiro no rechaza normalidad (p-valor = 0.365). Concluimos que los tests fueron superados y que por lo tanto el modelo es válido.

Pasamos ya, por lo tanto, a calcular las predicciones. Para ello, como ya hemos explicado en el caso del número de días laborables, nosotros daremos los nuevos valores para la variable explicativa y, por otro lado, los nuevos residuos se calcularán siguiendo el modelo $AR(1)$ que los define siguiendo el procedimiento que hemos estudiado en el Capítulo 2. Tras calcular las predicciones en R con un horizonte $h = 5$ y suponiendo que las temperaturas de las siguientes semanas han sido (14,3, 14,3, 12,3, 12,9, 11,9), hemos obtenido los valores recogidos en el Cuadro 4.2 y Figura 4.18.

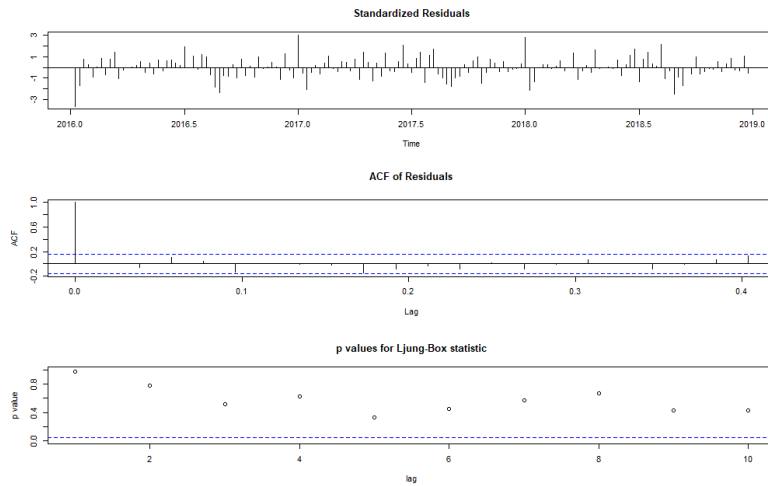


Figura 4.17: Representación gráfica de los test sobre los residuos del ajuste final.

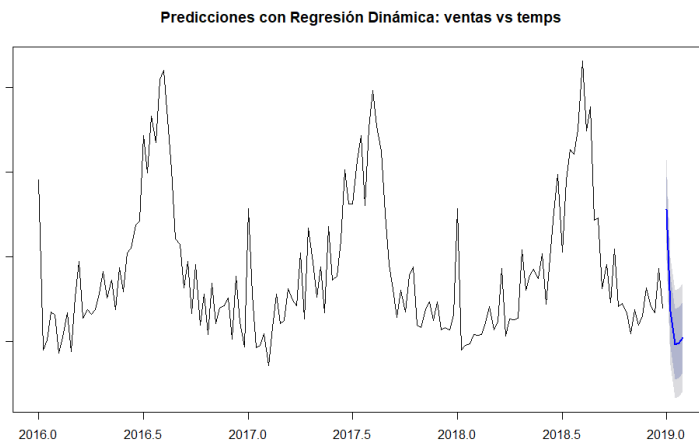


Figura 4.18: Representación gráfica de las predicciones con Regresión Dinámica de las ventas frente a las temperaturas.

	Point Forecast	Lo 80	Hi 80	Lo 95	Hi 95
29 Abr - 5 May	111672.12	94737.06	128607.18	85772.18	137572.05
6 May - 12 May	59090.65	40999.87	77181.43	31423.19	86758.11
13 May - 19 May	42100.72	23852.71	60348.73	14192.80	70008.64
20 May - 26 May	42873.42	24603.33	61143.51	14931.73	70815.12
27 May - 2 Jun	45676.88	27403.67	63950.09	17730.42	73623.34

Cuadro 4.2: Predicciones con Regresión Dinámica de las ventas frente a las temperaturas. Estos valores han sido transformados por un parámetro α para respetar la confidencialidad de los datos reales.

4.3. Series jerarquizadas

Como hemos explicado al inicio de este capítulo, por la estructura de mercado de nuestra empresa, las series de tiempo a menudo se pueden desagregar por distintos atributos, como la familia (cerveza, agua, vino, etc.), la marca (Estrella Galicia, 1906, etc.), el formato (lata, barril, vidrio, etc.), el envase (0.3l, 0.5l, 5l, etc) o incluso la zona geográfica. Esto hace que nuestras series sigan una estructura de agregación jerárquica.

En esta sección estudiaremos, siguiendo de cerca [Hyndman and Athanasopoulos \(2018\)](#), la predicción de grandes colecciones de series de tiempo que deben agregarse de algún modo. El objetivo de los métodos que veremos a continuación es que sus predicciones sean coherentes con la estructura jerárquica, es decir, que las predicciones resultantes tras la agregación sean realistas y fiables en cada nivel de la estructura jerárquica.

Empecemos modelando matemáticamente una estructura jerárquica, para luego poder definir claramente nuestro método de predicción. Lo haremos todo basado en el ejemplo de la Figura 4.19, donde se presenta una estructura jerárquica con dos niveles, $k = 2$. En la parte superior de la jerarquía, que llamaremos nivel 0, está el “total” y se corresponde con el nivel más agregado de los datos. Denotaremos por x_t a la observación de la serie total en el instante t . Luego, siguiendo con el ejemplo ilustrado, la serie total se desagrega en el nivel 1 en dos series distintas (A y B) que, a su vez, se desagregan en el nivel 2 en tres y dos series, respectivamente. Por debajo de nivel superior (nivel 0) denotaremos por $x_{j,t}$ la observación de la serie del nodo j en el instante t . Por último, denotaremos con la letra n el número total de series de nuestra estructura (en el ejemplo, $n = 8$) y con la letra m el número de series en el nivel más bajo de la estructura (en este caso, $m = 5$).

