Trabajo Fin de Máster

Forecast de ventas

Alejandro Bernárdez Ferradás

Máster en Técnicas Estadísticas

Curso 2022-2023

Propuesta de Trabajo Fin de Máster

Título en galego: Predición de vendas
Título en español: Predicción de ventas
English title: Sales forecast
Modalidad: Modalidad B
Autor/a: Alejandro Bernárdez Ferradás, Universidad de Vigo
Director/a: Jose Ameijeiras Alonso, Universidad de Santiago de Compostela
Tutor/a: Jorge López Muñiz, Hijos de Rivera, S.A.U.
Breve resumen del trabajo: Desarrollo de un modelo de forecasting que estime las ventas futuras de cerveza de la compañía, tanto por canal como a nivel agregado, para el año 2023. Para ello, se considerará el histórico mensual de ventas de cada una de las delegaciones, poniendo especial enfásis en el periodo de pandemia.
Recomendaciones:
Otras observaciones:

Don Jose Ameijeiras Alonso, profesor Axundante Doutor de la Universidad de Santiago de Compostela, y don Jorge López Muñiz, Responsable de Business Analytics de Hijos de Rivera, S.A.U., informan que el Trabajo Fin de Máster titulado

Forecast de ventas

fue realizado bajo su dirección por don Alejandro Bernárdez Ferradás 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 14 de julio de 2023.

El director: El tutor:

Don Jose Ameijeiras Alonso

AMEIJEIRAS ALONSO JOSE - 39450236S
Firmado digitalmente por AMEIJEIRAS ALONSO JOSE - 39450236S
Fecha: 2023.07.14 13:31:45 +02'00'

Don Jorge López Muñiz

El autor:

ALEJANDRO BERNÁRDEZ FERRADÁS - 53861634G
Firmado digitalmente por ALEJANDRO BERNÁRDEZ FERRADÁS - 53861634G
Fecha: 2023.07.14 13:59:58 +02'00'

Don Alejandro Bernárdez Ferradás

Declaración responsable. Para dar cumplimiento a la Ley 3/2022, de 24 de febrero, de convivencia universitaria, referente al plagio en el Trabajo Fin de Máster (Artículo 11, Disposición 2978 del BOE núm. 48 de 2022), **el/la autor/a declara** que el Trabajo Fin de Máster presentado es un documento original en el que se han tenido en cuenta las siguientes consideraciones relativas al uso de material de apoyo desarrollado por otros/as autores/as:

- Todas las fuentes usadas para la elaboración de este trabajo han sido citadas convenientemente (libros, artículos, apuntes de profesorado, páginas web, programas,...)
- Cualquier contenido copiado o traducido textualmente se ha puesto entre comillas, citando su procedencia.
- Se ha hecho constar explícitamente cuando un capítulo, sección, demostración,...sea una adaptación casi literal de alguna fuente existente.

Y, acepta que, si se demostrara lo contrario, se le apliquen las medidas disciplinarias que correspondan.

Agradecimientos

Pese a que no quiero extenderme mucho, en este espacio me gustaría mencionar a las personas que más me ayudaron y contribuyeron a sacar adelante este trabajo.

En primer lugar, quería agradecerle enormemente a la compañía Hijos de Rivera, S.A.U., la oportunidad de introducirme en el mercado laboral y ver de primera mano el funcionamiento de una empresa de tal entidad. En particular, quería destacar a mi tutor Jorge, el cual me brindó desde el principio toda su colaboración para adaptarme rápidamente a la dinámica de la empresa, además de ofrecerme sabios consejos. No obstante, tampoco me puedo olvidar de mis compañeros Adri, Patri y Juanjo, fieles compañeros de café durante las mañanas que, al margen de orientarme magníficamente en el desarrollo de este proyecto, también me enseñaron grandes lecciones sobre temas ajenos al ámbito laboral.

En segundo lugar, pero no menos importante, me gustaría resaltar la importancia de mis allegados, dándome siempre el impulso necesario para afrontar este reto. Soy plenamente consciente de que sin su apoyo no lo hubiera conseguido.

Finalmente, también debo reconocer la importancia de mis compañeros de máster a la hora de hacer mucho más llevadera esta etapa de mi vida.

En definitiva, aunque este sea mi trabajo, el mérito de que haya salido adelante es tanto mío como de todos ellos.

Índice general

Resumen

Resumen en español

Disponer de un modelo de forecasting fiable para la estimación de las futuras ventas de un producto o artículo constituye una de las grandes aspiraciones a nivel empresarial de cualquier compañía. Sin embargo, el estallido de la pandemia de la COVID-19 supuso un abrupto descenso de las ventas durante un lapso temporal prolongado, que, además, afectó duramente a la destreza predictiva de dichos modelos. Ante esta tesitura, se requiere elaborar un nuevo método de pronóstico que cuente con la capacidad de adecuarse al negativo impacto del coronavirus sobre las ventas. Así pues, el propósito fundamental de este trabajo será obtener predicciones precisas acerca del volumen de litros de cerveza vendidos por Hijos de Rivera, tanto por canal como a nivel agregado, para este 2023.

Para ello, inicialmente se recurrirá a los dos modelos clásicos y más habituales en el análisis de series de tiempo: los modelos Box-Jenkins y los modelos de suavizado exponencial. Posteriormente, se apelará a modelos más avanzados, como son los de regresión dinámica y los de series temporales jerárquicas. Sobre cada uno se efectuará una fundamentación teórica, la cual irá acompañada de su posterior aplicación práctica en una serie de datos reales de la empresa. Asimismo, también se llevará a cabo la implementación de una metodología destinada a analizar los cambios de tendencia, predominantemente transitorios, que tuvieron lugar en las series a partir de la COVID-19, con el objetivo de incrementar el rendimiento de los modelos predictivos anteriores.

English abstract

Having a reliable forecasting model for estimating future sales of a product or item is a major aspiration for any company. However, the outbreak of the COVID-19 pandemic led to a sharp decline in sales over an extended period, significantly impacting the predictive accuracy of existing models. In this scenario, there is a need to develop a new forecasting method that can adapt to the negative impact of the coronavirus on sales. The primary objective of this work is to obtain accurate predictions of the volume of beer sold by Hijos de Rivera, both by channel and at an aggregate level, for the year 2023.

Initially, two classical and commonly used time series models, namely the Box-Jenkins models and exponential smoothing models, will be employed. Subsequently, more advanced models such as dynamic regression and hierarchical time series models will be explored. Theoretical foundations will be provided for each model, followed by practical applications using real company data. Additionally, a methodology will be implemented to analyze the predominantly transient trend changes that occurred in the series due to COVID-19, with the aim of improving the performance of the previous predictive models.

Capítulo 1

Introducción

Antes de pasar a presentar y desarrollar los distintos modelos estadísticos empleados en el análisis de series temporales, resulta conveniente realizar una contextualización del campo productivo y la compañía dentro de las cuales se encuadra el estudio materializado a lo largo del presente trabajo. Así pues, inicialmente, se explicarán brevemente las principales características socioeconómicas del sector cervecero, tanto a nivel global como en un entorno más regional. Ulteriormente, y con un proceder similar, se enmarcará la actividad de la empresa Hijos de Rivera, a la vez que se confronta su desempeño frente al de otras firmas cerveceras.

Por otro lado, también se expondrá la cuestión primordial a resolver con este proyecto, que consiste, como bien hemos remarcado anteriormente, en la confección de un forecast sobre el volumen de ventas totales de cerveza en litros de la compañía Estrella Galicia. Para ello, se describirá la metodología seguida para la obtención del modelo de forecasting, es decir, aquel que cuenta con la mejor capacidad predictiva posible.

1.1. Sector cervecero

En primer lugar, debemos comenzar poniendo en liza que el sector cervecero vive un proceso de expansión en la mayor parte de los países, en lo que respecta a la producción y el consumo. Concretamente, en valores numéricos, la producción de cerveza a nivel mundial alcanzó los 190.000 millones de litros durante el año 2022. Al llevar a cabo una descomposión por países en términos de cantidad de cerveza elaborada, nos topamos con que el ranking se encuentra liderado por China, Estados Unidos y Brasil. Por contra, si realizamos una estratificación a nivel europeo, aquí la clasificación es encabezada por Alemania, con alrededor de 45 millones de litros producidos, perseguida a cierta distancia por España y Polonia (Orus, 2022).

En lo tocante al ámbito nacional, España, décimo productor de cerveza en el plano internacional, ha generado un total de 4.100 millones de litros en el año 2022. Asimismo, el sector cervecero ha logrado consolidarse como uno de los referentes en el panorama agroalimentario patrio; pues, a pesar de la importante recesión sufrida durante la pandemia, ha conseguido recuperar prácticamente las cifras pre-coronavirus. Además, también cabe señalar que se trata de una industria muy diversificada en base a los atributos del producto ofertado, constituyendo un sector que aglutina a una amplia gama de empresas. De entre todas ellas, sobresalen el grupo Mahou San Miguel, el grupo Damm y la corporación Hijos de Rivera, puesto que estas tres compañías forman parte del prestigioso "Top 40 Brewers", que engloba a los 40 mayores productores de cerveza del mundo en términos de volumen de producción.

Si ejecutamos una radiografía más específica de la industria, y en particular, sobre las ventas y el consumo, resulta significativo comentar que para el año 2022 se calculó un consumo total de 3.895 millones de litros de cerveza en nuestro país, según aparece recogido en (Cerveceros, 2022). Sin embargo, si tomamos en consideración el consumo per cápita habido en España, este se posiciona en las últimas posiciones del escalafón europeo, con una media de 58 litros por persona. En general, se puede definir a la cerveza como una bebida sumamente transversal entre los distintos grupos de edades, ya que alrededor del 83 % de la población de 18 a 65 años la consumen de manera habitual o esporádica. Diseccionando las ventas por canal, se ha roto la tendencia adquirida durante el periodo de la COVID-19, aunque por un estrecho margen, dado que la vía HORECA (acrónimo de "hoteles, restaurantes y cafeterías") representó una ligera superior proporción que alimentación en 2022.

1.2. Hijos de Rivera

Una vez hemos efectuado un enfoque genérico del sector cervecero, ya es momento de exponer la posición de la empresa Hijos de Rivera dentro del mercado en relación a otras cerveceras, todo ello de un modo breve y conciso.

Inicialmente, tenemos que comenzar recalcando que, si bien la compañía es conocida dentro del argot popular por el nombre de Estrella Galicia, debido a la denominación que adquiere su marca estrella, su verdadera razón social es Hijos de Rivera, S.A.U. Tal corporación fue fundada en el año 1906 por José María Rivera Corral, encargándose desde sus inicios de la producción, comercialización y distribución de cervezas. Pese a sus longevos 117 años de historia, ha logrado mantenerse hasta la actualidad como una entidad familiar y de capital español, sin presencia de inversión extranjera.

El crecimiento de la sociedad se ha dado de forma sostenida a lo largo del tiempo, a excepción del complejo periodo de la Guerra Civil española y su posterior etapa de posguerra en los años 40. Sin embargo, ha sido durante el siglo actual, y más específicamente en esta última década, donde la empresa ha experimentado una gran fase de expansión, la cual todavía prosigue en el instante presente. Estos hechos han llevado a la compañía a convertirse y consolidarse como uno de los grupos cerveceros más relevantes dentro del panorama nacional, contando con actividad comercial en un total de 68 países. Como dato reseñable, mencionar que se trata de la única gran compañía española del sector cervecero que consigue incrementar su facturación anual desde 2010, excluyendo el ejercicio de 2020 (ciclo de estallido de la pandemia del coronavirus).

Este acentuado desarrollo de la corporación ha propiciado la diversificación de la cartera de producción, puesto que al margen de la cerveza, adicionalmente también se dedica a distribuir y comercializar otro tipo de bebidas, tales como el agua mineral, el vermú o el vino. Empero, la cerveza continua constituyendo el producto referente, integrando aproximadamente alrededor de 2/3 de las ventas totales. Asimismo, dentro de la cerveza, Hijos de Rivera posee un amplio número de marcas, las cuales pueden visualizarse en la Figura 1.1. Del conjunto de ellas, las más demandadas son la llamada Estrella Galicia Especial y, en menor medida, la 1906.

Es aquí, dentro de este contexto descrito, donde surge la necesidad para la compañía de disponer de un modelo de forecasting para la previsión de sus ventas futuras en un horizonte temporal de un medio plazo, con el fin de poder anticiparse de manera más óptima a las demandas del mercado y planificar su producción en consecuencia.

Figura 1.1: Nombre de las marcas comerciales de cerveza de Hijos de Rivera S.A.U.
Fuente: Cerveceros (2022)

1.3. Descripción y planteamiento del proyecto

En un principio, hemos de pormenorizar el propósito fundamental de este trabajo y en qué aspectos específicos vamos a centrar el hilo conductor. La consigna esencial planteada en este documento es evidente. Esta consiste en desarrollar un método de forecasting que nos otorgue pronósticos precisos sobre los litros totales de cerveza que serán vendidos por la compañía para el año 2023. Mas, pese a que la meta final sea de carácter agregado, también se pretenden hallar predicciones fiables para los distintos canales de venta de la empresa. De este modo, si desglosamos las ventas de Hijos de Rivera por canal, podemos distinguir un total de seis mercados. No obstante, por motivos de confidencialidad de la compañía, solo estamos habilitados para desvelar dos de estos canales. Estos son:

- **HORECA**: Esta etiqueta engloba las ventas destinadas a empresas y negocios del territorio nacional que brindan servicios de comida y bebida fuera del hogar, tanto en restaurantes y bares, como en servicios de catering para eventos y hospedaje en hoteles.
- **Alimentación**: Bajo esta esfera se recopilan las ventas efectuadas a grandes distribuidoras alimentarias, supermercados y demás establecimientos mayoristas del país.

Como resulta obvio, no todos los canales tienen la misma importancia para la empresa. Dentro de ellos, el canal HORECA destaca como el de mayor relevancia, debido a que integra un porcentaje superior de las ventas de cerveza. Con un menor peso se posiciona el canal Alimentación, el cual acapara la otra gran proporción de ventas de la compañía. En contraposición, la huella sobre las ventas totales de los demás canales es secundaria y residual.

Si desmenuzamos, todavía más, el historial de ventas de cerveza en el espacio de cada canal, podemos ejecutar una segmentación por delegación de ventas. Cada una de estas delegaciones se sitúan, mayoritariamente, desperdigadas alrededor de todo el ámbito estatal español, aunque también se ubican unas pocas en países externos. En total se disponen de 46 delegaciones de ventas, repartidas de forma inequitativa entre los diferentes canales.

Delimitada la jerarquía de ventas en la que nos basaremos para nuestros análisis, ya podemos pasar a enumerar cómo se ha llevado a cabo el proceso de obtención de los datos.

Inicialmente, se recolectaron los datos relativos a nuestra propia cervecera. Dichos datos se encontraban almacenados en la plataforma SQL Server de la compañía. Todos los datos de interés fueron descargados en formato CSV, con la finalidad de procesarlos posteriormente en R (R Core Team (2023)). Cabe subrayar que este software fue el único entorno de programación utilizado en el desarrollo de este proyecto, siendo aplicados un vasto abanico de los paquetes que incorpora a lo largo del presente trabajo.

Tal y como fue indicado previamente, se reunió el histórico mensual de ventas de cerveza en litros considerando las diferentes delegaciones y canales existentes. No se realizó ninguna distinción conforme a la marca de cerveza o el material de ensamblaje en cuestión, con la intención de no descomponer en exceso los datos. Además, absolutamente todas las observaciones referidas a cada delegación comprenden el periodo que va desde enero de 2010 hasta octubre de 2022. Como consecuencia de que algunas delegaciones no contaban con registros de ventas para todos los meses de la serie (especialmente en los primeros años), se tomó la decisión de rellenar esas series con "1s" para que todas tuvieran la misma longitud y estuvieran definidas en el mismo espacio de tiempo[1].

Sin embargo, tenemos que precisar que durante este intervalo temporal ocurrió la pandemia de la COVID-19, la cual repercutió drásticamente en las ventas de la compañía acontecidas entre marzo de 2020 y mediados del 2021. Aunque su repercusión varió según la delegación de ventas de la que se tratase, en general se vivió un notable descenso en las ventas de cerveza, a excepción de determinadas delegaciones donde las ventas aumentaron en lugar de disminuir. Por tanto, independientemente de su efecto, estamos ante una series temporales fuertemente afectadas y marcadas por el factor covid.

Adicionalmente, se recurrieron a fuentes externas para añadir posibles variables correlacionadas con la venta de cerveza. Tras un profundo diagnóstico del sector, se tomaron otras cinco variables, quedando para cada delegación una base de datos con los registros mensuales de 6 variables y un total de 156 observaciones para cada variable. Dichas variables son:

- **Ventas de cerveza**: Ya explicada con anterioridad, computa las ventas de cerveza en litros para cada delegación.
- **Temperatura media**: Refleja el promedio mensual de temperaturas en ^{o}C que se experimentaron en la ciudad más representativa de cada delegación[2].
- **Precipitaciones acumuladas**: Expresa las precipitaciones acumuladas en mm en la ciudad más representativa de cada delegación.
- **Días de lluvia**: Plasma el número de días que llovió (se exige un umbral mínimo de 1mm de agua caída para que se considere día de lluvia) en la ciudad más representativa de cada delegación.
- **Pernoctaciones hoteleras**: Hace referencia al número de pernoctaciones habidas en los establecimientos hoteleros de la provincia más representativa de cada delegación (si es que contiene a más de una).
- **Días laborables**: Agrupa al conjunto de días laborables habidos en la comunidad autónoma más representativa de cada delegación (rara vez cubrirá más de una).