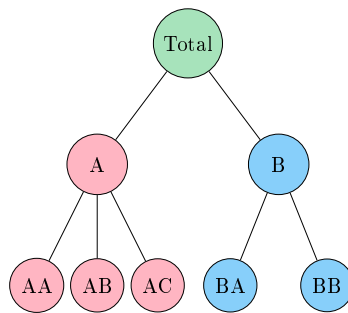


Figura 4.19: Diagrama de ejemplo de una jerarquía de series temporales.

Así, para todo instante t , se tiene la relación:

$$x_t = x_{AA,t} + x_{AB,t} + x_{AC,t} + x_{BA,t} + x_{BB,t},$$

y

$$x_{A,t} = x_{AA,t} + x_{AB,t} + x_{AC,t} \text{ y } x_{B,t} = x_{BA,t} + x_{BB,t}.$$

Equivalentemente, aunque más cómodo para representaciones generales futuras, la relaciones anteriores se pueden expresar matricialmente como sigue

$$\begin{pmatrix} x_t \\ x_{A,t} \\ x_{B,t} \\ x_{AA,t} \\ x_{AB,t} \\ x_{AC,t} \\ x_{BA,t} \\ x_{BB,t} \end{pmatrix} = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} x_{AA,t} \\ x_{AB,t} \\ x_{AC,t} \\ x_{BA,t} \\ x_{BB,t} \end{pmatrix}$$

o, abreviadamente,

$$\mathbf{x}_t = \mathbf{S}\mathbf{b}_t,$$

siendo \mathbf{x}_t el vector n -dimensional de todas las observaciones jerárquicas en el instante t y \mathbf{S} la matriz $n \times m$ que define la relación entre las distintas series.

Una vez definida la estructura de nuestras series, pasamos a la predicción de las mismas. Esto lo podemos llevar a cabo siguiendo varios enfoques que presentamos a continuación.

4.3.1. Enfoque “Bottom-up”

Este enfoque, como su nombre indica, consiste en generar en un primer momento las predicciones en cada una de las series del nivel más bajo y luego ir sumándolas para obtener las predicciones de todas las series de los demás niveles. En nuestro ejemplo guía, consistiría en predecir las observaciones con origen T y horizonte h : $\hat{x}_{AA,T}(h)$, $\hat{x}_{AB,T}(h)$, $\hat{x}_{AC,T}(h)$, $\hat{x}_{BA,T}(h)$ y $\hat{x}_{BB,T}(h)$. Así, para los niveles superiores

tendríamos:

$$\hat{x}_{A,T}(h) = \hat{x}_{AA,T}(h) + \hat{x}_{AB,T}(h) + \hat{x}_{AC,T}(h),$$

$$\hat{x}_{B,T}(h) = \hat{x}_{BA,T}(h) + \hat{x}_{BB,T}(h), \text{ y}$$

$$\hat{x}_T(h) = \hat{x}_{A,T}(h) + \hat{x}_{B,T}(h).$$

Equivalentemente, en forma matricial, se tendría

$$\begin{pmatrix} \tilde{x}_T(h) \\ \tilde{x}_{A,T}(h) \\ \tilde{x}_{B,T}(h) \\ \tilde{x}_{AA,T}(h) \\ \tilde{x}_{AB,T}(h) \\ \tilde{x}_{AC,T}(h) \\ \tilde{x}_{BA,T}(h) \\ \tilde{x}_{BB,T}(h) \end{pmatrix} = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{pmatrix} \begin{pmatrix} \hat{x}_{AA,T}(h) \\ \hat{x}_{AB,T}(h) \\ \hat{x}_{AC,T}(h) \\ \hat{x}_{BA,T}(h) \\ \hat{x}_{BB,T}(h) \end{pmatrix},$$

también escrito

$$\tilde{x}_T(h) = \mathbf{S}\hat{\mathbf{b}}_T(h).$$

Notación 4.4. Nótese que ahora distinguimos la notación entre las predicciones “base” (las que han sido calculadas en un primer momento) mediante el símbolo \hat{x} , y las predicciones “coherentes” obtenidas a partir de las anteriores mediante la notación \tilde{x} .

Una ventaja de este enfoque es que predice a partir del nivel más bajo de la estructura jerárquica, lo que permite que no se pierda información al hacer la agregación. Sin embargo, y como veremos más adelante con un ejemplo práctico con datos reales, los niveles más bajos son bastante ruidosos, por lo que su estimación se dificulta y la estimación final global acaba disparándose.

4.3.2. Enfoque “Top-down”

A diferencia del anterior enfoque, como se deduce por su nombre, ahora las primeras predicciones se harán sobre la serie “total” del nivel 0 y, en función de una serie

de pesos, se repartirá la predicción puntual de cada instante t entre las series de los niveles inferiores. Sea $\mathbf{p} = (p_1, \dots, p_m)'$ el vector de pesos que tiene cada una de las series desagregadas en la serie total. En nuestro ejemplo de la Figura 4.19, se tendría que

$$\tilde{x}_{AA,T}(h) = p_1 \hat{x}_T(h),$$

$$\tilde{x}_{AB,T}(h) = p_2 \hat{x}_T(h),$$

$$\tilde{x}_{AC,T}(h) = p_3 \hat{x}_T(h),$$

$$\tilde{x}_{BA,T}(h) = p_4 \hat{x}_T(h),$$

$$\tilde{x}_{BB,T}(h) = p_5 \hat{x}_T(h).$$

Matricialmente, seguiríamos la expresión

$$\tilde{\mathbf{x}}_T(h) = \mathbf{S}\mathbf{p}\hat{x}_T(h).$$

Lo importante en este enfoque es calcular los pesos que recibirá cada una de las series. Para ello, existen tres métodos distintos.

Proporciones históricas medias

Es el cálculo más intuitivo de los tres. Consiste en calcular qué proporción del total ha supuesto cada una de las series en cada instante t . Así,

$$p_i = \frac{1}{T} \sum_{t=1}^T \frac{x_{i,t}}{x_t},$$

con $i = 1, \dots, m$ y $t = 1, \dots, T$.

Proporciones de las medias históricas

En el caso anterior promediábamos la proporción que ha supuesto una serie desagregada en la serie total a lo largo de los tiempos t . Ahora, lo que haremos será calcular la media histórica de cada serie y ver qué proporción supone sobre la media histórica de la serie total. Así,

$$p_i = \frac{\sum_{t=1}^T \frac{x_{i,t}}{T}}{\sum_{t=1}^T \frac{x_t}{T}},$$

con $i = 1, \dots, m$.

Proporciones estimadas

Los dos anteriores métodos utilizan el histórico para obtener una proporción, pero sin tener en cuenta cómo estas van evolucionando a lo largo del tiempo. Esto hace que el enfoque “top-down” realice predicciones menos precisas en los niveles más bajos que el enfoque “bottom-up”. Para solucionar esto, se pueden utilizar proporciones basadas en pronósticos en lugar de datos históricos. El siguiente método fue propuesto en Athanasopoulos et al. (2009).

Consideremos el primer nivel de la jerarquía. En primer lugar calcularemos las predicciones con horizonte h para todas las series de este nivel, pero no las sumaremos ya que su agregación no será coherente. A estas predicciones las llamaremos “predicciones iniciales”. Lo que haremos con ellas será calcular la proporción que suponen sobre el total agregado de las predicciones iniciales de ese nivel. A esto lo llamaremos “proporciones estimadas” y las usaremos para desagregar las estimaciones iniciales en sus subniveles y obtener predicciones en la totalidad de la jerarquía. Así, podemos definir la proporción estimada de la serie j -ésima del nivel K como

$$p_j = \prod_{l=0}^{K-1} \frac{\hat{x}_{j,h}^{(l)}}{\hat{S}_{j,h}^{(l+1)}},$$

donde $j = 1, \dots, m$, $\hat{x}_{j,h}^{(l)}$ es la predicción inicial con horizonte h de la serie que está l niveles por encima de j y $\hat{S}_{j,h}^{(l+1)}$ es la suma de las predicciones iniciales con horizonte h de todas las series que están por debajo y directamente conectadas a la serie que está l niveles por encima de j .

4.3.3. Enfoque “Middle-out”

Este último enfoque combina técnicas del enfoque ascendente y descendente anterior. Empieza escogiendo un “nivel medio” para el que se calculan las predicciones iniciales de sus series y, luego, para las series de niveles superiores, se generan predicciones coherentes utilizando el enfoque “bottom-up” y para las series de los niveles inferiores se generan con el enfoque “top-down”. La implementación de R con el paquete `forecast` realiza la desagregación para niveles inferiores con el método de proporciones estimadas.

4.3.4. Aplicación de las series jerárquicas a Estrella Galicia

Para ilustrar este modelo, hemos considerado las ventas totales en litros de cerveza que vende nuestra compañía. Nuevamente, por confidencialidad, hemos decidido no mencionar los nombres reales de los componentes en los que se desagregará la jerarquía. Como podemos ver en la Figura 4.20, los litros totales vendidos los distinguiremos, en primer lugar, en función de la marca de la cerveza (distinguiremos entre

la Marca 1, Marca 2 y Marca 3) y, para cada una de ellas, los formatos en los que fueron vendidas (recordemos que siguiendo la nomenclatura de las Bases de Datos de la empresa, el formato denota si fue vendido en “Barril”, “Lata”, “Vidrio retornable”, etc.). Consideraremos, por tanto, hasta cinco tipos distintos de formatos en nuestra jerarquía: Formato A, Formato B, Formato C, Formato D y Formato E.

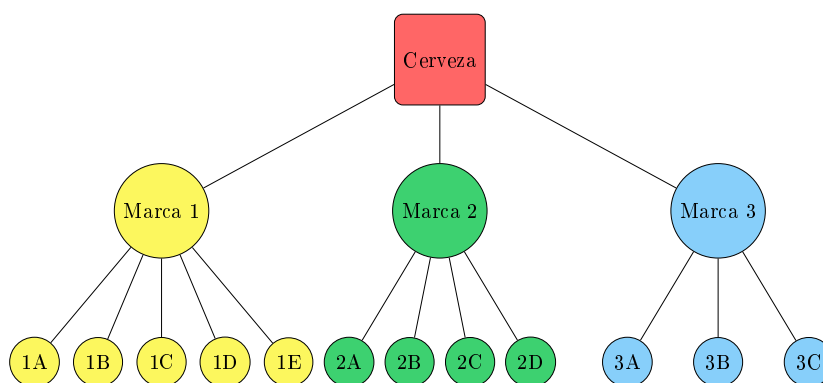


Figura 4.20: Árbol jerárquico que representa la estructura de agregación de nuestras series.

Por lo tanto, una vez clara la estructura, bastará con definir las series del nivel más bajo para luego, aplicando la función `hts` del paquete homónimo de R, obtener predicciones coherentes para toda la jerarquía.

Cabe destacar que, al igual que hay que seleccionar una metodología de agregación y desagregación de las predicciones (lo que hemos comentado en las subsecciones anteriores), también deberemos escoger un modelo para calcular las predicciones base iniciales. Para ello tenemos dos opciones, las presentadas en los anteriores capítulos de este trabajo: los modelos *ARIMA* y los modelos *ETS*. Tras hacer distintas pruebas, como estamos trabajando con un ejemplo muy simplificado de la estructura real de la empresa, no hemos apreciado diferencia relevante entre la utilización de un método u otro para la agregación y desagregación de nuestra jerarquía, pero sí en la utilización de modelos *ARIMA* o *ETS* para el cálculo de las predicciones base.

En la Figura 4.21, se muestran los resultados obtenidos tras la predicción realizada con el enfoque “top-down” y modelos *ETS*. En la Figura 4.22, se representan las calculadas con el enfoque “top-down” y modelos *ARIMA*. Como podemos observar, los modelos *ARIMA* devuelven predicciones más realistas. Las predicciones obtenidas para las series más desagregadas son las mostradas en el Cuadro 4.3 y para tener las de las series más agregadas bastaría con sumar como corresponde, pues sus predicciones ya han sido calculadas de forma que las agregaciones sean coherentes.