Para la extracción de las variables climatológicas se acudió a la página web de la AEMET (Agencia Estatal de Meteorología), en la cual fueron descargadas año a año (desde 2010 hasta 2022) los registros meteorológicos mensuales correspondientes a cada ciudad[3]. Todos estos archivos generados estaban en

[1] La razón elemental de haber rellenado las series con "1s" en vez de "0s" es que así se evitan posibles complicaciones futuras a la hora de realizar transformaciones sobre las series. Asimismo, remarcar que, dada la cuantía real del rango de valores en el que nos movemos, el efecto de completar con "1s" en lugar de "0s" resulta insignificante.

[2] Con "ciudad más representativa de cada delegación" aludimos a la urbe que concentra a un mayor número de población y que, por consiguiente, resulta previsible que abarque una mayor porción de las ventas dentro de cada delegación. De la misma manera, dicho modo de proceder será extrapolable a los estratos de provincia o comunidad autónoma.

[3] Para más detalles sobre el proceso de extracción de datos climatológicos visítese: https://www.aemet.es/es/datos_abiertos/AEMET_OpenData

formato JSON, por lo que tuvimos que valernos de la función fromJSON() del paquete jsonlite en R para poder hacer uso de las observaciones recabadas (Oons *et al.* (2023)). Finalmente, se apartaron las tres variables de interés para nuestro estudio.

En contraste, para conseguir información acerca de la variable "pernoctaciones hoteleras" nos servimos de la plataforma web del INE (Instituto Nacional de Estadística), donde también año a año, se bajaron los datos de la variable especificada en formato CSV. Ulteriormente, todos los documentos fueron transcritos a R por medio de la función read.csv2() del paquete R.utils (Bengtsson (2022)).

Finalmente, para la variable "días laborables" nos apoyamos en los datos compilados por el Ministerio de Asuntos Económicos y Transformación Digital en su portal en línea. Al igual que antes, se descargaron los datos año a año para cada ciudad en formato CSV y seguidamente, se importaron en R empleando un *modus operandi* análogo.

Capítulo 2

Series de tiempo

Como bien se señala en Cryer y Kung-Sik (2008), dentro del ámbito empresarial, los datos obtenidos a partir de las observaciones recopiladas durante un determinado periodo temporal son muy recurrentes. Algunos casos concretos podrían ser los precios de cierre diarios de las acciones, el nivel de producción mensual o el registro de los flujos de caja anuales. Por ende, y como consecuencia de lo anterior, surge la necesidad de llevar a cabo un exhaustivo análisis de las series temporales a considerar en cada caso.

Así pues, a lo largo de todo este capítulo, se hará una introducción a las series de tiempo. Para ello, se describirán los conceptos básicos que se deben tomar en cuenta en el modelaje de las mismas y las transformaciones existentes para conseguir su estacionariedad. Del mismo modo, también se presentarán los distintos criterios utilizados para analizar la calidad de los modelos, así como las diversas medidas de adecuación de las predicciones empleadas.

2.1. Conceptos básicos

Primeramente, debemos comenzar precisando que una **serie de tiempo** (o simplemente serie) es una colección de N observaciones de una variable X agrupadas de forma cronológica en sucesivos momentos temporales t, generalmente equispaciados (Brockwell y Davis (2002)). El intervalo regular de tiempo donde ha sido capturada cada una de las observaciones, dependiendo de la situación específica ante la que nos encontremos, puede ser de una hora, una semana, un año... Para examinar su evolución se construye el denominado gráfico secuencial, en el cual se representa cada observación, x_t, frente al instante, t, en que se observa. Un ejemplo de gráfico secuencial es el que se muestra a continuación en la Figura 2.1:

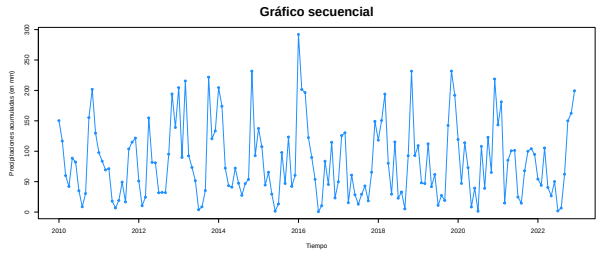

Figura 2.1: Precipitaciones acumuladas durante cada mes en la ciudad de Coruña en el periodo 2010-2022.

A la hora de evaluar y caracterizar la evolución de una serie temporal, es necesario identificar si se presentan los tres componentes fundamentales que atañen a las mismas. Estos son:

- **Tendencia**: Plasma el patrón de comportamiento del nivel de la serie (creciente o decreciente) con el transcurrir del tiempo. Si los cambios en la variable respuesta son constantes en el tiempo, entonces la tendencia es lineal. Por el contrario, cuando los cambios de la variable suceden de un modo no constante, de manera que la amplitud de la variación en la serie temporal aumenta a medida que también se incrementa el tiempo, la tendencia es curvilínea y no lineal. En concreto, para este último caso, la tendencia puede seguir una forma polinómica, logarítmica, exponencial o potencial, entre otras. Asimismo, destacar que, por lo general, en el mundo real, predominan las series con patrones de mutación inconstantes a lo largo del tiempo.
- **Estacionalidad**: También conocida como variación estacional, hace referencia a las fluctuaciones sistemáticas y repetidas que, por intervalos regulares de tiempo (días, semanas, meses,...), se producen en el nivel medio de una serie temporal. En la práctica, el mayor (o menor) impacto de dicho efecto sobre la serie suele variar entre los distintos periodos temporales, debido a la influencia de otros factores externos y modificaciones que puedan haber tenido lugar.
- **Heterocedasticidad**: Se manifiesta cuando la variabilidad de la serie se altera a lo largo del tiempo. En otras palabras, esta propiedad implica que la varianza no se mantiene constante en todo el periodo temporal de estudio, puesto que depende del nivel de la serie en cuestión. Bien es cierto que, mayoritariamente, en la realidad, la existencia de heterocedasticidad en la serie se encuentra muy intrínsecamente ligada a la también presencia de tendencia; mas, sin embargo, conviene resaltar que esta, en absoluto, es una condición *sine qua non*.

Los conceptos explicados previamente aparecen ilustrados gráficamente a través de las diferentes series temporales (todas ellas construidas a partir de datos empíricos) que se recogen en la Figura 2.2.

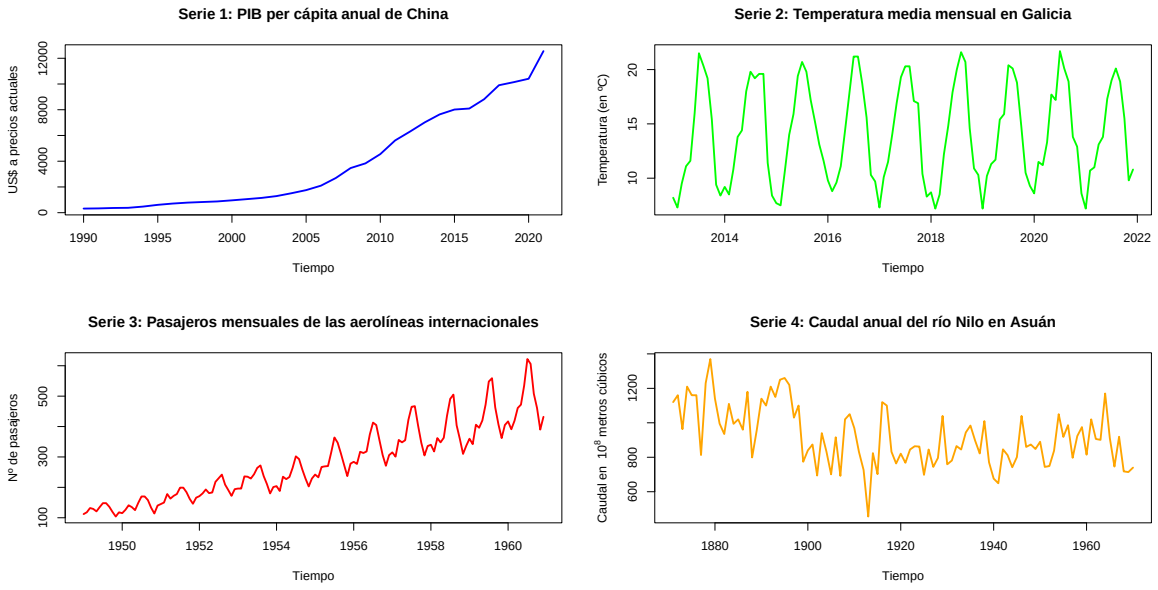

Figura 2.2: Cuatro ejemplos de series temporales. La serie 1 se caracteriza por tener una tendencia creciente y predominantemente no lineal. En la serie 2 se visualiza una fuerte componente estacional de carácter anual. En la serie 3 se observan las dos propiedades anteriores, es decir, tanto tendencia como estacionalidad. Finalmente, la serie 4 exhibe una notoria heterocedasticidad, pues su variabilidad cambia con el paso del tiempo.

Por otro lado, siguiendo lo recogido por Peña (2005), formalmente se define como **proceso estocástico** al conjunto de variables aleatorias, $\{X_t\}_{t\in C}$ (sea C el conjunto de valores posibles que puede tomar en el instante t), donde cada variable aleatoria refleja la observación en un momento específico del tiempo. Es decir, se trata de una familia de distribuciones de probabilidad, una para cada instante de tiempo, que describen el comportamiento de un fenómeno a lo largo del tiempo. De esta forma, a partir de una observación de un proceso estocástico, se establece que una serie temporal, $x_1, x_2, ..., x_t$, viene a constituir una realización o trayectoria parcial finita de este.

En el análisis de series temporales el proceso básico a seguir consistirá, por tanto, en estimar el modelo estocástico generador del conjunto de datos. Posteriormente, y basándonos en el proceso obtenido en cuestión, efectuaremos predicciones a futuro con respecto a la serie original. No obstante, hay que subrayar que en la praxis, salvo que trabajemos con datos ficticios, la serie nunca habrá sido construida tomando como base un determinado proceso estocástico. Por consiguiente, lo que realmente se pretende hallar es el modelo estocástico que se ajuste lo más óptimamente a la tipología de nuestra serie.

En consecuencia, la metodología a emplear en el estudio de series de tiempo, la cual aparece detallada en Box *et al.* (1994), se resume en las siguientes directrices:

1) Representación de la serie e identificación de la posible presencia de algunas de las componentes fundamentales. Frecuentemente, necesidad de transformar la serie para la verificación de la condición de estacionariedad.
2) Selección del modelo más adecuado para el conjunto de datos.
3) Ajuste del modelo elegido.
4) Diagnosis o análisis de los residuos.
5) Predicción de los valores futuros de la serie original.

Mencionado todo esto, procederemos a ir pormenorizando con más profundidad algunos de los pasos citados en los próximos apartados.

2.2. Estacionariedad y transformaciones para conseguirla

Para poder aplicar métodos de inferencia estadística sobre el registro de observaciones de un determinado proceso estocástico, habitualmente se hace necesario establecer una serie de suposiciones simplificadoras (y medianamente razonables) acerca de dicha estructura. Dentro del conjunto de todas ellas, la más elemental, indiscutiblemente, es la condición de estacionariedad.

2.2.1. Series estacionarias

En concordancia con lo expuesto en el párrafo previo, afirmaremos que estamos ante un **proceso estacionario** si se cumplen dos requisitos imprescindibles. El primero de ellos es que su media μ y su varianza σ^2 tienen que mantenerse constantes con el transcurrir del tiempo. Paralelamente, el segundo hace alusión a que el valor de la covarianza γ entre dos periodos únicamente debe depender de la distancia o rezago entre estos dos ciclos temporales. En lenguaje matemático, y dado un proceso estocástico $\{X_t\}_t$, los anteriores condicionantes quedarían plasmados del siguiente modo:

- Media: $\mathbb{E}(X_t) = \mu_t = \mu, \forall t.$
- Varianza: $\mathbb{V}ar(X_t) = \sigma_t^2 = \sigma^2, \forall t.$
- Covarianza: $Cov(X_t, X_{t+k}) = \gamma(t, t+k) = \gamma_k, \forall t, k.$

En definitiva, una serie verifica la propiedad de estacionariedad si y solo si tanto la media, la varianza y la autocovarianza (en diferentes rezagos) permanecen invariantes frente al tiempo. Por ende, constituyen un tipo de series que no exhiben ninguna de las componentes fundamentales; en otras palabras, que carecen de tendencia, estacionalidad y heterocedasticidad.

Tal y como se especifica en Cowpertwait y Metcalfe (2009), el proceso estocástico estacionario más reconocido es el denominado **ruido blanco**. De cualquier manera, su importante relevancia no radica en el hecho de que sea un modelo interesante en si mismo, sino en que se conforma como el precursor de otros muchos procesos útiles de estudio. En el contexto de un proceso ruido blanco, sea $\{a_t\}_t$ un conjunto de variables aleatorias, tendremos que la totalidad de las observaciones comparten que son independientes y se encuentran idénticamente distribuidas (i.i.d.) con media cero y varianza constante. Si expresamos esto último en notación matemática, tenemos que:

- $\mathbb{E}(a_t) = \mu = 0$
- $\mathbb{V}ar(a_t) = \sigma^2 = \sigma_a^2$
- $Cov(a_t, a_{t+k}) = \gamma_k = \begin{cases} \sigma_a^2, & si\ k = 0 \\ 0, & si\ k \neq 0 \end{cases}$

Si adicionalmente añadimos que las variables también siguen una distribución normal (i.e., $\{a_t\}_t \sim N(0, \sigma^2)$), entonces el proceso es llamado ruido blanco gaussiano. En la Figura 2.3 puede visualizarse un ejemplo de serie generada por un ruido blanco gaussiano:

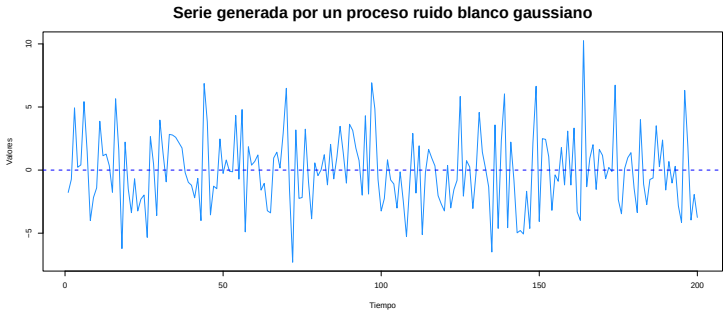

Figura 2.3: Serie simulada a partir de un proceso ruido blanco gaussiano con $\sigma_a^2 = 10$.

Con todo, conviene enfatizar que el gráfico secuencial precedente ha sido elaborado con datos simulados que, por tanto, no corresponden a mediciones reales. La razón de haber tomado observaciones simuladas se justifica por el motivo de que, en la práctica, es prácticamente imposible hallar una serie completamente estacionaria, puesto que acostumbran a mostrar tendencia y/o patrones repetitivos.

Hasta el momento hemos explicado cómo examinar la estacionariedad por medio del análisis gráfico. Empero, tal vía, aunque aporte conclusiones clarividentes en algunas situaciones, puede dificultar la extracción de resultados esclarecedores en otros escenarios. Es por ello, por lo que se requiere contar con un enfoque distinto, como es el caso del contraste de **Dickey-Fuller**. Este test sirve para calcular si una serie temporal presenta raíces unitarias, lo cual, a su vez, permite averiguar si esta cumple la hipótesis nula de estacionariedad. Puede consultarse más información sobre la prueba en Phillips y Perron (1988).

Con el fin de evaluar analíticamente la condición de estacionariedad sobre un conjunto de datos empíricos, recurrimos a la serie ilustrada previamente en la Figura 2.1. Para dicho propósito, nos valemos de la función `adf.test()`, perteneciente al paquete `tseries` (Trapletti *et al.* (2023)) del software R. La salida arrojada se plasma en la Figura 2.4:

```
        Augmented Dickey-Fuller Test

data:  Precipitaciones_Coruña
Dickey-Fuller = -7.3134, Lag order = 5, p-value = 0.01
alternative hypothesis: stationary
```

Figura 2.4: Prueba de Dickey-Fuller para contrastar la estacionariedad de una serie.

A la vista de la salida de código, para un nivel de significación del 5 % se concluye que se rechaza la H_0, de forma que la serie no es estacionaria.

Como habíamos remarcado anteriormente y, además, hemos logrado atestiguar con este ejemplo particular, la amplia mayoría de series con observaciones reales incumplen el supuesto de estacionariedad. Dado que la estacionariedad es una exigencia imprescindible para el correcto ajuste de algunos de los modelos más elementales de series temporales, se vuelve indispensable realizar alguna transformación sobre la propia serie original con el fin de volverla estacionaria. A grandes rasgos, se distinguen dos categorías de transformaciones: por un lado, las transformaciones de estabilización del nivel, y, por el otro, las transformaciones de estabilización de la varianza.

Sin embargo, antes de adentrarnos a comentar ambas modalidades de transformación, se precisa aclarar el significado de los conceptos de función de autocorrelaciones simples (`fas`) y función de autocorrelaciones parciales (`fap`), ya que poseerán una considerable utilidad en análisis posteriores.

La primera de ellas es una medida empleada para valorar el grado de dependencia lineal entre los valores de la serie en un momento establecido y en los instantes anteriores (retardos). Puede adoptar valores entre $[-1, 1]$ y su representación gráfica es de gran provecho a la hora de mensurar la tendencia y detectar patrones estacionales. Más formalmente, en términos matemáticos, la `fas` se expresaría así:

$$\rho(t, t+k) = Cor(X_t, X_{t+k}) = \frac{Cov(X_t, X_{t+k})}{\sigma_t \sigma_{t+k}} = \frac{\gamma(t, t+k)}{\sqrt{\gamma(t,t)\gamma(t+k, t+k)}}$$

Mientras que, en contraposición, la `fap`, a pesar de que determina lo mismo que la `fas`, se diferencia de ella en el aspecto de que sí elimina la influencia de las autocorrelaciones en los retardos intermedios. Esta circunstancia posibilita identificar la correlación directa entre el valor de un determinado instante y los valores en retardos específicos, sin el potencial impacto de los valores en otros retrasos. Al igual que la `fas`, toma valores entre $[-1, 1]$. Desde un punto de vista matemático, se definiría de la siguiente manera:

$$\alpha(t, t+k) = \frac{Cov\left(X_t - \widehat{X}_t^{(t,t+k)}, X_{t+k} - \widehat{X}_{t+k}^{(t,t+k)}\right)}{\sqrt{\mathbb{V}ar\left(X_t - \widehat{X}_t^{(t,t+k)}\right)\mathbb{V}ar\left(X_{t+k} - \widehat{X}_{t+k}^{(t,t+k)}\right)}}$$

donde $\widehat{X}_j^{(t,t+k)}$ denota al mejor predictor lineal de X_j construido a partir de las variables medidas en los instantes comprendidos entre t y $t+k$.

Expuestos ambos términos, ahora ya resulta oportuno pasar a desglosar los diversos tipos de transformaciones habidas para obtener series estacionarias.

2.2.2. Transformaciones de estabilización de la varianza

Las transformaciones de estabilización de la varianza, o también conocidas como transformaciones funcionales, se utilizan cuando la varianza de la serie cambia a lo largo del tiempo con frecuencia, es decir, en casos de series con heterocedasticidad. Así pues, se tratan de herramientas muy eficaces para aplicar sobre series cuya varianza varía fuertemente en función del nivel o la tendencia. Si bien existen diferentes transformaciones de esta índole útiles para algunos escenarios específicos, tales como la transformación de raíces cuadradas, la más generalizada y válida para múltiples contextos es la **transformación de Box-Cox**.

Matemáticamente, se establece que las transformaciones de Box-Cox de un proceso estocástico cualesquiera, $\{X_t\}_{t\in C}$, constituyen aquellas que transmutan, para cada $t = \{1, ..., T\}$, la observación x_t en

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

donde el parámetro λ se suele escoger mediante la minimización de la suma de los cuadrados de los errores (método de Guerrero) o el cálculo del modelo lineal más verosímil ajustado a la serie (método de máxima verosimilitud). Por consiguiente, estas transformaciones paramétricas se centran en hallar la potencia óptima que mejor estabilice la varianza de la serie, todo con el objetivo de mejorar la normalidad de los datos. Cabe resaltar que únicamente puede ejecutarse sobre valores positivos. Debido a esto, para casos de series que incumplan la mencionada premisa, se requerirá agregar una misma constante a todos los valores para convertirlos en positivos.

Particularmente, cuando se disponen de datos positivos con una relación multiplicativa entre el nivel y la varianza de la serie ($\sigma_t = k\mu_t$), la variabilidad se estabiliza con un $\lambda = 0$, o dicho de otro modo, usando la transformación logarítmica. Esta última quizás se trate de la más recurrentemente empleada en la práctica, especialmente, dentro del ámbito financiero.

A continuación, se recoge en la Figura 2.5 un ejemplo de transformación de Box-Cox para el conjunto de datos AirPassengers, el cual se encuentra disponible en R dentro del paquete datasets. Para llevar a cabo la transformación nos hemos valido, respectivamente, de las funciones BoxCox.lambda() y BoxCox() del paquete forecast (Hyndman y Athanasopoulos (2023)). La técnica utilizada para la elección de λ fue el método de Guerrero. Finalmente, subrayar el éxito de la intervención, puesto que la serie pasa de contar con una variabilidad cambiante a lo largo del tiempo (gráfico de arriba) a volverse homocedástica (gráfico de abajo). Para más detalles se recomienda revisar Box y Cox (1964).

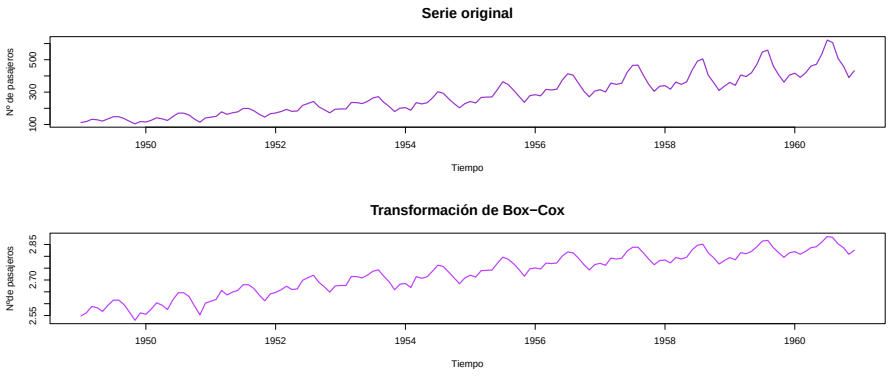

Figura 2.5: Transformación de Box-Cox con un parámetro $\hat{\lambda} = -0{,}29$.

2.2.3. Transformaciones de estabilización del nivel

Una buena parte de las series temporales reales presentan tendencia o componente estacional (o ambas inclusive). La metodología más habitual para anular su efecto es la diferenciación. Si, a su vez, efectuamos una subdivisión en base a los tipos de diferenciación habidas, tenemos dos clases de diferenciación: la diferenciación regular y la diferenciación estacional.

En lo tocante a la **diferenciación regular**, se determina el operador de diferencia regular de orden `d` a través de

$$\nabla^d = (1 - B)^d$$

donde B denota el operador retardo definido por $B^d x_t = x_{t-d}$.

De esta forma, en función del orden `d` de diferenciación apropiado para cada serie, se aplican `d` diferencias entre observaciones sucesivas. Mas, regularmente, con la inclusión de una diferenciación regular de primer orden (comúnmente denominada diferenciación simple) basta para suprimir la tendencia de la serie.

Gráficamente, la valoración de si necesita ser diferenciada la serie no resulta aconsejable realizarla solamente a la vista del gráfico secuencial, ya que estudiar la `fas` también es pertinente. Si atendemos a la `fas`, en líneas generales, se precisará de diferenciación regular cuando dispongamos de autocorrelaciones positivas con un decrecimiento lento y lineal. Este fenómeno puede visualizarse en la Figura 2.6, donde se muestra el gráfico secuencial y la fas de la serie transformada por Box-Cox del conjunto AirPassengers (parte superior) frente a la serie resultante tras llevar a cabo la diferenciación simple (parte inferior). Claramente se percibe que la tendencia creciente de la serie ha sido eliminada con la diferenciación, pues la serie deja de tener un patrón de comportamiento creciente para adquirir uno constante con el transcurrir del tiempo.

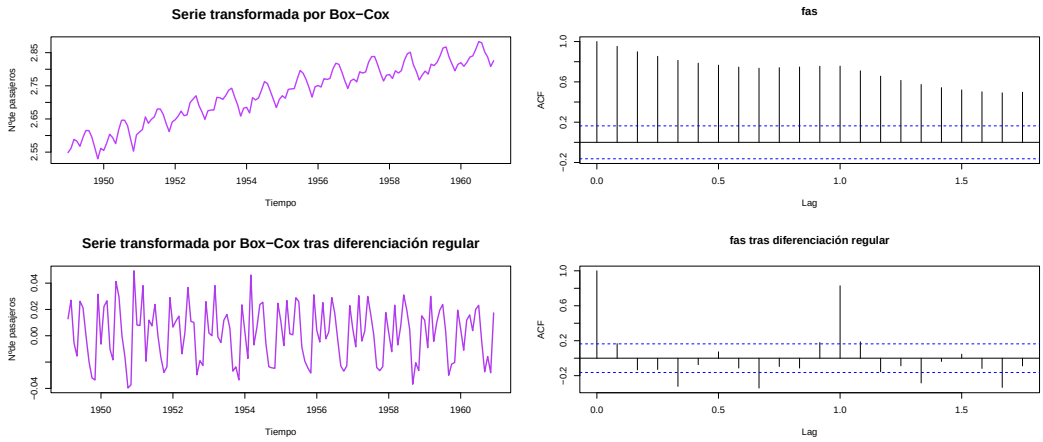

Figura 2.6: Comparación del gráfico secuencial y la fas de la serie transformada por Box-Cox tras aplicar diferenciación regular de orden `d=1`.

En lo que respecta a la **diferenciación estacional**, el operador de diferencia estacional de orden D y periodo `s` viene fijado por

$$\nabla_s^D = (1 - B^s)^D$$

donde B denota el operador de retardo definido por $B^s x_t = (1 - B^s)\, x_t$ para $D = 1$.

Su mecanismo de funcionamiento es prácticamente idéntico al de la diferenciación regular, salvo en el hecho de que el cálculo de las diferencias se realiza considerando las observaciones separadas

por el mismo intervalo de tiempo estacional. Por tanto, corresponde recurrir a ella cuando se detectan patrones o fluctuaciones regulares repetidas en intervalos fijos.

Retomando a nuestro conjunto anterior de AirPassengers, se puede contemplar en la `fas` de la serie transformada y diferenciada regularmente (Figura 2.6) un pico en el valor de la autocorrelación del retardo número 12. Esto es síntoma de presencia de estacionalidad en la serie, por lo que surge el menester de ejecutar una diferenciación estacional con periodo `s=12`. Tras diferenciarla estacionalmente, se obtiene lo plasmado en la Figura 2.7:

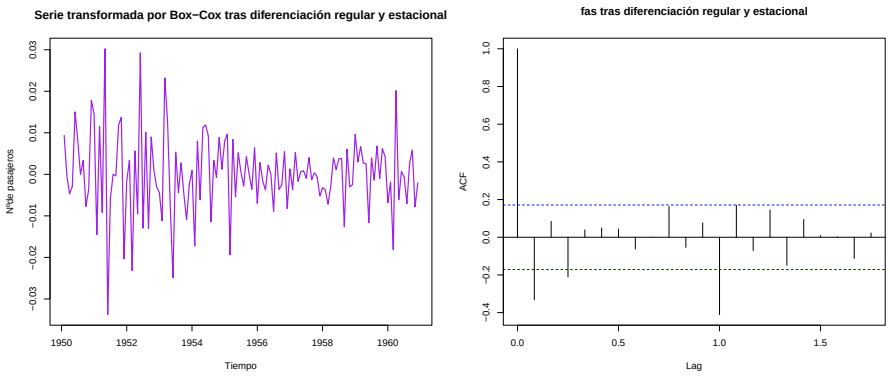

Figura 2.7: Gráfico secuencial y fas de la serie transformada por Box-Cox tras aplicar diferenciación regular (`d=1`) y diferenciación estacional (`D=1`) con `s=12`.

Finalmente, destacar que, tras la implementación de diversas transformaciones sobre la serie original, se logra una serie transformada más apta para ser ajustada por cualquier clase de modelo. Asimismo, también debemos recalcar que, para coyunturas donde sea imprescindible estabilizar tanto la varianza como la media, se hará primeramente la transformación de estabilización de la varianza. Por el contrario, el orden en la aplicación de la diferenciación regular y diferenciación estacional, o viceversa, resulta indiferente.

2.3. Criterios de selección de modelos

A lo largo de los próximos capítulos caracterizaremos y detallaremos algunos modelos destinados al ajuste de series temporales. Ante esta tesitura, dentro de cada familia de modelos, aparecerá la disyuntiva de tener que escoger cuál es el idóneo. Dicha elección es evidente que ha de sustentarse en fundamentos estadísticos, es decir, en determinados criterios de selección de modelos. Por último, señalar que toda la explicación de este apartado está cimentada en las publicaciones de Burnham y Anderson (2004) y Konishi y Kitagawa (2008).

Aunque, en una primera instancia, se puede pensar que el coeficiente de determinación R^2 se trata de un buen indicador de la bondad de ajuste de los modelos de series temporales; no lo es por varias razones:

a) La autocorrelación temporal de los datos viola el principio de independencia de las observaciones, cuyo cumplimiento se requiere para el correcto cálculo del R^2. Como consecuencia de esto, en un gran número de casos puede darse una sobreestimación del grado de bondad de ajuste del modelo[1]. De ahí que, en el contexto de series temporales, el R^2 nos señale erróneamente un elevado grado de ajuste de los modelos a los datos, cuando muchos en realidad no lo tienen.

[1]El sobreajuste (o *overfitting*) hace referencia a una situación en la que, en teoría el modelo se ajusta adecuadamente bien al conjunto de datos, pero, donde al mismo tiempo, mantiene un desempeño muy deficiente en la predicción de

b) El coeficiente R^2 asume varianza constante en toda la serie, de modo que no toma en cuenta la posible existencia de heterocedasticidad en los datos. Dado que la mayoría de las series temporales de datos reales no satisfacen la condición de homocedasticidad, el R^2 probablemente efectuará una evaluación incorrecta del grado de ajuste del modelo.

c) La posible influencia de las componentes de tendencia y estacionalidad no son apreciadas por el coeficiente R^2, centrándose únicamente en la variabilidad explicada por el modelo, sin considerar los patrones de la serie temporal. Este hecho, por tanto, repercute negativamente en la calidad de ajuste del R^2 a los datos.

Como hemos podido comprobar, al trabajar con datos temporales hay que optar por otros criterios de selección más específicos, propios del análisis de series de tiempo. Concretamente, nos concentraremos en tres criterios que veremos seguidamente. Sin embargo, cabe resaltar que, independientemente de cómo hayan sido construidos, todos ellos comparten que cuanto menor sea su valor, mejor será el ajuste de la serie. Asimismo, los tres penalizan el aumento de la complejidad del modelo, es decir, la inclusión de más parámetros o variables (sobreajuste).

- **Criterio de Información de Akaike (AIC):**

Sea la suma de los residuos cuadráticos,

$$\texttt{SSE} = \sum_{t=i}^{T} \epsilon_t^2$$

el criterio de información de Akaike se define como

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

donde k refleja la cantidad de parámetros estimados en el modelo (excluyendo los errores residuales) y T el número de observaciones recogidas en la serie. A grandes rasgos, podemos decir que constituye el criterio más frecuentemente usado en la práctica, puesto que equilibra adecuadamente el buen ajuste de los datos con una complejidad moderada del modelo.

- **Criterio de Información de Akaike corregido (AICc):**

Acostumbra a emplearse en coyunturas donde el tamaño de T es reducido con el fin de corregir la propensión del `AIC` a elegir modelos con un amplio número de parámetros estimados. Así pues, penaliza en mayor medida el *overfitting* cuando T es limitado. La mencionada corrección del sesgo da lugar a la siguiente expresión:

$$\texttt{AICc} = \texttt{AIC} + \frac{2k^2 + 2k}{T - k - 1}$$

- **Criterio de Información Bayesiano (BIC):**

También conocido criterio de Schwarz, a diferencia del `AIC`, sí penaliza intensamente el sobreajuste aún cuando T es elevado. En notación matemática quedaría expresado así:

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

nuevos valores. Normalmente esta circunstancia suele ocurrir cuando se ajusta un modelo desmesuradamente complejo, que incorporar un número excesivo de parámetros o variables. Por ende, siempre interesa encontrar un equilibrio entre la complejidad del modelo y su facultad para capturar la estructura y patrones de los datos, evitando tanto modelos demasiado simples como demasiado complejos

Expuestos los criterios de selección, tenemos que poner de manifiesto que tanto el criterio `AIC` como el `BIC` tienden a devolver exactamente el mismo modelo. Sin embargo, aún así debemos exponer que el criterio BIC es más consistente, ya que penaliza más fuertemente la complejidad del modelo, ayudando, de esta manera, a evitar el sobreajuste. En contraste, y fruto de lo anterior, el criterio `BIC` no es asintóticamente eficiente, es decir, no escoge en base a la minimización del error de predicción esperado (al contrario de los criterios `AIC` y `AICc`).

2.4. Medidas de adecuación de las predicciones

De forma análoga a como hicimos en la sección previa, procederemos a desglosar minuciosamente las distintas medidas de adecuación habidas para evaluar la capacidad predictiva de los modelos a estudiar en las venideras secciones. El objetivo fundamental con la utilización de las mismas será establecer comparaciones entre los modelos según su rendimiento, con el propósitos de quedarnos con el que nos ofrezca predicciones más fiables a futuro (finalidad de cualquier forecasting). Mas, es vital tener presente que estas medidas se fundamentan en confrontar los valores predichos con los valores reales observados. Esto implica, a su vez, que las medidas solo pueden evaluar la calidad de las predicciones conforme a los datos disponibles del pasado. Por consiguiente, no proporcionan una garantía de precisión futura y no contemplan eventuales modificaciones en el comportamiento de la serie. Igualmente, también pueden ser bastante sensibles a valores atípicos o extremos (*outliers*). Remarcados todos estos condicionantes, pasamos a explicar una por una las diferentes medidas.

- **Error medio (ME):**

A partir de $\hat{x}_T(h)$, una predicción con origen en T y horizonte h, y a partir de x_{T+h}, el valor real de lo que se pretendía pronosticar, se establece el error de predicción:

$$e_T(h) = x_{T+h} - \hat{x}_T(h)$$

Precisado esto, puede definirse el `ME` como

$$\texttt{ME} = \frac{1}{H}\sum_{h=1}^{H} e_T(h)$$

donde H representa el tamaño del horizonte de predicción, o en otras palabras, el número total de predicciones halladas.