Figura 4.21: Predicciones jerárquicas mediante el método “top-down” y los modelos *ETS*.Figura 4.22: Predicciones jerárquicas mediante el método “top-down” y los modelos *ARIMA*.

	1A	1B	1C	1D	1E		
2019-15	22724.62	58538.73	115923.11	135403.96	100368.12		
2019-16	22025.32	58927.59	119686.25	138134.07	95683.68		
2019-17	23749.12	61407.78	141647.41	144325.00	117533.49		
2019-18	23703.87	61595.73	126146.64	130312.07	114834.02		
2019-19	23844.96	64933.93	134675.59	142896.85	97670.95		
2019-20	22695.76	60590.61	127780.59	134090.33	103622.64		
2019-21	23164.08	68817.77	131208.08	144472.47	105952.88		
2019-22	22302.05	62774.74	133460.55	133101.37	107710.89		
	2A	2B	2C	2D	3A	3B	3C
2019-15	3977.28	15099.12	13023.18	26680.75	3785.44	12173.95	8149.48
2019-16	4857.65	15798.72	14488.07	26088.60	4132.29	12594.40	7511.74
2019-17	5769.67	17154.15	15149.60	27939.66	3986.80	15389.34	8401.33
2019-18	3982.95	15336.94	13509.06	27449.87	4136.95	14248.26	8707.53
2019-19	4561.90	16518.09	15236.53	24190.74	4108.44	15239.48	7981.89
2019-20	5375.96	15199.00	14131.54	26619.52	3913.62	13868.67	8501.70
2019-21	4643.56	16034.43	15025.11	25918.35	4502.11	13764.84	8739.77
2019-22	4685.67	15528.33	13971.78	26385.01	3901.20	13963.65	8938.00

Cuadro 4.3: Predicciones puntuales coherentes para las series del nivel más desagregado de la jerarquía. Estos valores han sido transformados por un parámetro α para respetar la confidencialidad de los datos reales.

Capítulo 5

Implementación y conclusiones

Tras todo lo estudiado en los capítulos anteriores, hemos llegado al punto final en el que debemos escoger el mejor enfoque de predicción para implementar en Estrella Galicia. Como hemos comentado anteriormente, está especialmente interesada en la predicción a corto plazo (diaria o semanalmente), por lo que la idea de limitarnos a los métodos clásicos como los modelos *ARIMA* o los modelos *ETS* no es suficiente para nuestra cervecera debido a varios motivos.

- Por una parte, las series semanales y diarias han empezado a presentar una volatilidad muy considerable en los últimos años, lo que hace difícil a estos modelos, por sí solos, aportar intervalos de confianza lo suficientemente ajustados como para ser útiles en la toma de decisión de la empresa. Esto, sumado al concepto de estructura jerárquica que tiene nuestra empresa, hace que si calculamos las predicciones finales como una simple agregación de las predicciones base, las últimas acaben siendo mucho más elevadas de lo que cabría esperar, debido justamente a la suma de error acumulado en cada predicción base.
- Por otra parte, como es común en el estudio de series temporales semanales y diarias, es muy complicado que la estacionalidad sea capaz de recoger ciertos factores clave decisivo en las ventas. Como hemos comentado en el capítulo anterior, el cambio de fecha continuo de algunas festividades como Semana Santa o la posición en el día de la semana de otras como Fin de año, hace que las ventas de esas semanas varíen considerablemente y que no se repita un patrón continuo año tras año.

Es por ello que hemos considerado modelos más complejos como las Series Jerárquicas o la Regresión Dinámica. El problema que nos hemos encontrado a la hora de implementar la Regresión Dinámica, es que tendríamos que disponer de una base de datos lo suficientemente extensa, tanto en histórico como en localizaciones geográficas.

cas, como para poder realizar una regresión sobre nuestro registro de ventas y, luego, que se actualizase casi a diario para poder hacer unas predicciones a corto plazo.

Pero disponer de una base de datos sobre las temperaturas no es el único punto crítico de este enfoque. Algo muy importante a tener en cuenta sobre los datos disponibles en Estrella Galicia es que las ventas no son a nivel consumidor, sino a nivel de fábrica. El nivel más bajo que queda registrado es el del proveedor local (o cadena de supermercados) al que se ha efectuado la venta, proveedor o cadena que luego se encargará de distribuir la mercancía a los bares o supermercados que se lo soliciten. Es por ello que si queremos considerar un modelo con temperaturas, lo más preciso sería considerar las temperaturas por localidad, pero a este nivel las series acaban siendo picos seguidos de muchos ceros: el proveedor se aprovisiona con la mercancía (registramos una venta y se produce un pico en la serie) y luego lo va distribuyendo entre sus clientes en los días o semanas siguientes (distribución de la que no disponemos de información y, por lo tanto, para nosotros suponen ceros en la serie de ventas). Esto hace inviable la utilización, a nivel local, de un modelo como la Regresión Dinámica que enfrente ventas con temperaturas. Por lo tanto, tendríamos que considerar, por ejemplo, las ventas a nivel de Delegación, con la pérdida de información en que eso supone, pues empezariamos a tener que considerar la temperatura media entre zonas con registros que pueden ser bastante distintos.

Por todo esto, hemos considerado que la mejor opción es la implementación de una estructura jerárquica que nos permita calcular unas predicciones puntuales fiables y, además, a todos los niveles en los que la empresa está interesada.

La estructura que finalmente hemos implementado es la siguiente:

Nivel $k = 0$: En el nivel más alto y agregado, tendremos los litros totales de cerveza en general.

Nivel $k = 1$: En el primer nivel de la jerarquía tendremos las distintas marcas de cerveza que gestiona la cervecera: Estrella Galicia, 1906, Estrella Galicia Light, Estrella Galicia Pilsen, Estrella Galicia 0,0 y Shandy Estrella Galicia.

Nivel $k = 2$: En el segundo nivel distinguiremos, dentro de cada marca, los formatos en los que fueron vendidos los litros de cerveza. En la globalidad de este nivel encontraremos los nodos: Barril, Barril no retornable, Box, Lata, Tanque, Vidrio no retornable y Vidrio retornable.

Nivel $k = 3$: Finalmente, en el nivel más bajo de la jerarquía, posicionaremos el material en el que fue embotellada la cerveza. En total, en este nivel tendremos más de 150 nodos. Algunos ejemplos pueden ser: “EG BARRIL 50 L”, “1906 NR 24 LATA PACK 33CL”, “SHANDY 24B 4X6 PACK 25CL”, etc.

Por motivos de confidencialidad, no enseñaremos el código, ya que es difícil no mostrar cierta información interna, pero sí explicaremos brevemente el esqueleto de la implementación:

1. La información de entrada consistirá en un archivo `.csv`, generado con SQL en conexión con las Bases de Datos de Estrella Galicia, que recogerá el registro en bruto de todas las ventas de cerveza en litros.
2. Dicho archivo de entrada será tratado mediante el script de python `etl.py`. En él, se recogerá la información de entrada y se transformará para adaptarla al modelo que luego queremos aplicar. Por ejemplo, para que nuestro modelo jerárquico sea consistente, es necesario que todas las series tengan la misma longitud y estén definidas en el mismo período temporal. Sin embargo, en el histórico de ventas podemos apreciar como no todas las semanas se venden todos los materiales, esto genera que las series no tengan la misma longitud y que tengamos que “rellenarlas”.
3. Una vez tratados nuestros datos y almacenados de forma cómoda para su explotación, elaboramos el script de R, llamado `modelo.R`, que recogerá nuestras series y calculará las predicciones de forma coherente siguiendo los parámetros que el usuario haya escogido. Por defecto, hemos decidido que el “método jerárquico” sea el Middle-out, que el “método individual” sean los modelos *ARIMA* y que el horizonte de predicción sea $h = 8$. Una vez haya acabado de predecir, los resultados se escribirán en el fichero `predicciones_borrador.txt`.
4. Finalmente, la tabla con las predicciones puntuales que nos devuelve R tendremos que transformarla para adaptarla nuevamente a la estructura de las Bases de Datos de la empresa. En concreto, lo que haremos con el script de python `escritura_final.py` será recuperar los códigos originales de marca, formato y material de cada producto y escribir cada combinación al lado de sus predicciones correspondientes, quedando así listo el archivo `predicciones_ok.txt` para el uso de la empresa.