La principal ventaja que posee este coeficiente con respecto a otros es que sí nos advierte de si las predicciones se orientan hacia la sobreestimación (`ME` positivo), o bien, hacia la subestimación (`ME` negativo). La gran desventaja en su uso es la dependencia que tiene de las unidades de los datos en cuestión, careciendo de capacidad para confrontar métodos de predicción aplicados sobre series calibradas con diferentes unidades (o en distintas magnitudes).

- **Error medio absoluto (MAE):**

Al igual que el coeficiente anterior, mide el tamaño promedio del error en la predicción, con la exclusiva diferencia de que ahora se toman las diferencias absolutas entre las predicciones y los valores reales. Matemáticamente se expresaría así:

$$\texttt{MAE} = \frac{1}{H}\sum_{h=1}^{H} |e_T(h)|$$

Considerando que siempre adopta valores positivos, nos informa solamente del error de pronóstico cometido, sin entrar a valorar la tendencia del sesgo de las predicciones. Del mismo modo que el `ME`, también está vinculado a las unidades de medida sobre las que se construye la serie.

- **Raíz del error cuadrático medio (RMSE):**

Construido de un modo similar a los coeficientes anteriores, calcula la raíz cuadrada de la distancia promedio al cuadrado habida entre cada una de las predicciones y sus respectivos valores reales. Su expresión matemática es la subsiguiente:

$$\texttt{RMSE} = \sqrt{\frac{1}{H}\sum_{h=1}^{H} e_T^2(h)}$$

A la vista de la fórmula, se extrae que el `RMSE`, en comparación al `MAE`, le otorga un mayor peso a los $e_T(h)$ más altos, al encontrarse elevados al cuadrado. Es por ello, por lo que irremediablemente el `RMSE` será idéntico o superior al `MAE`, pues a mayor varianza de cada uno de los errores, más disparidad entre ambos criterios.

- **Media de los errores porcentuales absolutos (MAPE):**

Si el error porcentual de predicción es definido como

$$p_T(h) = \frac{100 e_T(h)}{x_{T+h}}$$

entonces se puede determinar que

$$\texttt{MAPE} = \frac{1}{H}\sum_{h=1}^{H} |p_T(h)|$$

Su atributo más significativo es su condición de adimensionalidad, es decir, que no depende de las unidades de medida. Esta circunstancia le confiere legitimidad para confrontar la adecuación de diversos procedimientos de predicción cuantificados en distintas unidades. No obstante, su importante inconveniente se refleja en el deficiente desempeño al usarlo en series con valores próximos o iguales a 0 en el periodo de interés, comportándose con notable inestabilidad. En última instancia, recalcar que cuanto más inferior sea el valor del `MAPE`, mayor precisión en el pronóstico tendremos.

- **Media de los errores escalados absolutos (MASE):**

Primeramente, se fija el error escalado, $q_T(h)$, el cual se construye ajustando $e_T(h)$ por medio del `MAE` conseguido tras ejecutar una predicción *naïve*[2] sobre la muestra histórica (sin la inclusión de los futuros valores) y en un horizonte temporal fijado en 1. Habiéndose dispuesto esto, se establece que

$$\texttt{MASE} = \frac{1}{H}\sum_{h=1}^{H} |q_T(h)|$$

Como sucedía con el `MAPE`, el `MASE` también se trata de una medida insensible a la escala de los datos. Este parangona el método de predicción empleado a distintos horizontes con el del *naïve* a horizonte 1. Por ende, un $\texttt{MASE} < 1$ señala que la vía de pronóstico utilizada aporta un rendimiento superior al del predictor *naïve*. En contraposición, esto es, cuando $\texttt{MASE} > 1$, ocurre justamente lo opuesto. Como

[2]El enfoque de predicción *naïve* consiste en predecir el valor x_t a través del valor más recientemente observado, x_{t-k}. Con este mecanismo de pronóstico, el valor futuro de una serie temporal coincidirá con el último valor observado en la serie.

síntesis final, podemos concluir que lo preferible será inclinarse por el método que de lugar a un menor `MASE`.

Exhibidas las diferentes medidas de adecuación de las predicciones, puede emanar la duda de cuál es la más conveniente de emplear en los análisis. La respuesta en este sentido es clara: todo dependerá del contexto y las características particulares de los datos en cuestión. Bien es cierto que regularmente, independientemente de en lo que nos basemos, extraeremos las mismas deducciones; mas pueden darse casuísticas donde se nos presenten resultados contradictorios. En esas coyunturas habrá de valorarse la tipología de nuestro conjunto de datos y qué preferimos priorizar (por ejemplo, minimización del impacto de los valores atípicos o, por contra, minimización del error cuadrático medio).

Antes de rematar con esta sección, debemos comentar dos aspectos relevantes más. En lo tocante al cálculo de los coeficientes mencionados dentro del software R, destacar que los cinco son proporcionados por la función `accuracy()` del paquete `forecast` Hyndman y Athanasopoulos (2023). Por otra banda, en lo concerniente a las fuentes consultadas, estas básicamente fueron Makridakis *et al.* (1984) y Hyndman y Koehler (2006).

Capítulo 3

Modelos univariantes de series temporales

Habiendo presentado ya toda la información íntegra que atañe a las series en su conjunto, además de diversos aspectos genéricos a considerar en cualquier clase de serie temporal, llegó el momento de explicar algunos de los modelos utilizados para describir esas series de tiempo.

Siguiendo con la explicación del párrafo precedente, debemos señalar que, en concreto, durante este capítulo nos focalizaremos en la modelización de series univariantes. Como su propia nomenclatura indica, estos modelos se centran en capturar los patrones y características inherentes a una única variable a lo largo del tiempo, sin contemplar la influencia de potenciales variables exógenas. Su principal cometido es, por tanto, identificar la estructura subyacente de la serie con la finalidad de poder a llevar a cabo predicciones de los valores futuros de esta.

En total, nos serviremos de dos modelos univariantes para hallar *forecast* sobre el volumen de ventas de cerveza de la compañía Hijos de Rivera. Tales modelos serán: el modelo `SARIMA` (encuadrado dentro de la metodología Box-Jenkins) y el modelo `ETS` (modelo de suavizado exponencial).

3.1. Modelos Box-Jenkins

En primer lugar, en el transcurso de este bloque nos dedicaremos a desarrollar la colección de modelos cuyo uso se encuentra más extendido en el análisis de series temporales; estos son, los modelos Box-Jenkins. Diseñados por los estadísticos George Box y Gwilym Jenkins en la década de los años 70, este conjunto de modelos ha logrado probar su efectividad en una amplia variedad de aplicaciones prácticas (Box *et al.* (1994)). Sus bondades más reseñables son: por un lado, su fuerte flexibilidad para adecuarse a una extensa diversidad de series diferentes, y por el otro, su capacidad para proporcionar proyecciones futuras confiables.

Empero, al margen de los modelos elaborados por estas dos ilustres personalidades, se debe señalar su contribución con la denominada, en honor a ambos, metodología Box-Jenkins. Este enfoque, ilustrado anteriormente (véase la subsección 2.1), no solo se aplica específicamente sobre los modelos tipo Box-Jenkins, sino que se ha extendido a prácticamente toda la clase de modelos de series habidos. Es por ello, por lo que en nuestros estudios también nos secundaremos en estas directrices, tomando como referencia lo recogido en Makridakis y Hibon (1997) y Velasco y del Puerto (2009).

3.1.1. Modelos AR, MA y ARMA

En el marco de la modelización Box-Jenkins, los modelos más elementales son los `AR`, `MA`, y la unión de los dos anteriores `ARMA`. Empleados para modelar series ya originalmente estacionarias (es decir, que no necesitan ser transformadas previamente), actúan como base en la construcción del resto de modelos Box-Jenkins más complejos. Compartido esto, procedemos a describir por separado cada uno.

- **Modelos AR:**

Primeramente, decimos que una serie generada a partir de un proceso estacionario, $\{X_t\}_{t\in C}$, sigue un modelo autorregresivo de orden `p`, `AR(p)`, si se satisface la ecuación:

$$X_t = \phi_1 X_{t-1} + \phi_2 X_{t-2} + ... + \phi_p X_{t-p} + a_t$$

donde $\phi_1, ..., \phi_p$ son parámetros del modelo con $\phi_p \neq 0$ y $\{a_t\}_t$ es un proceso de ruido blanco.

Traducida al lenguaje habitual, esta ecuación viene a plasmar la idea de que el valor actual de la serie es una combinación lineal de los `p` valores pasados más recientes de sí misma, más el término de innovación, a_t, el cual incorpora toda la información adicional no explicada a través de los valores pasados. En otras palabras, el proceso `AR(p)` modela la relación lineal existente entre una observación actual y las observaciones antecesoras.

Asimismo, dicho proceso también puede expresarse en función del operador de retardo, denotado por B:

$$\phi_p(B)X_t = (1 - \phi_1 B - \phi_2 B^2 - ... - \phi_p B^p)X_t = a_t$$

donde $\phi_p(B)$ recibe es el polinomio autorregresivo y $(\phi_1, \phi_2, ..., \phi_p)$ es el vector de parámetros autorregresivos.

A priori, la media de todo proceso `AR(p)` será nula, debido a su condición de estacionariedad:

$$(1 - \phi_1 - \phi_2 - ... - \phi_p)\ E(X_t) = 0 \Rightarrow E(X_t) = 0$$

Sin embargo, este modelo puede generalizarse para representar series con medias distintas de cero. De este manera, el modelo `AR(p)` que incluye un parámetro constante, c :

$$X_t = c + \phi_1 x_{t-1} + \phi_2 x_{t-2} + ... + \phi_p x_{t-p} + a_t$$

tiene media:

$$(1 - \phi_1 - \phi_2 - ... - \phi_p)\ E(X_t) = c \Rightarrow E(X_t) = \frac{c}{1 - \phi_1 - \phi_2 - ... - \phi_p}$$

Para todo modelo `AR(p)`, se constata que la `fap` se anula para todo retardo superior a `p`, pero no en el propio retardo `p`. Concretamente, si las variables del proceso $\{X_t\}_t$ son i.i.d., tienen varianza finita y el tamaño T de la serie es lo suficientemente grande, en cualquier `AR(p)` se verifica que:

$$\texttt{AR(p)} \Rightarrow \hat{\alpha}_k \approx N\left(0, \frac{1}{\sqrt{T}}\right), \forall k > p$$

donde $\hat{\alpha}_k$ denota la `fap` muestral.

De este modo, para un nivel de significación del 5 %, siempre y cuando la serie haya sido generada por un proceso `AR(p)` debería corroborarse que, para cada $k > p$:

$$\hat{\alpha}_k \in \left(-\frac{1{,}96}{\sqrt{T}}, \frac{1{,}96}{\sqrt{T}}\right)$$

En la Figura 3.1 se muestra el gráfico secuencial y las correspondientes `fas` y `fap` de una serie simulada a partir de un proceso `AR(2)` de tamaño $T = 100$. A la vista de la gráfica de la `fap`, se observa claramente como la última $\hat{\alpha}_k$ en salir del intervalo $\left(-\frac{1{,}96}{\sqrt{100}}, \frac{1{,}96}{\sqrt{100}}\right)$ es $\hat{\alpha}_2$, dado que el 2° retardo se trata del último que supera las bandas discontinuas azules. Por consiguiente, este proceso `AR(2)` expresa el valor actual, X_t, como una función lineal de `p=2` valores pasados, X_{t-1} y X_{t-2}.

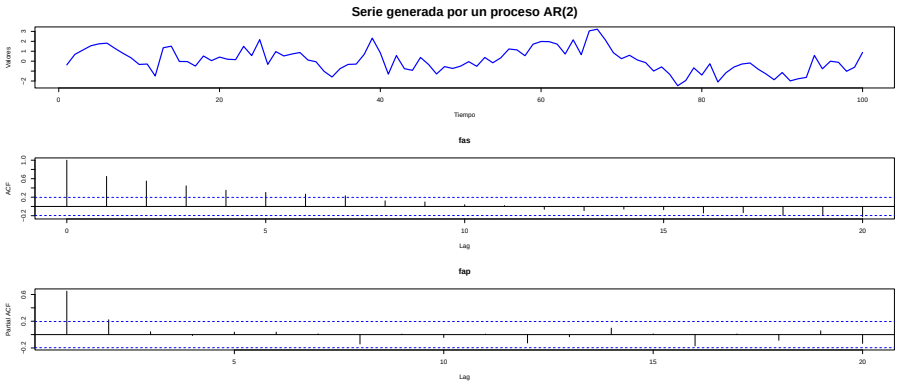

Figura 3.1: Serie simulada a través de un proceso `AR(2)` de parámetros $\phi_1 = 0{,}5$ y $\phi_2 = 0{,}25$.

- **Modelos MA:**

Complementariamente a los modelos autorregresivos, tenemos los modelos de medias móviles. Afirmaremos que una serie generada a partir de un proceso estacionario, $\{X_t\}_{t\in C}$, sigue un modelo de medias móviles de orden `q`, `MA(q)`, si se verifica que:

$$X_t = a_t + \theta_1 a_{t-1} + \theta_2 a_{t-2} + ... + \theta_q a_{t-q}$$

donde $\theta_1, ..., \theta_q$ son parámetros del modelo con $\theta_q \neq 0$ y $\{a_t\}_t$ es un proceso de ruido blanco.

Interpretando la igualdad anterior, podemos inducir que esta clase de procesos se basan en modelizar la dependencia temporal de la serie a través de los términos de error ruido blanco pasados. De esta forma, a medida que se incrementa el orden `q` del modelo, más residuos pasados han de considerarse en la predicción actual.

Análogamente a como acaecía con el proceso `AR(p)`, el proceso `MA(q)` también puede definirse de acuerdo al operador de retardo, B, como sigue:

$$\theta_q(B)a_t = (1 - \theta_1 B - \theta_2 B^2 - ... - \theta_q B^q)a_t = X_t$$

donde $\theta_q(B)$ se denomina polinomio de medias móviles y $(\theta_1, \theta_2, ..., \theta_q)$ se designa vector de parámetros de medias móviles.

Ante la inclusión de un parámetro constante, $E(X_t) \neq 0$, adoptando el modelo `MA(q)` resultante esta notación matemática:

$$X_t = c + a_t + \theta_1 a_{t-1} + \theta_2 a_{t-2} + ... + \theta_q a_{t-q}$$

Para cualquier modelo `MA(q)`, se comprueba que la `fas` se anula para todo retardo superior a `q`, pero no en el propio retardo `q`. Concretamente, si las variables del proceso $\{X_t\}_t$ son i.i.d., tienen varianza finita y el tamaño T de la serie es lo suficientemente grande, en cualquier `MA(q)` se verifica que:

$$\texttt{MA(q)} \Rightarrow \hat{\rho}_k \approx N\left(0, \frac{\sqrt{1 + 2\left(\rho_1^2 + ... + \rho_q^2\right)}}{\sqrt{T}}\right), \forall k > q$$

donde $\hat{\rho}_k$ denota la `fas` muestral.

De este modo, para un nivel de significación del 5 %, siempre y cuando la serie haya sido generada por un proceso `MA(q)` ha de corroborarse que, para cada $k > q$:

$$\hat{\rho}_k \in \left(\pm 1{,}96 \cdot \frac{\sqrt{1 + 2\left(\hat{\rho}_1^2 + ... + \hat{\rho}_q^2\right)}}{\sqrt{T}} \right)$$

Con la Figura 3.2 se ilustra el gráfico secuencial y las respectivas `fas` y `fap` de una serie simulada a partir de un proceso `MA(2)` de tamaño $T = 100$. A la vista de la gráfica de la `fas`, se visualiza claramente como la última $\hat{\rho}_k$ en salir del intervalo $\left(-\frac{1{,}96}{\sqrt{100}}, \frac{1{,}96}{\sqrt{100}}\right)$ es $\hat{\rho}_2$, puesto que el 2° retardo se trata del último que rebasa las bandas azules discontinuas. Por ende, este proceso `MA(2)` explica el valor actual, X_t, como una función lineal de `q=2` valores pasados de un proceso de ruido blanco, a_{t-1} y a_{t-2}.

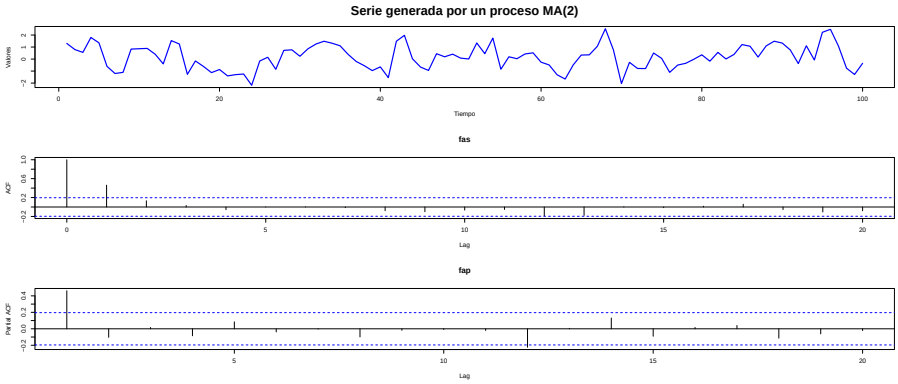

Figura 3.2: Serie simulada a través de un proceso `MA(2)` de parámetros $\phi_1 = 0{,}5$ y $\phi_2 = 0{,}25$.

- **Modelos ARMA:**

La última categoría de modelos Box-Jenkins válidos para modelar series estacionarias son los modelos `ARMA`, los cuales, en realidad, equivalen a una conjunción dentro de un mismo proceso de una estructura `AR(p)` y `MA(q)`. En términos matemáticos, un modelo autorregresivo de media móvil de orden `(p,q)` con $\mu = 0$, o en forma abreviada, `ARMA(p,q)`, se representa por:

$$X_t = \phi_1 X_{t-1} + ... + \phi_p X_{t-p} + a_t + \theta_1 a_{t-1} + ... + \theta_q a_{t-q}$$

donde $\phi_1, ..., \phi_p, \theta_1, ..., \theta_q$ son parámetros del modelo con $\phi_p \neq 0$ y $\theta_q \neq 0$, y $\{a_t\}_t$ es un proceso ruido blanco. De manera simplificada, y por medio del operador de retardo, el proceso `ARMA(p,q)` también podría expresarse en forma compacta:

$$(1 - \phi_1 B - ... - \phi_p B^p) X_t = (1 - \theta_1 B - ... - \theta_q B^q) a_t$$
$$\phi_p(B) X_t = \theta_q(B) a_t$$

donde $\phi_p(B)$ es el polinomio autorregresivo y $\theta_q(B)$ el polinomio de medias móviles.

La condiciones de estacionariedad y causalidad del modelo `ARMA(p,q)` vienen impuestas por la parte autorregresiva, ya que la de medias móviles siempre es estacionaria. Para que satisfagan ambas hipótesis es estrictamente necesario que el módulo de las raíces de $\phi_p(B)$ se halle fuera del círculo unidad.

En contraste, el supuesto de invertibilidad viene dictado por la componente de medias móviles, pues, como ya su propia denominación advierte, la estructura autorregresiva siempre es invertible. En consecuencia, un proceso `ARMA(p,q)` cumplimenta dicha propiedad sí y sólo sí el módulo de las raíces $\theta_q(B)$ no se encuentra dentro del círculo unidad.

La identificación de procesos `ARMAs` mixtos (i.e., modelos `ARMA(p,q)` con `p`$\neq$`0` y `q`$\neq$`0`) a partir de la interpretación gráfica de la `fas` y la `fap` muestrales es una ardua tarea, de ahí que no se maneje esta metodología en la práctica. Bien es cierto que existen otras propuestas para conseguir llevar a cabo esto, como el método de la menor correlación canónica (Galián (1983)), mas ninguna de ellas resulta demasiado popular. Usualmente lo más recurrente será modelizar la serie probando diferentes procesos `ARMA`, quedándonos con aquel que ofrezca un `AIC`, `AICc` o `BIC` menor.

Finalizada toda la descripción de los modelos Box-Jenkins estacionarios, tenemos que resaltar que en el mundo real las series acostumbrar a contar con tendencia y patrones repetitivos. Por tanto, se hace inexorablemente imprescindible extender esta clase de modelos `ARMA` para que sí puedan tomar en consideración algunos otros componentes.

3.1.2. Modelos ARIMA

Se establece que un proceso `ARIMA(p,d,q)` es aquel que, luego de aplicar `d` diferencias regulares para eliminar la tendencia de la serie[1], se transforma en un proceso `ARMA(p,q)`. Por consiguiente, se trata del modelo resultante de la fusión de la diferenciación regular con los modelos autorregresivos y de medias móviles. Esto se traduce en que:

$$\{X_t\}_t \text{ es } \texttt{ARIMA(p,d,q)} \Rightarrow (1-B)^d \; es \; \texttt{ARMA(p,q)}$$

De manera completa, un modelo `ARIMA(p,d,q)` puede escribirse así:

$$\underset{\uparrow}{(1-\phi_1 B - ... - \phi_p B^p)} \; \underset{\uparrow}{(1-B)^d X_t} \;=\; c + \underset{\uparrow}{(1+\theta_1 B + ... + \theta_q B^q)}\, a_t$$

$$\texttt{AR(p)} \qquad\qquad \texttt{d diferencias} \qquad\qquad \texttt{MA(q)}$$

No obstante, equivalentemente, en una versión más reducida tendríamos:

$$\phi_p(B)\,(1-B)^d X_t = c + \theta_q(B)\,a_t$$

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

Para exponer el proceso de identificación de un modelo `ARIMA(p,d,q)` a través del análisis gráfico tomaremos como referencia la serie recogida en la Figura 3.3. En particular, nos fundamentaremos en un *dataset* que contiene información sobre la producción anual de tabaco en Estados Unidos.

En los gráficos concernientes a la serie original se detecta una fuerte tendencia alcista, observándose un lento decrecimiento de los valores positivos en la `fas`. Sin embargo, no se distingue componente estacional. Es por ello, por lo que se hace indispensable diferenciar regularmente la serie (`d=1`) y transformarla aplicándole logaritmos (transformación Box-Cox con $\lambda = 0$). Ahora, sobre la serie alterada y convertida en estacionaria, se selecciona el modelo `ARMA(p,q)` correspondiente, que para este caso, se infiere que podría ser un `ARMA(0,1)`.

Finalmente, para ratificar y comprobar que es adecuado el modelo seleccionado, recurrimos a la función `auto.arima()` del paquete `forecast`, la cual nos devuelve el mejor modelo `ARIMA` en base al criterio de selección elegido (`AICc` para esta instancia). El modelo obtenido, un `ARIMA(0,1,1)`, coincide con el que habíamos determinado antes, por lo que puede proponerse como generador de la serie temporal, a expensas de que supere la diagnosis de sus residuos.

[1] El proceso de diferenciación regular ya ha sido desmenuzado con detalle. Puede consultarse el apartado 2.2.3 si se desea repasarlo de nuevo.

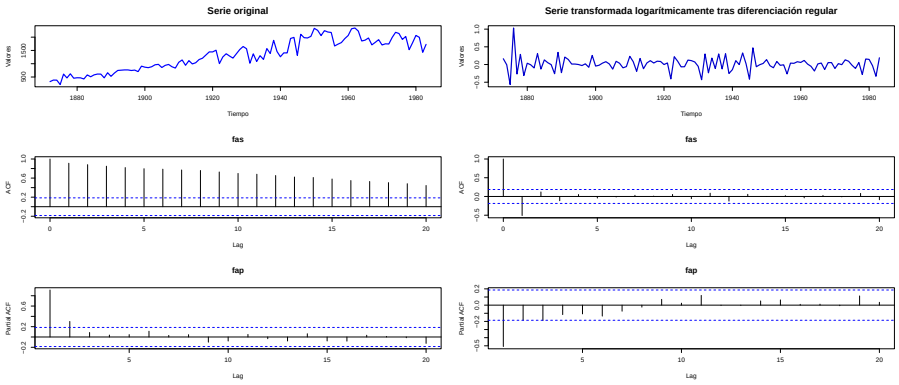

Figura 3.3: A la izquierda se ilustra el gráfico secuencial de una serie de datos reales con su correspondiente `fas` y `fap`. Mientras que a la derecha, se reflejan los mismos gráficos después de haber sido transformada logarítmicamente y diferenciada regularmente la serie. En base a estos últimos, se sugiere que la serie puede ser modelizable por un `ARIMA(0,1,1)`.

Como se ha constatado, los modelos `ARIMA` se ajustan bastante bien a algunas series temporales existentes en la realidad, las cuales presentan una componente tendencial. Empero, hemos de ampliar este tipo de modelos para que tengan la capacidad de adecuarse a series con componente estacional, también muy asiduas en la práctica.

3.1.3. Modelos SARIMA

Los modelos `SARIMA`, o también comúnmente llamados modelos `ARIMA estacionales`, constituyen una extensión de los modelos `ARIMA` utilizados en la modelización de series temporales que exhiben periodicidad, ya sea mensual, trimestral, anual, etc. Así pues, se erigen como una clase de modelos formados por dos partes diferenciadas: una no estacional `(p,d,q)` y otra estacional $\texttt{(P,D,Q)}_s$.

Podemos definir un proceso $\texttt{ARIMA(p,d,q)}\times\texttt{(P,D,Q)}_s$ como aquél que, después de aplicarle `d` diferencias regulares y `D` diferencias estacionales de periodo `s` para suprimir la tendencia y la estacionalidad[2], respectivamente, muta en un proceso $\texttt{ARMA(p,q)}\times\texttt{(P,Q)}_s$. En notación matemática, esto se transcribiría de la siguiente forma:

$$\{X_t\}_t \text{ es } \texttt{ARIMA(p,d,q)}\times\texttt{(P,D,Q)}_s \Rightarrow (1-B)^d\,(1-B^s)^D \;\; es \;\; \texttt{ARMA(p,q)}\times\texttt{(P,Q)}_s$$

En su versión extendida, un modelo $\texttt{ARIMA(p,d,q)}\times\texttt{(P,D,Q)}_s$ puede expresarse así:

$$\underset{\substack{\uparrow\\ \texttt{AR reg.}}}{(1-\phi_pB^p)}\;\underset{\substack{\uparrow\\ \texttt{AR est.}}}{(1-\Phi_pB^{ps})}\;\underset{\substack{\uparrow\\ \texttt{Dif. reg.}}}{(1-B)^d}\;\underset{\substack{\uparrow\\ \texttt{Dif. est.}}}{(1-B^s)^D}X_t = \; c+\underset{\substack{\uparrow\\ \texttt{MA reg.}}}{(1+\theta_qB^q)}\;\underset{\substack{\uparrow\\ \texttt{MA est.}}}{(1+\Theta_qB^{qs})}\,a_t$$

Mas, análogamente, si estipulamos que $\nabla^d = (1-B)^d$ y que $\nabla_s^D = (1-B^s)^D$, la igualdad anterior puede simplificarse tal y como sigue:

$$\phi_p(B)\,\Phi_P(B^s)\,\nabla^d\,\nabla_s^D X_t = c+\theta_q(B)\,\Theta_Q(B^s)\,a_t$$

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

[2]El proceso de diferenciación estacional también ha sido detallado en profundidad. Puede consultarse el apartado 2.2.3 si se necesita revisarlo.

Al igual que hicimos para los modelos `ARIMA`, también mostraremos gráficamente cómo se efectúa el procedimiento de detección del modelo $\texttt{ARIMA(p,d,q)}\times\texttt{(P,D,Q)}_s$ que más adecuadamente se ajusta a una serie temporal específica. En concreto, nos basaremos en la serie representada en la Figura 3.4, donde aparecen registrados los datos referentes a la matriculación mensual de motocicletas en Galicia durante el periodo de 1985 a 2004.

Atendiendo a los gráficos relativos a la serie original, se aprecia que la tendencia es fluctuante conforme transcurre el tiempo. Asimismo, también se atestigua como la `fas` presenta una fuerte correlación positiva en el retardo estacional, siendo, por tanto, manifiestamente patente la estacionalidad de la serie. Debido a esto, y después de que haya sido transformada logarítmicamente la serie inicial, ha de realizarse una diferenciación regular y estacional sobre la misma. Una vez ejecutados todos los pasos relatados, ya resulta factible determinar el modelo $\texttt{ARMA(p,q)}\times\texttt{(P,Q)}_s$ pertinente para este caso, el cual se trata de un $\texttt{ARMA(0,1)}\times\texttt{(0,1)}_{12}$.

Para concluir nos valemos de la función `auto.arima()` con el fin de corroborar que el modelo escogido previamente concuerda con el proporcionado por dicha función. Puesto que sí sucede, se deduce que la serie alterada puede ser compatible con un modelo $\texttt{ARIMA(0,1,1)}\times\texttt{(0,1,1)}_{12}$.

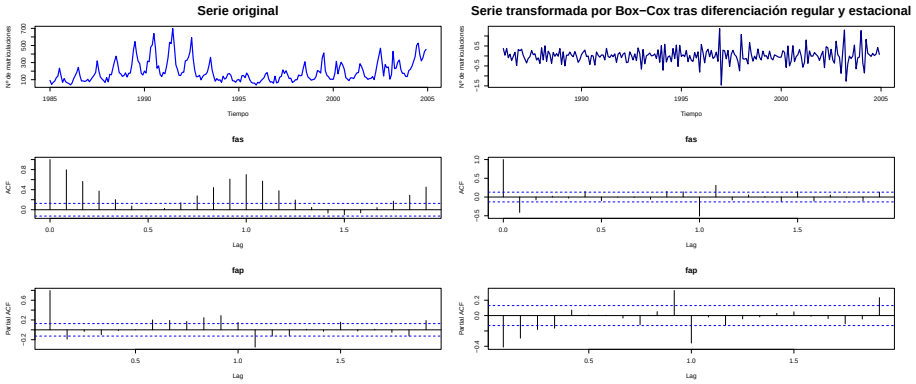

Figura 3.4: A la izquierda se ilustra el gráfico secuencial de una serie de datos reales con su correspondiente `fas` y `fap`. Mientras que a la derecha, se reflejan los mismos gráficos después de que la serie haya sido transformada logarítmicamente y diferenciada tanto regularmente como estacionalmente. En base a estos últimos, se sugiere que la serie puede ser modelizable por un $\texttt{ARIMA(0,1,1)}\times\texttt{(0,1,1)}_{12}$.

Dentro de la metodología Box-Jenkins, si nos enfocamos en la modelización de series univariantes construidas con datos verídicos, el proceso $\texttt{ARIMA(p,d,q)}\times\texttt{(P,D,Q)}_s$ suele emplearse frecuentemente, dado que abundan las series con componente estacional y tendencia. En resumen, podría decirse que se consolida como el modelo generalizador de todos los procesos caracterizados a lo largo de esta sección.

Desplegadas las diferentes variantes fundamentales de modelos Box-Jenkins, tenemos que pasar a explicar cómo se lleva a cabo la estimación y ajuste del modelo, la diagnosis o análisis de sus residuos y el cálculo de las predicciones de sus valores futuros. No obstante, conviene destacar que tales procedimientos resultan extrapolables a una vasta gama de modelos, como por ejemplo los `ETS`.

3.1.4. Estimación y ajuste del modelo

Primeramente, es preciso puntualizar que, tanto en este apartado como en los venideros, nos focalizaremos en describir y aplicar algunas de las fases de la metodología Box-Jenkins sobre los modelos `SARIMA` en concreto. Tomando en consideración que prácticamente todos los datos temporales de ventas de nuestra cervecera se ajustan a esta tipología de modelos, lo idóneo es vincular el enfoque teórico de cada paso a su posterior aplicación práctica sobre una determinada serie de ventas de la compañía.

El punto de partida en el que nos hallamos es el siguiente: tras la identificación del modelo `ARIMA(p,d,q)×(P,D,Q)`$_s$ como posible generador de la serie, se necesita estimar el valor de sus parámetros $(c, \phi_1, ..., \phi_p, \theta_1, ..., \theta_q, \Phi_1, ..., \Phi_p, \Theta_1, ..., \Theta_q)$. Para ello se puede recurrir al **método de Mínimos Cuadrados** (*Least Squares Estimation*) o bien al **método de Máxima Verosimilitud** (*Maximum Likehood Estimation*).

El primero de ellos tiene como cometido primordial minimizar la suma de las innovaciones cuadradas. En este, a partir de un vector de parámetros $(\hat{c}, \hat{\phi}_1, ..., \hat{\phi}_p, \hat{\theta}_1, ..., \hat{\theta}_q, \hat{\Phi}_1, ..., \hat{\Phi}_p, \hat{\Theta}_1, ..., \hat{\Theta}_q)$, se obtienen los residuos, o dicho de otro modo, la diferencia entre los valores observados y los correspondientes valores ajustados:

$$\hat{a}_t = X_t - \left(\hat{c} + \hat{\phi}_p X_{t-p} + \hat{\Phi}_p X_{t-p} + \hat{\theta}_q \hat{a}_{t-q} + \hat{\Theta}_q \hat{a}_{t-q}\right)$$

Por consiguiente, lo que se busca es encontrar los parámetros que minimicen la función S, de forma que

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

Empero, si $p > 0$, se da una dificultad asociada a la estimación por mínimos cuadrados, ya que $(\hat{a}_1, \hat{a}_2, ..., \hat{a}_p)$ depende de los valores no observados de $(X_0, X_{-1}, ..., X_{1-p})$. Para subsanarla ha de aplicarse el denominado **método de Mínimos Cuadrados Condicionados**, una variación de la técnica anterior, donde se minimiza la función

$$S_c(\hat{c}, \hat{\phi}_1, ..., \hat{\phi}_p, \hat{\theta}_1, ..., \hat{\theta}_q, \hat{\Phi}_1, ..., \hat{\Phi}_p, \hat{\Theta}_1, ..., \hat{\Theta}_q) = \sum_{t=p+1}^{T} \hat{a}_t^2$$

condicionada a que

$$\hat{a}_p = \hat{a}_{p-1} = ... = \hat{a}_{p+1-q} = 0$$

Por otro lado, a través del método de Máxima Verosimilitud la estimación de los parámetros $(c, \phi_1, ..., \phi_p, \theta_1, ..., \theta_q, \Phi_1, ..., \Phi_p, \Theta_1, ..., \Theta_q)$ y σ_a^2 se logra a través de los valores $(\hat{c}, \hat{\phi}_1, ..., \hat{\phi}_p, \hat{\theta}_1, ..., \hat{\theta}_q, \hat{\Phi}_1, ..., \hat{\Phi}_p, \hat{\Theta}_1, ..., \hat{\Theta}_q)$ y $\hat{\sigma}_a^2$ que maximizan su función de verosimilitud L. En lenguaje matemático, esto sería:

$$(\hat{c}, \hat{\phi}_p, \hat{\theta}_q, \hat{\Phi}_p, , \hat{\Theta}_q, \hat{\sigma}_a^2) = arg\ máx_{\hat{c}, \hat{\phi}_p, \hat{\theta}_q, \hat{\Phi}_p, \hat{\Theta}_q, \hat{\sigma}_a^2}\ L\left(\hat{c}, \hat{\phi}_p, \hat{\theta}_q, \hat{\Phi}_p, \hat{\Theta}_q, \hat{\sigma}_a^2\right)$$

Bajo ciertas condiciones de regularidad, las estimaciones de los parámetros (salvo σ_a^2) obtenidas por el método de Máxima Verosimilitud son asintóticamente óptimas[3]. Por contra, el estimador de máxima verosimilitud de σ_a^2 solamente debe validar la condición de consistencia.