Apéndice A

Tablas recopilatorias de los modelos ETS

Trend	Seasonal		
	N	A	M
N	$\ell_t = \alpha x_t + (1 - \alpha)\ell_{t-1}$	$\ell_t = \alpha(x_t - s_{t-m}) + (1 - \alpha)\ell_{t-1}$	$\ell_t = \alpha \left(\frac{x_t}{s_{t-m}} \right) + (1 - \alpha)\ell_{t-1}$
	$s_t = \gamma(\alpha x_t - \ell_{t-1}) + (1 - \gamma)s_{t-m}$	$s_t = \gamma(\alpha x_t - \ell_{t-1}) + (1 - \gamma)s_{t-m}$	$s_t = \gamma \left(\frac{x_t}{\ell_{t-1}} \right) + (1 - \gamma)s_{t-m}$
	$\hat{x}_{t+h t} = \ell_t$	$\hat{x}_{t+h t} = \ell_t + s_{t-m+h_m^*}$	$\hat{x}_{t+h t} = \ell_t s_{t-m+h_m^*}$
A	$\ell_t = \alpha x_t + (1 - \alpha)(\ell_{t-1} + b_{t-1})$	$\ell_t = \alpha(\alpha x_t - s_{t-m}) + (1 - \alpha)(\ell_{t-1} + b_{t-1})$	$\ell_t = \alpha \left(\frac{x_t}{s_{t-m}} \right) + (1 - \alpha)(\ell_{t-1} + b_{t-1})$
	$b_t = \beta(\ell_t - \ell_{t-1}) + (1 - \beta)b_{t-1}$	$b_t = \beta(\ell_t - \ell_{t-1}) + (1 - \beta)b_{t-1}$	$b_t = \beta(\ell_t - \ell_{t-1}) + (1 - \beta)b_{t-1}$
	$s_t = \gamma(\alpha x_t - \ell_{t-1} - b_{t-1}) + (1 - \gamma)s_{t-m}$	$s_t = \gamma(\alpha x_t - \ell_{t-1} - b_{t-1}) + (1 - \gamma)s_{t-m}$	$s_t = \gamma \left(\frac{x_t}{\ell_{t-1} + b_{t-1}} \right) + (1 - \gamma)s_{t-m}$
	$\hat{x}_{t+h t} = \ell_t + hb_t$	$\hat{x}_{t+h t} = \ell_t + hb_t + s_{t-m+h_m^*}$	$\hat{x}_{t+h t} = (\ell_t + hb_t)s_{t-m+h_m^*}$
A _d	$\ell_t = \alpha x_t + (1 - \alpha)(\ell_{t-1} + \phi b_{t-1})$	$\ell_t = \alpha(\alpha x_t - s_{t-m}) + (1 - \alpha)(\ell_{t-1} + \phi b_{t-1})$	$\ell_t = \alpha \left(\frac{x_t}{s_{t-m}} \right) + (1 - \alpha)(\ell_{t-1} + \phi b_{t-1})$
	$b_t = \beta(\ell_t - \ell_{t-1}) + (1 - \beta)\phi b_{t-1}$	$b_t = \beta(\ell_t - \ell_{t-1}) + (1 - \beta)\phi b_{t-1}$	$b_t = \beta(\ell_t - \ell_{t-1}) + (1 - \beta)\phi b_{t-1}$
	$s_t = \gamma(\alpha x_t - \ell_{t-1} - \phi b_{t-1}) + (1 - \gamma)s_{t-m}$	$s_t = \gamma(\alpha x_t - \ell_{t-1} - \phi b_{t-1}) + (1 - \gamma)s_{t-m}$	$s_t = \gamma \left(\frac{x_t}{\ell_{t-1} + \phi b_{t-1}} \right) + (1 - \gamma)s_{t-m}$
	$\hat{x}_{t+h t} = \ell_t + \phi_h b_t$	$\hat{x}_{t+h t} = \ell_t + \phi_h b_t + s_{t-m+h_m^*}$	$\hat{x}_{t+h t} = (\ell_t + \phi_h b_t)s_{t-m+h_m^*}$
M	$\ell_t = \alpha x_t + (1 - \alpha)(\ell_{t-1} b_{t-1})$	$\ell_t = \alpha(\alpha x_t - s_{t-m}) + (1 - \alpha)(\ell_{t-1} b_{t-1})$	$\ell_t = \alpha \left(\frac{x_t}{s_{t-m}} \right) + (1 - \alpha)(\ell_{t-1} b_{t-1})$
	$b_t = \beta \left(\frac{\ell_t}{\ell_{t-1}} \right) + (1 - \beta)b_{t-1}$	$b_t = \beta \left(\frac{\ell_t}{\ell_{t-1}} \right) + (1 - \beta)b_{t-1}$	$b_t = \beta \left(\frac{\ell_t}{\ell_{t-1}} \right) + (1 - \beta)b_{t-1}$
	$s_t = \gamma(\alpha x_t - \ell_{t-1} b_{t-1}) + (1 - \gamma)s_{t-m}$	$s_t = \gamma(\alpha x_t - \ell_{t-1} b_{t-1}) + (1 - \gamma)s_{t-m}$	$s_t = \gamma \left(\frac{x_t}{\ell_{t-1} b_{t-1}} \right) + (1 - \gamma)s_{t-m}$
	$\hat{x}_{t+h t} = \ell_t b_t^h$	$\hat{x}_{t+h t} = \ell_t b_t^h + s_{t-m+h_m^*}$	$\hat{x}_{t+h t} = \ell_t b_t^h s_{t-m+h_m^*}$
M _d	$\ell_t = \alpha x_t + (1 - \alpha)(\ell_{t-1} b_{t-1}^{\phi})$	$\ell_t = \alpha(\alpha x_t - s_{t-m}) + (1 - \alpha)(\ell_{t-1} b_{t-1}^{\phi})$	$\ell_t = \alpha \left(\frac{x_t}{s_{t-m}} \right) + (1 - \alpha)(\ell_{t-1} b_{t-1}^{\phi})$
	$b_t = \beta \left(\frac{\ell_t}{\ell_{t-1}} \right) + (1 - \beta)b_{t-1}^{\phi}$	$b_t = \beta \left(\frac{\ell_t}{\ell_{t-1}} \right) + (1 - \beta)b_{t-1}^{\phi}$	$b_t = \beta \left(\frac{\ell_t}{\ell_{t-1}} \right) + (1 - \beta)b_{t-1}^{\phi}$
	$s_t = \gamma(\alpha x_t - \ell_{t-1} b_{t-1}^{\phi}) + (1 - \gamma)s_{t-m}$	$s_t = \gamma(\alpha x_t - \ell_{t-1} b_{t-1}^{\phi}) + (1 - \gamma)s_{t-m}$	$s_t = \gamma \left(\frac{x_t}{\ell_{t-1} b_{t-1}^{\phi}} \right) + (1 - \gamma)s_{t-m}$
	$\hat{x}_{t+h t} = \ell_t b_t^{\phi h}$	$\hat{x}_{t+h t} = \ell_t b_t^{\phi h} + s_{t-m+h_m^*}$	$\hat{x}_{t+h t} = \ell_t b_t^{\phi h} s_{t-m+h_m^*}$

Cuadro A.1: Fórmulas para el cálculo recursivo de las predicciones puntuales de todos los métodos.

Trend	Seasonal			
	N	A	M	M
N	$\mu_t = \ell_{t-1}$	$\mu_t = \ell_{t-1} + s_{t-m}$	$\mu_t = \ell_{t-1} s_{t-m}$	$\mu_t = \ell_{t-1} s_{t-m}$
	$\ell_t = \ell_{t-1} + \alpha \varepsilon_t$	$\ell_t = \ell_{t-1} + \alpha \varepsilon_t$	$\ell_t = \ell_{t-1} + \alpha \frac{\varepsilon_t}{s_{t-m}}$	$\ell_t = \ell_{t-1} + \alpha \frac{\varepsilon_t}{s_{t-m}}$
		$s_t = s_{t-m} + \gamma \varepsilon_t$		$s_t = s_{t-m} + \gamma \frac{\varepsilon_t}{\ell_{t-1}}$
A	$\mu_t = \ell_{t-1} + b_{t-1}$	$\mu_t = \ell_{t-1} + b_{t-1} + s_{t-m}$	$\mu_t = (\ell_{t-1} + b_{t-1}) s_{t-m}$	$\mu_t = (\ell_{t-1} + b_{t-1}) s_{t-m}$
	$\ell_t = \ell_{t-1} + b_{t-1} + \alpha \varepsilon_t$	$\ell_t = \ell_{t-1} + b_{t-1} + \alpha \varepsilon_t$	$\ell_t = \ell_{t-1} + b_{t-1} + \alpha \frac{\varepsilon_t}{s_{t-m}}$	$\ell_t = \ell_{t-1} + b_{t-1} + \alpha \frac{\varepsilon_t}{s_{t-m}}$
	$b_t = b_{t-1} + \beta \varepsilon_t$	$b_t = b_{t-1} + \beta \varepsilon_t$	$b_t = b_{t-1} + \beta \frac{\varepsilon_t}{s_{t-m}}$	$b_t = b_{t-1} + \beta \frac{\varepsilon_t}{s_{t-m}}$
		$s_t = s_{t-m} + \gamma \varepsilon_t$	$s_t = s_{t-m} + \gamma \frac{\varepsilon_t}{\ell_{t-1} + b_{t-1}}$	$s_t = s_{t-m} + \gamma \frac{\varepsilon_t}{\ell_{t-1} + b_{t-1}}$
Ad	$\mu_t = \ell_{t-1} + \phi b_{t-1}$	$\mu_t = \ell_{t-1} + \phi b_{t-1} + s_{t-m}$	$\mu_t = (\ell_{t-1} + \phi b_{t-1}) s_{t-m}$	$\mu_t = (\ell_{t-1} + \phi b_{t-1}) s_{t-m}$
	$\ell_t = \ell_{t-1} + \phi b_{t-1} + \alpha \varepsilon_t$	$\ell_t = \ell_{t-1} + \phi b_{t-1} + \alpha \varepsilon_t$	$\ell_t = \ell_{t-1} + \phi b_{t-1} + \alpha \frac{\varepsilon_t}{s_{t-m}}$	$\ell_t = \ell_{t-1} + \phi b_{t-1} + \alpha \frac{\varepsilon_t}{s_{t-m}}$
	$b_t = \phi b_{t-1} + \beta \varepsilon_t$	$b_t = \phi b_{t-1} + \beta \varepsilon_t$	$b_t = \phi b_{t-1} + \beta \frac{\varepsilon_t}{s_{t-m}}$	$b_t = \phi b_{t-1} + \beta \frac{\varepsilon_t}{s_{t-m}}$
		$s_t = s_{t-m} + \gamma \varepsilon_t$	$s_t = s_{t-m} + \gamma \frac{\varepsilon_t}{\ell_{t-1} + \phi b_{t-1}}$	$s_t = s_{t-m} + \gamma \frac{\varepsilon_t}{\ell_{t-1} + \phi b_{t-1}}$
M	$\mu_t = \ell_{t-1} b_{t-1}$	$\mu_t = \ell_{t-1} b_{t-1} + s_{t-m}$	$\mu_t = \ell_{t-1} b_{t-1} s_{t-m}$	$\mu_t = \ell_{t-1} b_{t-1} s_{t-m}$
	$\ell_t = \ell_{t-1} b_{t-1} + \alpha \varepsilon_t$	$\ell_t = \ell_{t-1} b_{t-1} + \alpha \varepsilon_t$	$\ell_t = \ell_{t-1} b_{t-1} + \alpha \frac{\varepsilon_t}{s_{t-m}}$	$\ell_t = \ell_{t-1} b_{t-1} + \alpha \frac{\varepsilon_t}{s_{t-m}}$
	$b_t = b_{t-1} + \beta \frac{\varepsilon_t}{\ell_{t-1}}$	$b_t = b_{t-1} + \beta \frac{\varepsilon_t}{\ell_{t-1}}$	$b_t = b_{t-1} + \beta \frac{\varepsilon_t}{s_{t-m} \ell_{t-1}}$	$b_t = b_{t-1} + \beta \frac{\varepsilon_t}{s_{t-m} \ell_{t-1}}$
		$s_t = s_{t-m} + \gamma \varepsilon_t$	$s_t = s_{t-m} + \gamma \frac{\varepsilon_t}{\ell_{t-1} b_{t-1}}$	$s_t = s_{t-m} + \gamma \frac{\varepsilon_t}{\ell_{t-1} b_{t-1}}$
Md	$\mu_t = \ell_{t-1} b_{t-1}^\phi$	$\mu_t = \ell_{t-1} b_{t-1}^\phi + s_{t-m}$	$\mu_t = \ell_{t-1} b_{t-1}^\phi s_{t-m}$	$\mu_t = \ell_{t-1} b_{t-1}^\phi s_{t-m}$
	$\ell_t = \ell_{t-1} b_{t-1}^\phi + \alpha \varepsilon_t$	$\ell_t = \ell_{t-1} b_{t-1}^\phi + \alpha \varepsilon_t$	$\ell_t = \ell_{t-1} b_{t-1}^\phi + \alpha \frac{\varepsilon_t}{s_{t-m}}$	$\ell_t = \ell_{t-1} b_{t-1}^\phi + \alpha \frac{\varepsilon_t}{s_{t-m}}$
	$b_t = b_{t-1}^\phi + \beta \frac{\varepsilon_t}{\ell_{t-1}}$	$b_t = b_{t-1}^\phi + \beta \frac{\varepsilon_t}{\ell_{t-1}}$	$b_t = b_{t-1}^\phi + \beta \frac{\varepsilon_t}{s_{t-m} \ell_{t-1}}$	$b_t = b_{t-1}^\phi + \beta \frac{\varepsilon_t}{s_{t-m} \ell_{t-1}}$
		$s_t = s_{t-m} + \gamma \varepsilon_t$	$s_t = s_{t-m} + \gamma \frac{\varepsilon_t}{\ell_{t-1} b_{t-1}^\phi}$	$s_t = s_{t-m} + \gamma \frac{\varepsilon_t}{\ell_{t-1} b_{t-1}^\phi}$

Cuadro A.2: Fórmulas para todos los modelos ETS con errores aditivos.