Señalados y pormenorizados ambos métodos, tenemos que enfatizar que en la práctica suele usarse el método de Máxima Verosimilimitud con mayor asiduidad, dado que es más flexible y proporciona estimadores asintóticamente óptimos. En R la estimación de los valores de los parámetros del modelo son brindados tanto por la función auto.arima() como por la función Arima() (ambas pertenecientes al paquete forecast) de manera automática. Sin embargo, sí ha de cotejarse "manualmente" que todos los parámetros incluidos en el modelo son significativamente distintos de 0, puesto que, en caso contrario, estos serán excluidos del modelo al carecer de relevancia.

[3]Un estimador se considera asintóticamente óptimo cuando verifica las propiedades de consistencia (el estimador converge en probabilidad al valor verdadero del parámetro a medida que el tamaño de la muestra tiende hacia infinito) y de eficiencia asintótica (el estimador tiene la menor varianza asintótica posible entre todos los estimadores consistentes).

Según se recoge en Shumway *et al.* (2000), para contrastar este requerimiento se denota por γ a cualquiera de los anteriores parámetros (excepto σ_a^2), por $\hat{\gamma}$ a su estimación máximo verosímil y por $\hat{\sigma}_{\hat{\gamma}}$ a la desviación típica del estimador correspondiente. Si planteamos siguiente el contraste de hipótesis:

$$\begin{cases} H_0: \ \gamma = 0 \\ H_1: \ \gamma \neq 0 \end{cases}$$

tendremos que:

$$\hat{\gamma} \text{ es distinto de 0 a un nivel de significación del } 5\,\% \Leftrightarrow |\hat{\gamma}| \geq 1{,}96 \cdot \hat{\sigma}_{\hat{\gamma}}$$

Así pues, una vez estimado y ajustado el modelo, ya se puede avanzar a la próxima etapa.

3.1.5. Diagnosis o análisis de los residuos

La fase de diagnosis comprende al conjunto de pruebas de bondad de ajuste que se efectúan sobre el modelo. El propósito es constatar si se cumplen las hipótesis básicas asumidas respecto a él. La condición esencial exigida es que las innovaciones del modelo, $\{a_t\}_t$, sean ruido blanco, es decir, que cuenten con media cero, que tengan varianza constante y además estén incorreladas. Cabe señalar que dicho requerimiento es taxativo, de modo que su invalidación anula al modelo ajustado como posible generador de la serie de estudio. Por ello, en caso de que esto ocurra, deberemos retornar a identificar un nuevo modelo y volver a ejecutar los procedimientos posteriores.

La verificación de la incorrelación de los residuos puede realizarse a través del análisis gráfico de la `fas` y de la `fap`, o bien, a través del contraste de independencia Ljung-Box. Mas, resulta fundamental subrayar que, dado que las innovaciones a_t no son observables, la totalidad del proceso de revisión se practica sobre sus estimaciones $\hat{a}_t$.

Mediante el método gráfico, la comprobación se produce añadiendo dos líneas paralelas a distancia $1{,}96/\sqrt{T}$ del origen, observándose si todos los coeficientes $\hat{\rho}_k$ se sitúan dentro de estas bandas de confianza. Aunque es cierto que contar con un coeficiente fuera de estos límites en un retardo elevado no es problemático, sí que lo es que se de en uno de los retardos iniciales, ya que esto último constituye un claro síntoma de incorrelación de los residuos.

Otro camino es valerse de la prueba de **Ljung-Box** para examinar la independencia de los residuos y, por ende, su no correlación[4] (Ljung y Box (1978)). Sea H el retardo máximo considerado en la prueba, el estadístico de contraste del test se denota así:

$$Q_H = T \cdot (T+2) \sum_{k=1}^{H} \frac{\hat{\rho}_k^2}{T-k}$$

Puesto que este estadístico se distribuye de acuerdo a una χ^2 con $(H - p - q)$ grados de libertad, rechazaremos la hipótesis nula si el valor obtenido de Q es superior al percentil 0,95 de la distribución χ^2.

Para aplicar el contraste de Ljung-Box en R recurrimos a la función `tsdiag()` del paquete `stats`. Sea cualesquiera la serie objeto de estudio, se recomienda usar ambos enfoques para corroborar que se han extraído las mismas conclusiones.

[4] La independencia implica inexorablemente que también exista incorrelación entre los residuos, pero el recíproco no siempre acontece.

Por otra banda, también ha de emplearse el contraste de media nula para confirmar que se garantiza que las innovaciones a_t son ruido blanco (Peña (2005)). Sea una muestra, $y_1, ..., y_T$, afirmaremos que esta proviene de variables aleatorias i.i.d. con media cero y varianza finita cuando:

$$\frac{\bar{y}}{\hat{s}_y/\sqrt{T}} \approx t_{T-1} \approx N(0,1)$$

donde $\bar{y}$ refleja la media muestral y $\hat{s}_y$ la cuasivarianza muestral.

En R dicha prueba de hipótesis viene dada por la función `t.test(x,mu=0)`. Una vez aquí, siempre que se verifiquen tanto este requisito como el anterior (premisas *sine que non*), ya se puede aseverar que el modelo `SARIMA` seleccionado es válido para modelizar la serie temporal.

A pesar de que el incumplimiento de la hipótesis de normalidad no anula al modelo en cuestión como posible generador de la muestra, sí que resulta preferible que haya normalidad en los residuos, debido a que así se garante que no se ha dejado información sin modelizar y que las innovaciones son gaussianas. Al igual que antes, podemos acudir a la representación gráfica o a determinados contrastes para valorar la presencia de normalidad.

En lo tocante a la representación gráfica, acostumbra a construirse el **gráfico Q-Q** (Cuantil-Cuantil), donde se representan los cuantiles muestrales frente a los cuantiles de una distribución $N(0,1)$. En este, la no linealidad indicará ausencia de normalidad, o en caso contrario, lo opuesto.

Una de las pruebas para evaluar si los residuos siguen una distribución normal es el contraste de **Jarque-Bera**, el cual se basa en la asimetría y curtosis de los datos (Jarque y Bera (1980)). Su estadístico de contraste es el siguiente:

$$JB = \frac{T}{6}\left(S^2 + \frac{1}{4}\cdot(K-3)^2\right)$$

donde T es el tamaño muestral, S la asimetría de la muestra y K la curtosis de la muestra. Para más pormenores sobre los diversos enfoques de cálculo de los coeficientes de asimetría y de curtosis puede consultarse Groeneveld y Meeden (1984).

El test se ejecuta en R por medio de la función `jarque.bera.test()` del paquete `tseries`.

El otro gran contraste de normalidad es el de **Shapiro-Wilk** (Shapiro y Wilk (1965)). Para este caso su estadístico de contraste se define tal que así:

$$\omega = \frac{\left(\sum_{t=1}^{[T/2]} b_{t,T}\left(y_{(T-t+1)} - y_{(t)}\right)\right)^2}{T\cdot s_y^2}$$

donde $y_{(t)}$ denota al estadístico ordenado de orden t y las constantes $b_{t,T}$ vienen dadas a partir de la inversa de la distribución normal estándar.

A través de la función `shapiro.test()` se realiza en R.

Presentada la etapa de análisis de los residuos, solamente falta especificar la última fase de la metodología Box-Jenkins; esta es, la de predicción de los valores futuros de la serie.

3.1.6. Predicción de los valores futuros de la serie

Para poder efectuar predicciones acerca de los futuros valores de la variable X en periodos venideros se utiliza la información histórica disponible de la serie temporal. En particular, en los modelos `SARIMA`, los pronósticos se generan mediante un enfoque recursivo, en el cual se utiliza cada nueva observación para predecir la siguiente. Es decir, sea un horizonte temporal h, lo que se hará es predecir los siguientes h valores de la serie en cuestión $(x_{T+1}, x_{T+2}, ..., x_{T+h})$. Formalmente, cada una de estas predicciones serán denotadas por $\hat{x}_T(h)$.

Aunque el propósito primordial durante esta etapa es hallar predicciones sobre los valores futuros del proceso, también es importante saber cuán precisas son dichos pronósticos utilizando los intervalos de predicción.. La construcción de estos intervalos, básicamente, puede hacerse por dos vías: vía asintótica y vía bootstrap.

En lo referente a la vía **asintótica**, es vital tener en cuenta que para seguir este camino ha de validarse la hipótesis de gaussianidad de las innovaciones, o lo que es mismo, los residuos deben distribuirse de acuerdo a una distribución normal. Esto es debido al hecho de que en ausencia de normalidad, no es factible determinar el nivel de confianza de los intervalos. Mediante esta técnica, partiendo de la expresión:

$$X_t = c + a_t + \psi_1 a_{t-1} + \psi_2 a_{t-2} + ...$$

donde los ψ_i son parámetros desconocidos (en función de ϕ_j, θ_k, Φ_l y Θ_m), se calcula el intervalo de predicción al 95 % para el valor de X_{T+k}:

$$\left(\hat{X}_T(h) \pm 1{,}96 \cdot \sqrt{\sigma_a^2 \cdot (1 + \psi_1^2 + ... + \psi_{h-1}^2)}\right)$$

Empero, si la suposición de normalidad falla, los intervalos anteriores carecen de fiabilidad. Ante esta coyuntura, lo más habitual es acudir a la vía **bootstrap**[5]. En ella, primeramente, se simulan los próximos k valores de la serie: $\{x_{T+1}, x_{T+2}, ..., x_{T+k}\}$. Posteriormente, se repite de nuevo este paso una infinidad de veces:

$$\left\{x_{T+1}^{(1)}, x_{T+2}^{(1)}, ..., x_{T+k}^{(1)}\right\}, ..., \left\{x_{T+1}^{(B)}, x_{T+2}^{(B)}, ..., x_{T+k}^{(B)}\right\}$$

Finalmente, los extremos del correspondiente intervalo de predicción se establecerían, para cada horizonte $1 \leq h \leq k$, a partir de los percentiles de los valores $\left\{x_{T+h}^{(j)}\right\}_{i=1}^{B}$.

Ya caracterizados y expuesta minuciosamente la metodología Box-Jenkins de manera teórica, llegó el momento de pasar a plasmarla en base a nuestros datos de estudio.

3.1.7. Aplicación de los modelos Box-Jenkins a los datos de Hijos de Rivera

A lo largo de este apartado, vamos a aplicar una buena parte de la teoría desarrollada previamente sobre una serie de ventas mensuales de cerveza de la compañía Hijos de Rivera. Por razones obvias, en el presente documento resulta inviable mostrar el proceso de implementación de los modelos `SARIMA` en la totalidad de series temporales de las que disponemos. Debido a esto, se ha optado por ilustrar la metodología Box-Jenkins a través de una serie específica de ventas.

Como bien enfatizábamos en la introducción, por motivos de confidencialidad tampoco hay posibilidad de revelar los valores concretos de la serie ni categorizar con exactitud su procedencia; aunque,

[5]El método bootstrap se fundamenta en el muestreo con reemplazo, lo cual conlleva que cada vez que se genera una muestra bootstrap, se seleccionen observaciones de la muestra original de forma aleatoria y con reemplazo. Como resultado, cada muestra bootstrap será ligeramente diferente de las demás. Sin embargo, cabe reseñar que esta es una característica deseable, ya que este hecho nos permitirá evaluar la incertidumbre y construir intervalos de confianza para el estadístico en cuestión.

no obstante, sí que podemos detallar que, en particular, la serie hace referencia a la evolución histórica de los litros de cerveza que han sido vendidos cada mes en la delegación de ventas A de la empresa. Asimismo, también comentar que han sido recopiladas las observaciones que van desde enero del 2010 hasta octubre del año 2022 (últimos registros accesibles) para la construcción de la serie.

Remarcado lo anterior, hemos de proceder al análisis de la serie seleccionada. Así que, inicialmente, comenzamos representando el gráfico secuencial de la misma, el cual aparece reflejado en la Figura 3.5:

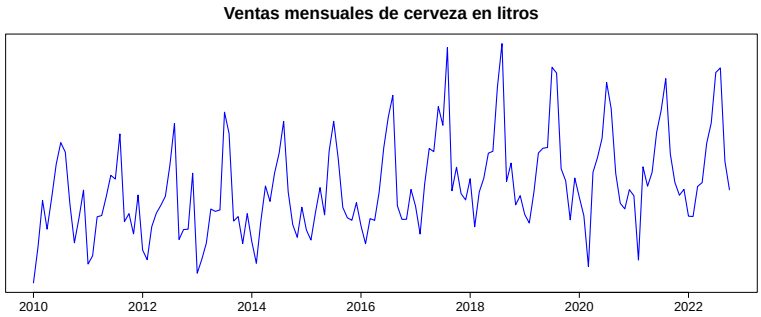

Figura 3.5: Gráfico secuencial de la serie original de ventas de la delegación A.

A la vista de la gráfica, se aprecia que la serie sigue una tendencia medianamente creciente con el paso del tiempo, exceptuando el periodo coincidente con los momentos críticos de la pandemia del coronavirus (desde marzo de 2020 hasta mediados de la primavera de 2021). Por otro lado, se observa presencia de cierta heterocedasticidad en la serie, incrementándose la variabilidad especialmente durante la etapa del coronavirus. Finalmente, en relación a la componente estacional, se evidencia una notaria estacionalidad de carácter mensual, que resulta todavía más palpable si atendemos al diagrama de cajas plasmado en la Figura 3.6. Por medio de este diagrama también se detecta una significativa variación mensual interanual de las ventas, pese a que en el gráfico no se muestren los valores atípicos habidos dentro de cada mes (es decir, se han excluido los *outliers* que se sitúan fuera de los límites del diagrama de cajas).

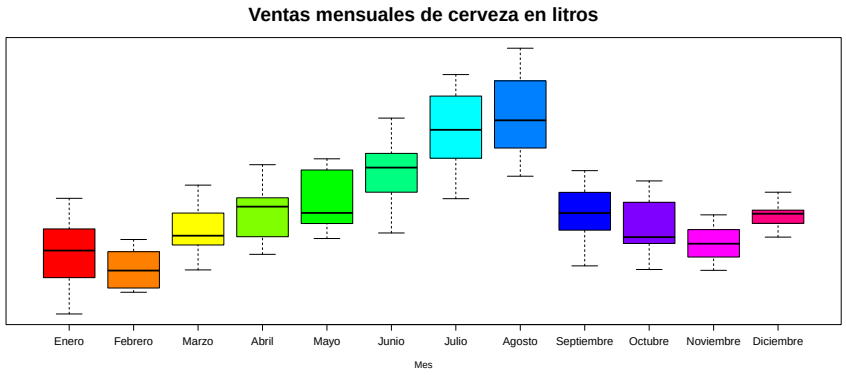

Figura 3.6: Diagrama de cajas de las ventas de cerveza en relación a cada uno de los meses.

Al margen de la evaluación gráfica, la suposición de que la serie no cumple la condición de estacionariedad queda refrendada a través de la prueba de Dickey-Fuller, donde, como era previsible, para un nivel de significación del 5 %, se rechaza (p-valor = 0,001) la H_0 de que la serie sea estacionaria.

Puesto que se ha resuelto que la serie no es estacionaria, debemos aplicarle distintas alteraciones para que así lo sea. De este modo, lo primero que necesitamos es transformar la serie para conseguir su homocedasticidad. Teniendo en cuenta que todos los valores observados de la serie presentan signo positivo, ejecutaremos una transformación Box-Cox sobre los datos. Empleando la función `BoxCox.lambda()`, hallamos que el valor de lambda óptimo es $\hat{\lambda} = 0{,}105$. Dado que dicho valor no difiere apenas del que tendríamos utilizando la transformación logarítmica, cotejamos gráficamente ambas transformaciones para comprobar si dan lugar a series similares. Tal y como se puede percibir en la Figura 3.7, las series resultantes exhiben un patrón de comportamiento prácticamente idéntico, por lo que nos decantamos por trabajar con la serie logarítmica, ya que más adelante será más sencilla de interpretar.

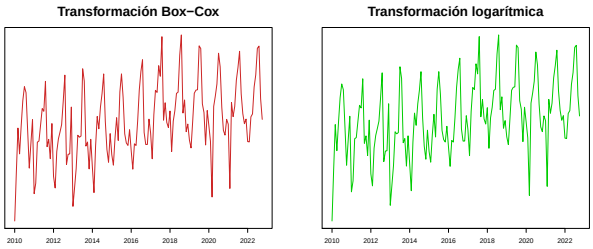

Figura 3.7: Gráficos secuenciales de la serie después de haber sido transformada por Box-Cox ($\hat{\lambda} = 0,105$) y logarítmicamente, respectivamente.

Aunque la transformación logarítmica estabiliza levemente la variabilidad de la serie, esta última continua disparándose (si bien en menor medida) en el periodo de la COVID-19. Complementariamente a su gráfico secuencial, si añadimos su respectivo gráfico de la `fas` (Figura 3.8), se reafirma la imprescindibilidad de diferenciar regularmente y estacionalmente la serie.

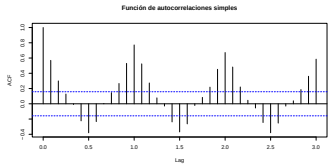

Figura 3.8: `Fas` de la serie logarítmica.

Por consiguiente, primeramente empezamos efectuando una diferenciación regular (`d=1`) sobre la serie transformada logarítmicamente para eliminar la tendencia, obteniéndose el gráfico secuencial y la `fas` expuestos en la Figura 3.9:

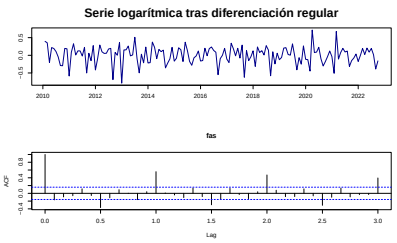

Figura 3.9: Gráfico secuencial (arriba) y `fas` (abajo) de la serie logarítmica tras la diferenciación regular.

La tendencia ha sido suprimida; mas, todavía se visualiza la existencia de una componente estacional de periodo `s=12`, pues los retardos estacionales cuentan con valores destacadamente elevados. Por tanto, se precisa llevar a cabo una diferenciación estacional de orden `D=1`, que da resultado al siguiente gráfico secuencial con sus correspondientes `fas` y `fap` (Figura 3.10):

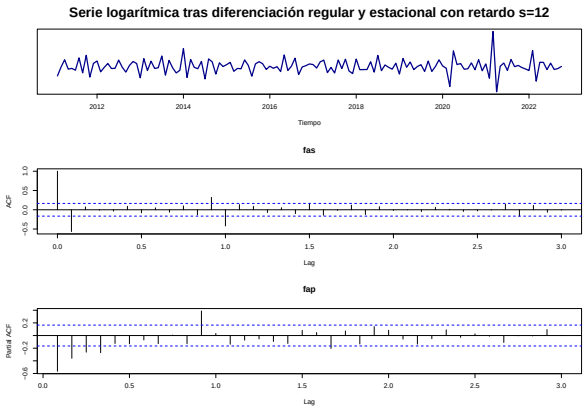

Figura 3.10: Gráfico secuencial, `fas` y `fap` de la serie logarítmica tras diferenciación regular y estacional.

Con la desaparición de la componente estacional, la serie transformada se ha convertido en estacionaria. Sabedores de que estamos antes un modelo `SARIMA` con `d=1`, `D=1` y `s=12`, tenemos que identificar los órdenes del proceso $\texttt{ARMA(p,q)}\times\texttt{(P,Q)}_{12}$ resultante. Para ello, se requiere observar en la Figura 3.10 la `fas` y la `fap` de la serie modificada.

En lo referente a la identificación de la estructura regular, hemos de prestar atención a los retardos iniciales. Si examinamos la gráfica de la `fas`, se vislumbra que el último retardo en salirse de las bandas con claridad es el primero, de forma que sugeriríamos un modelo `MA(1)` para la parte regular; mientras que, si, en cambio, inspeccionamos la gráfica de la `fap`, el último retardo en traspasar los límites es el cuarto, pudiéndose proponer así un modelo `AR(4)` para la parte regular. Debido a que la intención siempre es la de estimar el modelo más sencillo posible, nos decantamos por el modelo `MA(1)` como factible generador de la parte regular.

En contraste, para determinar los órdenes `P` y `Q` de la parte estacional hay que fijarse en los retardos estacionales (`s`, `2s`, `3s`,...), tomando en consideración que `s=12`. Con un procedimiento similar al del párrafo anterior, concluimos que el modelo $\texttt{MA(1)}_{12}$ es un potencial generador de la parte estacional.

Por ende, y recapitulando las deducciones obtenidas a partir del estudio de las gráficas, la serie logarítmica de ventas parece compatible con un modelo $\texttt{ARIMA(0,1,1)}\times\texttt{(0,1,1)}_{12}$. Sin embargo, debemos ratificar que se trata del modelo más adecuado en base a los criterios de selección de modelos (para este caso concreto usaremos la vía `AICc`, porque es más asintóticamente eficiente que el criterio `BIC` y castiga más poderosamente el *overfitting* que el criterio `AIC`). La función `auto.arima()` devuelve el modelo $\texttt{ARIMA(1,1,2)}\times\texttt{(0,1,1)}_{12}$, el cual es parcialmente distinto del identificado mediante el análisis gráfico. Ante esta situación, optamos por quedarnos con este nuevo modelo sugerido.

Sea $Y_t = log\,(X_t)$, si definimos el modelo $\texttt{ARIMA(1,1,2)}\times\texttt{(0,1,1)}_{12}$ para la serie logarítmica, este quedaría expresado en notación matemática así:

$$(1-\phi_1\,B)\,(1-B)\,(1-B^{12})\,Y_t = \;(1+\theta_1\,B+\theta_2\,B^2)\,(1+\Theta_1\,B^{12})\,a_t$$

Equivalentemente, en su versión ampliada se denotaría de esta manera:

$$Y_t = (1+\phi_1)\,Y_{t-1} - \phi_1\,Y_{t-2} + Y_{t-12} - (1+\phi_1)\,Y_{t-13} + a_t + \theta_1\,a_{t-1} + \theta_2\,a_{t-2} + \Theta_1\,a_{t-12} + \\ + \theta_1\,\Theta_1\,a_{t-13} + \theta_2\,\Theta_1\,a_{t-14}$$

Nótese que se ha desechado de incluir una constante, c, a raíz de que su inserción implicaría contar con una tendencia polinómica de grado `d+D=2`, la cual podría derivar en que se cometiesen grandes errores en la etapa de predicción.

Ahora, a partir del modelo seleccionado, estimamos por Máxima Verosimilitud el valor de sus parámetros, valiéndonos de la función Arima() del paquete forecast de R. La salida generada se muestra en la Figura 3.11:

```
Series: serie
ARIMA(1,1,2)(0,1,1)[12]
Box Cox transformation: lambda= 0

Coefficients:
        ar1      ma1     ma2     sma1
        0.4468  -1.5800  0.6871  -0.7389
s.e.    0.1971   0.1636  0.1521   0.0782

sigma^2 = 0.0152:  log likelihood = 91.59
AIC=-173.18   AICc=-172.73   BIC=-158.44
```

Figura 3.11: Valores estimados de los parámetros del modelo `SARIMA` ajustado.

Traduciendo el código de la salida de `R` a lenguaje matemático, inferimos que:

$$\hat{\phi}_1 = 0{,}4468\,(\sigma = 0{,}1971)$$
$$\hat{\theta}_1 = -1{,}5800\,(\sigma = 0{,}1636)$$
$$\hat{\theta}_2 = 0{,}6871\,(\sigma = 0{,}1521)$$
$$\hat{\Theta}_1 = -0{,}7389\,(\sigma = 0{,}0782)$$

Tras acreditar que efectivamente todos los parámetros incluidos en el modelo resultan significativamente distintos de 0 a un nivel de significación del 5 % (no es necesario prescindir de ninguno), sustituyendo dichas estimaciones en el modelo en versión extendida tendríamos que:

$$Y_t = 1{,}4468\,Y_{t-1} - 0{,}4468\,Y_{t-2} + Y_{t-12} - 1{,}4488\,Y_{t-13} + a_t - 1{,}5800\,a_{t-1} + 0{,}6871\,a_{t-2} - 0{,}7389\,a_{t-12} + 1{,}1675\,a_{t-13} - 0{,}5077\,a_{t-14}$$

Finalizada la fase de estimación de los parámetros, pasamos a desarrollar la diagnosis de los residuos del modelo con el objetivo de validar si sus innovaciones, $\{a_t\}_t$, son ruido blanco. Asimismo, también evaluaremos si estas son gaussianas.

Sirviéndonos del contraste de independencia de Ljung-Box (para el cual se obtuvo un p-valor = 0,5848) y de los gráficos de la Figura 3.12 corroboramos que se cumplen las condiciones de independencia e incorrelación de los residuos. De forma análoga, constatamos que no se rechaza la H_0 de media nula, por lo que podemos asegurar que el modelo $\texttt{ARIMA(1,1,2)}\times\texttt{(0,1,1)}_{12}$ es válido para modelar la serie logarítmica de ventas. Por el contrario, a la vista del gráfico Q-Q (Figura 3.13) y de los contrastes de Shapiro-Wilk y de Jarque-Bera (p-valores resultantes de $2{,}2\,e^{-16}$ y $1{,}53\,e^{-5}$, respectivamente), se extrae que no se satisface la hipótesis de normalidad de los residuos y que, por tanto, las innovaciones no son gaussianas. Atendiendo al gráfico secuencial de la serie original (Figura 3.5), se deduce que la razón fundamental de que no haya normalidad es por causa de la drástica contracción vivida en las ventas durante el periodo de la pandemia. Bien es cierto que, aunque las diversas transformaciones realizadas sobre la serie mitigan este efecto (Figura 3.10), este todavía prosigue siendo de carácter considerable.

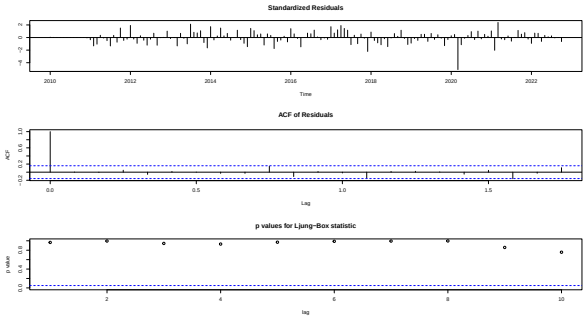

Figura 3.12: Comprobación de la hipótesis de independencia de los residuos.

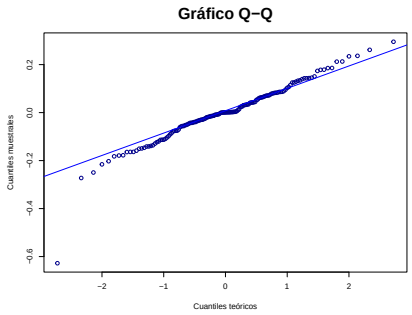

Figura 3.13: Gráfico Q-Q.

Para terminar con la metodología Box-Jenkins, tenemos que estimar cuáles son las predicciones de los valores futuros de la serie fundamentándonos en el modelo ajustado previamente. Como consecuencia de que las innovaciones no son gaussianas, únicamente podremos construir los intervalos de predicción vía bootstrap. El horizonte temporal de pronóstico será $h = 14$ con la finalidad de abarcar lo que restaba de 2022 y todo el año 2023 en su conjunto.

Así pues, con la función forecast() del paquete homónimo estimamos las predicciones recogidas en el cuadro 3.1[6] (son redondeadas a valores enteros para facilitar su compresión), las cuales son representadas gráficamente en la Figura 3.14. Además, también calculamos las principales medidas de adecuación de las predicciones (Figura 3.15).

Fechas	Predicción	Inf.80	Sup.80	Inf.95	Sup.95
Nov 2022	217250	188573	279007	195447	303424
Dec 2022	241912	210021	278999	197848	304206
Jan 2023	209979	1812116	241241	169809	260942
Feb 2023	177281	153275	204408	142215	221960
Mar 2023	220271	190953	254896	177911	276816
Apr 2023	253696	218990	293793	202354	320380
May 2023	279275	240692	324250	219474	355470
Jun 2023	314542	271496	365708	246772	398117
Jul 2023	376298	323079	441927	294824	482219
Aug 2023	393562	337423	460245	305001	502865
Sep 2023	263031	223884	308239	201071	334314
Oct 2023	237932	202917	280588	184815	307423
Nov 2023	218259	183815	261906	161997	285859
Dec 2023	245471	205817	293251	186619	320101

Cuadro 3.1: Predicciones e intervalos de confianza obtenidos (tanto al 80 % como al 95 %) por la vía bootstrap para la serie de ventas de la delegación A.

[6]Hemos de recalcar que absolutamente todos los valores revelados han sido transformados por un parámetro λ para respetar la confidencialidad de los datos reales.

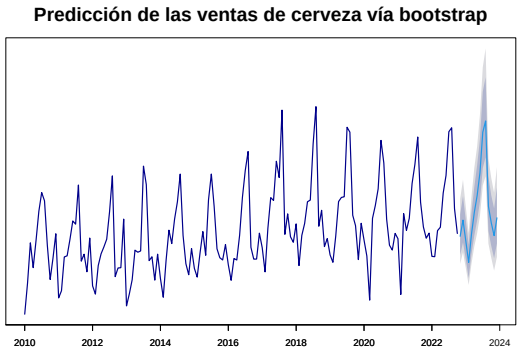

Figura 3.14: Gráfico secuencial de la serie acompañado de las predicciones y los intervalos de confianza estimados (tanto al 80 % como al 95 %) vía bootstrap para los 14 próximos meses.

	ME	RMSE	MAE	MPE	MAPE	MASE	ACF1
Training set	756.4982	39008.33	28793.83	-0.6880857	8.547733	0.7575428	0.0237935

Figura 3.15: Medidas de adecuación de las predicciones halladas.

A la luz de las medidas de adecuación de las predicciones computadas, podemos afirmar que la capacidad predictiva del modelo anterior es relativamente buena. Por ejemplo, el `MAPE` toma un valor inferior al 10 %, lo que, en términos generales, suele interpretarse como un elevado nivel de precisión. Asimismo, el `MASE` < 1, de modo que los pronósticos determinados poseen un rendimiento superior a los del predictor *naïve*. Empero, y pese a que las medidas de adecuación no lo recogen, resulta importante resaltar que dichas estimaciones de las predicciones para los próximos meses probablemente pueden estar infraestimadas, es decir, que presenten un valor menor al que realmente podrían tener en realidad. Esto es debido a la influencia de los meses donde caen dramáticamente las ventas por la huella de la pandemia del coronavirus.

3.2. Modelos de suavizado exponencial

Inicialmente, se debe poner de manifiesto que a lo largo de toda la sección nos centraremos en abordar el otro gran conjunto de modelos destinados al estudio de series de tiempo: los modelos `ETS` (*Exponential Smoothing State Space*). Cabe subrayar que esta clase de modelos emergen a partir de los métodos de pronóstico de suavizado exponencial formulados a finales de la década de los años 50. Particularmente, ha de destacarse la obra de estadísticos como Charles C. Holt o John F. Winters, los cuales contribuyeron enormemente al desarrollo de uno de los enfoques más exitosos en el campo de la predicción (Holt (1957) y Winters (1960)).

Resumidamente, los métodos de suavizado exponencial se basan en efectuar pronósticos futuros por medio de la ponderación de las observaciones pasadas, de forma que los pesos de cada observación van decayendo exponencialmente a medida que estas son más antiguas. En otras palabras, estos métodos se sustentan en la idea de que los valores más recientes de la serie temporal tienen una mayor relevancia para predecir el futuro que los valores más antiguos. En ellos la totalidad de las ponderaciones dependen de un parámetro constante, conocido como parámetro de suavizado. Las mayores virtudes de este tipo de métodos residen en su simplicidad estructural, su capacidad de adaptabilidad y su sencilla interpretación intuitiva.

Para afrontar la explicación, primero apuntaremos los elementos básicos en los que se descompone un modelo `ETS`. Después, ahondaremos en los cimientos de los métodos de suavizado exponencial más relevantes, con el fin de ligar dicha exposición a los modelos estadísticos subyacentes a los métodos descritos[7]. En última instancia, serán aplicadas buena parte de estas nociones teóricas sobre una serie real (específicamente, sobre la ya utilizada en la sección previa). Para ello, nos cimentaremos en Hyndman *et al.* (2002) y Hyndman *et al.* (2008).

3.2.1. Descomposición de los elementos de un modelo ETS

Tal y como su propio acrónimo señala, fundamentalmente, un modelo `ETS` está compuesto por tres elementos: **tendencia** (`T`), **estacionalidad** (`S`) y **error** (`E`). La tendencia y estacionalidad ya fueron ilustradas anteriormente[8], mientras que con "error" aludimos al componente estocástico del modelo, en otras palabras, a la variabilidad aleatoria que no puede ser elucidada por otras componentes. En contraste a los modelos Box-Jenkins, los modelos `ETS` no requieren verificar que la serie sea homocedástica para su correcta implementación, puesto que se adaptan a variabilidades constantes y cambiantes con el transcurso del tiempo.

A su vez, estas tres componentes pueden combinarse de diversas maneras para el ajuste de nuestra serie en cuestión. Se distinguen dos grandes métodos de descomposición:

– **Descomposición aditiva**: La serie se expresa como la suma de sus componentes:

$$\mathtt{X} = \mathtt{T} + \mathtt{S} + \mathtt{E}$$

En este enfoque, el efecto de la estacionalidad es invariable a lo largo del tiempo y se añade como sumando a la serie observada.

[7] Tenemos que puntualizar que método de predicción y modelo estadístico se refieren a conceptos diferentes, aunque ambos se encuentren inheremente relacionados. En concreto, mientras que los métodos de suavizado exponencial son simplemente técnicas de pronóstico para predicciones puntuales; los segundos son representaciones que proporcionan una base teórica y estadística para comprender y analizar estos métodos, tratándose, por tanto, de estructuras considerablemente más complejas.

[8] Los términos de tendencia y estacionalidad junto con el de heterocedasticidad ya fueron examinados pormenorizadamente. Puede consultarse el apartado 2.1 si se necesita revisarlos de nuevo.