Trend	Seasonal		
	N	A	M
N	$\mu_t = \ell_{t-1}$	$\mu_t = \ell_{t-1} + s_{t-m}$	$\mu_t = \ell_{t-1} s_{t-m}$
	$\ell_t = \ell_{t-1}(1 + \alpha \varepsilon_t)$	$\ell_t = \ell_{t-1} + \alpha(\ell_{t-1} + s_{t-m})\varepsilon_t$	$\ell_t = \ell_{t-1}(1 + \alpha \varepsilon_t)$
	$s_t = s_{t-m} + \gamma(\ell_{t-1} + s_{t-m})\varepsilon_t$	$s_t = s_{t-m} + \gamma(\ell_{t-1} + s_{t-m})\varepsilon_t$	$s_t = s_{t-m}(1 + \gamma \varepsilon_t)$
A	$\mu_t = \ell_{t-1} + b_{t-1}$	$\mu_t = \ell_{t-1} + b_{t-1} + s_{t-m}$	$\mu_t = (\ell_{t-1} + b_{t-1})s_{t-m}$
	$\ell_t = (\ell_{t-1} + b_{t-1})(1 + \alpha \varepsilon_t)$	$\ell_t = \ell_{t-1} + b_{t-1} + \alpha(\ell_{t-1} + b_{t-1} + s_{t-m})\varepsilon_t$	$\ell_t = (\ell_{t-1} + b_{t-1})(1 + \alpha \varepsilon_t)$
	$b_t = b_{t-1} + \beta(\ell_{t-1} + b_{t-1})\varepsilon_t$	$b_t = b_{t-1} + \beta(\ell_{t-1} + b_{t-1} + s_{t-m})\varepsilon_t$	$b_t = b_{t-1} + \beta(\ell_{t-1} + b_{t-1})\varepsilon_t$
A_d	$\mu_t = \ell_{t-1} + \phi b_{t-1}$	$\mu_t = \ell_{t-1} + \phi b_{t-1} + s_{t-m}$	$\mu_t = (\ell_{t-1} + \phi b_{t-1})s_{t-m}$
	$\ell_t = (\ell_{t-1} + \phi b_{t-1})(1 + \alpha \varepsilon_t)$	$\ell_t = \ell_{t-1} + \phi b_{t-1} + \alpha(\ell_{t-1} + \phi b_{t-1} + s_{t-m})\varepsilon_t$	$\ell_t = (\ell_{t-1} + \phi b_{t-1})(1 + \alpha \varepsilon_t)$
	$b_t = \phi b_{t-1} + \beta(\ell_{t-1} + \phi b_{t-1})\varepsilon_t$	$b_t = \phi b_{t-1} + \beta(\ell_{t-1} + \phi b_{t-1} + s_{t-m})\varepsilon_t$	$b_t = \phi b_{t-1} + \beta(\ell_{t-1} + \phi b_{t-1})\varepsilon_t$
M	$\mu_t = \ell_{t-1} b_{t-1}$	$\mu_t = \ell_{t-1} b_{t-1} + s_{t-m}$	$\mu_t = \ell_{t-1} b_{t-1} s_{t-m}$
	$\ell_t = \ell_{t-1} b_{t-1} + (1 + \alpha \varepsilon_t)$	$\ell_t = \ell_{t-1} b_{t-1} + \alpha(\ell_{t-1} b_{t-1} + s_{t-m})\varepsilon_t$	$\ell_t = \ell_{t-1} b_{t-1} + (1 + \alpha \varepsilon_t)$
	$b_t = b_{t-1} + (1 + \beta \varepsilon_t)$	$b_t = b_{t-1} + \beta \frac{(\ell_{t-1} b_{t-1} + s_{t-m})\varepsilon_t}{\ell_{t-1}}$	$b_t = b_{t-1} + (1 + \beta \varepsilon_t)$
M_d	$\mu_t = \ell_{t-1} b_{t-1}^\phi$	$\mu_t = \ell_{t-1} b_{t-1}^\phi + s_{t-m}$	$\mu_t = \ell_{t-1} b_{t-1}^\phi s_{t-m}$
	$\ell_t = \ell_{t-1} b_{t-1}^\phi + (1 + \alpha \varepsilon_t)$	$\ell_t = \ell_{t-1} b_{t-1}^\phi + \alpha(\ell_{t-1} b_{t-1}^\phi + s_{t-m})\varepsilon_t$	$\ell_t = \ell_{t-1} b_{t-1}^\phi + (1 + \alpha \varepsilon_t)$
	$b_t = b_{t-1}^\phi + (1 + \beta \varepsilon_t)$	$b_t = b_{t-1}^\phi + \beta \frac{(\ell_{t-1} b_{t-1}^\phi + s_{t-m})\varepsilon_t}{\ell_{t-1}}$	$b_t = b_{t-1}^\phi + (1 + \beta \varepsilon_t)$
		$s_t = s_{t-m} + \gamma(\ell_{t-1} b_{t-1}^\phi + s_{t-m})\varepsilon_t$	$s_t = s_{t-m}(1 + \gamma \varepsilon_t)$

Cuadro A.3: Fórmulas para todos los modelos ETS con errores multiplicativos.

Bibliografía

- Aneiros, G. (2017/2018). Series de tiempo. Apuntes de la materia.
- Athanasopoulos, G., Ahmed, R. A., and Hyndman, R. J. (2009). Hierarchical forecasts for australian domestic tourism. *International Journal of Forecasting*, 25(1):146–166.
- Box, G. E. and Cox, D. R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society: Series B (Methodological)*, 26(2):211–243.
- Diebold, F. X. and Mariano, R. S. (2002). Comparing predictive accuracy. *Journal of Business & economic statistics*, 20(1):134–144.
- Hyndman, R., Koehler, A. B., Ord, J. K., and Snyder, R. D. (2008). *Forecasting with exponential smoothing: the state space approach*. Springer Science & Business Media.
- Hyndman, R. J. and Athanasopoulos, G. (2018). *Forecasting: principles and practice*. OTexts.
- Hyndman, R. J. and Koehler, A. B. (2006). Another look at measures of forecast accuracy. *International journal of forecasting*, 22(4):679–688.
- Hyndman, R. J., Koehler, A. B., Ord, J. K., and Snyder, R. D. (2005). Prediction intervals for exponential smoothing using two new classes of state space models. *Journal of Forecasting*, 24(1):17–37.
- Hyndman, R. J., Koehler, A. B., Snyder, R. D., and Grose, S. (2002). A state space framework for automatic forecasting using exponential smoothing methods. *International Journal of forecasting*, 18(3):439–454.
- Taylor, J. W. (2003). Exponential smoothing with a damped multiplicative trend. *International journal of Forecasting*, 19(4):715–725.
- Velasco, M. G. and del Puerto García, I. M. (2009). *Series temporales*. Universidad de Extremadura, Servicio de Publicaciones.