– **Descomposición multiplicativa**: Se considera que la serie es producto de sus tres componentes:

$$\mathtt{X} = \mathtt{T} \times \mathtt{S} \times \mathtt{E}$$

En esta planteamiento, el impacto de la estacionalidad varía proporcionalmente con la tendencia al paso del tiempo.

Ambos procedimientos de descomposición son facilitados por la función decompose del paquete stats en R. Sin embargo, en realidad, también puede darse cualquier otra combinación entre estas dos, como por ejemplo:

$$\mathtt{X} = (\mathtt{T} + \mathtt{S}) \times \mathtt{E}$$

En cualquier caso, la elección de cuál es la técnica de descomposición más óptima estará supeditada a la naturaleza de la serie y a la relación entre su nivel y variabilidad.

3.2.2. Clasificación de los métodos de suavizado exponencial

La categorización de los métodos de suavizado exponencial se adecuará a las características de las agrupaciones de tendencia (`T`) y estacionalidad (`S`), sin importar la manera en la que se agregue el error (`E`), pues no crea, en absoluto, ninguna diferencia en las predicciones puntuales que se obtengan (Gardner Jr (1985)). Es por ello, por lo que comúnmente se ignoran en la definición de los métodos (no ocurre en los modelos `ETS` que veremos luego).

En lo que respecta a la tendencia, esta es entendida como una conjunción del nivel (ℓ) y la pendiente (b), parámetros los cuales deben ser estimados. Según como los integremos se disciernen cinco variedades de tendencias futuras. De este modo, sea T_h la tendencia predicha para h tiempos futuros y ϕ un parámetro de suavizado ($0 < \phi < 1$), tenemos estos tipos de tendencia:

– **Nula** (`N`): $T_h = \ell$. No hay tendencia futura esperada. Se espera que la serie experimente una evolución en torno al nivel al nivel horizontal l.

– **Aditiva** (`A`): $T_h = \ell + b \cdot h$. Se prevé una evolución lineal de la serie, con pendiente b.

– **Aditiva suavizada** ($\mathtt{A}_d$): $T_h = \ell + (\phi + \phi^2 + ... + \phi^h) \cdot b$. Con la inclusión del parámetro de suavizado, se vaticina una evolución lineal que se atenuará, aproximándose a una constante en algún momento futuro (la pendiente b va disminuyendo con el tiempo).

– **Multiplicativa** (`M`): $T_h = \ell \cdot b^h$. Se augura una evolución exponencial en la serie a medida que avanza el tiempo (la pendiente b aumentará o menguará en un mayor grado conforme pase el tiempo).

– **Multiplicativa suavizada** ($\mathtt{M}_d$): $T_h = \ell \cdot b^{(\phi+\phi^2+...+\phi^h)}$. Al igual que acontecía con la tendencia aditiva suavizada, el parámetro ϕ moderará la evolución exponencial de la serie.

No obstante, cuando planteamos un método de suavizado exponencial, además de establecer la variedad de tendencia a incorporar, ha de añadirse la componente estacional (`S`) de forma aditiva ($\mathtt{T} + \mathtt{S}$) o de forma multiplicativa ($\mathtt{T} \times \mathtt{S}$) o, en el supuesto de que no lo haya, de forma nula (`N`). En consecuencia, podemos establecer un total de quince métodos producto de las distintas combinaciones entre tendencia y estacionalidad. Estos aparecen reflejados en el cuadro 3.2. Pese a que aquí hemos englobado a los métodos con tendencia multiplicativa, bien es cierto, que en ocasiones, estos se omiten de ser mostrados debido a que tienden a brindar pronósticos deficientes.

Tendencia \ Estacionalidad	Nula	Aditiva	Multiplicativa
Nula	(N,N)	(N,A)	(N,M)
Aditiva	(A,N)	(A,A)	(A,M)
Aditiva suavizada	(A_d,N)	(A_d,A)	(A_d,M)
Multiplicativa	(M,N)	(M,A)	(M,M)
Multiplicativa suavizada	(M_d,N)	(M_d,A)	(M_d,M)

Cuadro 3.2: Clasificación *two-way* de los métodos de suavizado exponencial.

Cabe enfatizar que algunos de los métodos exhibidos gozan de denominación propia. En nuestro caso profundizaremos en los tres métodos más reconocidos en el marco teórico del suavizado exponencial, ya que, como podremos corroborar más adelante, todos ellos siguen un mismo patrón estructural, el cual puede ser amplificado para el conjunto de métodos de esta índole.

- **Método de Suavizado Exponencial Simple:**

También llamado método SES, constituye el más simple de los métodos de suavizado exponencial (N,N). Resulta idóneo para construir pronósticos de datos que no cuentan ni con una tendencia clara ni ningún patrón estacional (Ostertagova y Ostertag (2012)). Fruto de que la tendencia y la componente estacional sean nulas, este método se define a través de dos únicas ecuaciones: una para el nivel (ℓ) y otra para las predicciones. Matemáticamente dichas ecuaciones se denotarían así:

$$\begin{aligned} &\text{Ecuación de nivel:} && \ell_t = \alpha \cdot x_t + (1-\alpha) \cdot \ell_{t-1} \\ &\text{Ecuación de pronóstico:} && \hat{x}_{t+h|t} = \ell_t \end{aligned}$$

donde ℓ_t hace referencia a una estimación del nivel de la serie en t y donde α es el parámetro de suavizado para el nivel ($0 \leq \alpha \leq 1$).

En la primera ecuación, en esencia, lo que se hace es calcular el nivel que tenía la serie hasta el instante $t-1$, para a posteriori corregirlo levemente con la nueva observación añadida; en tanto que, en la segunda se computa una función de pronóstico "plana", donde la totalidad de las predicciones toman el mismo valor (concretamente el de la componente del último nivel).

- **Método de tendencia lineal de Holt:**

El método lineal de Holt (A,N) es aquél que carece de componente estacional, pero que, en contraposición, sí presenta una tendencia aditiva (Gardner Jr y McKenzie (1985)). Como resultado, ahora intervienen dos elementos no observables: el nivel (ℓ) y la tendencia (b). Es por ello, por lo que este método involucra a una ecuación de pronóstico y dos de suavizado:

$$\begin{aligned} &\text{Ecuación de nivel:} && \ell_t = \alpha \cdot x_t + (1-\alpha) \cdot (\ell_{t-1} + b_{t-1}) \\ &\text{Ecuación de tendencia:} && b_t = \beta^* \cdot (\ell_t - \ell_{t-1}) + (1-\beta^*) \cdot b_{t-1} \\ &\text{Ecuación de pronóstico:} && \hat{x}_{t+h|t} = \ell_t + h \cdot b_t \end{aligned}$$

donde ℓ_t es una estimación del nivel de la serie en el tiempo t, b_t una estimación de la tendencia (pendiente) de la serie en el tiempo t, α el parámetro de suavizado para el nivel ($0 \leq \alpha \leq 1$) y β^* el parámetro de suavizado para la tendencia ($0 \leq \beta^* \leq 1$).

Tal y como en el método SES, la ecuación de nivel viene a ser una combinación convexa entre la observación en el instante t (x_t) y el nivel del instante anterior (ℓ_{t-1}), mas añadiendo ahora el efecto de la pendiente b_{t-1}. En cambio, la ecuación de tendencia es una unión convexa de la pendiente en el instante anterior y la diferencia de niveles en esos instantes. Finalmente, los pronósticos son estimados como una función lineal de h.

- **Método estacional aditivo de Holt-Winters:**

Según se documenta en Chatfield (1978), en el método de Holt-Winters aditivo (`A,A`) se implican todos los elementos no observables posibles: el nivel (ℓ), la pendiente (b) y la estacionalidad (s). Lo conforman, por tanto, cuatro ecuaciones:

$$\begin{aligned}
&\text{Ecuación de nivel:} && \ell_t = \alpha \cdot (x_t - s_{t-m}) + (1-\alpha) \cdot (\ell_{t-1} + b_{t-1}) \\
&\text{Ecuación de tendencia:} && b_t = \beta^* \cdot (\ell_t - \ell_{t-1}) + (1-\beta^*) \cdot b_{t-1} \\
&\text{Ecuación estacional:} && s_t = \gamma \cdot (x_t - \ell_{t-1} - b_{t-1}) + (1-\gamma) \cdot s_{t-m} \\
&\text{Ecuación de pronóstico:} && \hat{x}_{t+h|t} = \ell_t + h \cdot b_t + s_{t+h-m(k+1)}
\end{aligned}$$

donde γ denota el parámetro de suavizado para la componente estacional, m la frecuencia de la estacionalidad y k es la parte entera de $(h-1)/m$ que asegura que las estimaciones de los índices estacionales usados para la previsión proceden del último año de la muestra.

En contraste a la ecuación de tendencia, que es idéntica a la del método lineal de Hont, se encuentra la ecuación de nivel, la cual se construye como un promedio ponderado entre la observación ajustada estacionalmente ($x_t - s_{t-m}$) y el promedio no estacional ($l_{t-1} + b_{t-1}$) para el tiempo t. Por otra banda, la nueva ecuación agregada, la estacional, expresa el promedio ponderado entre el índice estacional actual ($x_t - \ell_{t-1} - b_{t-1}$) y el índice estacional del mismo periodo del año pasado (i.e., m periodos anteriores).

Desarrolladas las ecuaciones de los métodos más emblemáticos, en el próximo apartado nos dedicaremos a detallar los modelos que se fundamentan en el conjunto de estos métodos.

3.2.3. Clasificación de los modelos de suavizado exponencial

Como citábamos al principio de este capítulo, de cada método visto en el punto anterior surgen dos modelos diferentes: uno incorporando errores de manera aditiva (`A`) y otro manera multiplicativa (`M`). Debido a esto, los también denominados *State Space Models*, se denotan de acuerdo a la naturaleza de cada una de las tres componentes de la serie.

A pesar de que los modelos `ETS` generan exactamente los mismos pronósticos que sus respectivos métodos, estos, por el contrario, sí permiten hallar intervalos de predicción. Asimismo, debemos poner en relieve que, si se opta por un idéntico parámetro de suavizado, los pronósticos puntuales no variarán, independientemente del tipo de error (`A` o `M`) utilizado; empero, en contraste, los intervalos de predicción divergirán.

A continuación, en la página siguiente, se recogen dos tablas (cuadro 3.3 y cuadro 3.4) en las cuales se registran las ecuaciones explícitas para cada uno de los distintos modelos que se pueden crear. No obstante, se han excluido aquellos que incluyen tendencias multiplicativas (`M`) y multiplicativas suavizadas ($\texttt{M}_d$), dado que en la realidad prácticamente no se disponen de series que se amolden a esos atributos. Tales modelos pueden ser ejecutados en R ajustando la función `ets()` del paquete `forecast`.

Las predicciones puntuales que nos ofrecen todos estos modelos se pueden obtener fácilmente iterando sus ecuaciones explícitas para $t+1, ..., t+h$ y fijando $a_{t+i} = 0$ para todo $i = 1, ..., h$. La mayoría de las veces estas predicciones van a coincidir con la media condicional $u_t(h)$, salvo aquellos con estacionalidad multiplicativa. Como regla general, destacar que la distribución de las predicciones es gaussiana para los modelos lineales y no gaussiana para los no lineales.

Concluido el desglose de la base teórica en la que se enmarcan los métodos y modelos de suavizado exponencial, es momento de llevar a la práctica algunos de las ideas enunciadas.

T \ E	N	A	M
N	$x_t = \ell_{t-1} + a_t$ $\ell_t = \ell_{t-1} + \alpha a_t$	$x_t = \ell_{t-1} + s_{t-m} + a_t$ $\ell_t = \ell_{t-1} + \alpha a_t$ $s_t = s_{t-m} + \gamma a_t$	$x_t = \ell_{t-1} s_{t-m} + a_t$ $\ell_t = \ell_{t-1} + \alpha a_t / s_{t-m}$ $s_t = s_{t-m} + \gamma a_t / l_{t-1}$
A	$x_t = \ell_{t-1} + b_{t-1} + a_t$ $\ell_t = \ell_{t-1} + b_{t-1} + \alpha a_t$ $b_t = b_{t-1} + \beta a_t$	$x_t = \ell_{t-1} + b_{t-1} + s_{t-m} + a_t$ $\ell_t = \ell_{t-1} + b_{t-1} + \alpha a_t$ $b_t = b_{t-1} + \beta a_t$ $s_t = s_{t-m} + \gamma a_t$	$x_t = (\ell_{t-1} + b_{t-1}) s_{t-m} + a_t$ $\ell_t = \ell_{t-1} + b_{t-1} + \alpha a_t / s_{t-m}$ $b_t = b_{t-1} + \beta a_t / s_{t-m}$ $s_t = s_{t-m} + \gamma a_t / (l_{t-1} + b_{t-1})$
$\mathtt{A}_d$	$x_t = \ell_{t-1} + \phi b_{t-1} + a_t$ $\ell_t = \ell_{t-1} + \phi b_{t-1} + \alpha a_t$ $b_t = \phi b_{t-1} + \beta a_t$	$x_t = \ell_{t-1} + \phi b_{t-1} + s_{t-m} + a_t$ $\ell_t = \ell_{t-1} + \phi b_{t-1} + \alpha a_t$ $b_t = \phi b_{t-1} + \beta a_t$ $s_t = s_{t-m} + \gamma a_t$	$x_t = (\ell_{t-1} + \phi b_{t-1}) s_{t-m} + a_t$ $\ell_t = \ell_{t-1} + \phi b_{t-1} + \alpha a_t / s_{t-m}$ $b_t = \phi b_{t-1} + \beta a_t / s_{t-m}$ $s_t = s_{t-m} + \gamma a_t / (l_{t-1} + \phi b_{t-1})$

Cuadro 3.3: Ecuaciones explícitas de los modelos ETS con errores aditivos.

T \ E	N	A	M
N	$x_t = \ell_{t-1}(1 + a_t)$ $\ell_t = \ell_{t-1}(1 + \alpha a_t)$	$x_t = (\ell_{t-1} + s_{t-m})(1 + a_t)$ $\ell_t = \ell_{t-1} + \alpha(\ell_{t-1} + s_{t-m}) a_t$ $s_t = s_{t-m} + \gamma(\ell_{t-1} + s_{t-m}) a_t$	$x_t = \ell_{t-1} s_{t-m}(1 + a_t)$ $\ell_t = \ell_{t-1}(1 + \alpha a_t)$ $s_t = s_{t-m}(1 + \gamma a_t)$
A	$x_t = (\ell_{t-1} + b_{t-1})(1 + a_t)$ $l_t = (\ell_{t-1} + b_{t-1})(1 + \alpha a_t)$ $b_t = b_{t-1} + \beta(\ell_{t-1} + b_{t-1}) a_t$	$x_t = (\ell_{t-1} + b_{t-1} + s_{t-m})(1 + a_t)$ $\ell_t = \ell_{t-1} + b_{t-1} + \alpha(\ell_{t-1} + b_{t-1} + s_{t-m}) a_t$ $b_t = b_{t-1} + \beta(\ell_{t-1} + b_{t-1} + s_{t-m}) a_t$ $s_t = s_{t-m} + \gamma(l_{t-1} + b_{t-1} + s_{t-m}) a_t$	$x_t = (\ell_{t-1} + b_{t-1}) s_{t-m}(1 + a_t)$ $\ell_t = (\ell_{t-1} + b_{t-1})(1 + \alpha a_t)$ $b_t = b_{t-1} + \beta(\ell_{t-1} + b_{t-1}) a_t$ $s_t = s_{t-m}(1 + \gamma a_t)$
$\mathtt{A}_d$	$x_t = (\ell_{t-1} + \phi b_{t-1})(1 + a_t)$ $l_t = (\ell_{t-1} + \phi b_{t-1})(1 + \alpha a_t)$ $b_t = \phi b_{t-1} + \beta(\ell_{t-1} + \phi b_{t-1}) a_t$	$x_t = (\ell_{t-1} + \phi b_{t-1} + s_{t-m})(1 + a_t)$ $\ell_t = \ell_{t-1} + \phi b_{t-1} + \alpha(\ell_{t-1} + \phi b_{t-1} + s_{t-m}) a_t$ $b_t = \phi b_{t-1} + \beta(\ell_{t-1} + \phi b_{t-1} + s_{t-m}) a_t$ $s_t = s_{t-m} + \gamma(\ell_{t-1} + \phi b_{t-1} + s_{t-m}) a_t$	$x_t = (\ell_{t-1} + \phi b_{t-1}) s_{t-m}(1 + a_t)$ $\ell_t = (\ell_{t-1} + \phi b_{t-1})(1 + \alpha a_t)$ $b_t = \phi b_{t-1} + \beta(\ell_{t-1} + \phi b_{t-1}) a_t$ $s_t = s_{t-m}(1 + \gamma a_t)$

Cuadro 3.4: Ecuaciones explícitas de los modelos ETS con errores multiplicativos.

3.2.4. Aplicación de los modelos de suavizado exponencial a los datos de Hijos de Rivera

Durante este apartado aplicaremos los modelos ETS sobre la serie de ventas mensuales de cerveza ya trabajada en la sección previa (véase la Figura 3.5), con el propósito de poder cotejar la adecuación de las predicciones aportadas por cada modelo. Por ende, para esta tarea seguiremos de cerca la metodología Box-Jenkins, ya expuesta en detalle, adaptándola en este escenario al tratamiento de los modelos ETS. En virtud de que para esta clase de modelos no se asumen condiciones tan estrictas (recordemos que, por ejemplo, no se requiere que la serie verifique la hipótesis de homocedasticidad), el proceso de análisis del mismo se simplificará bastante.

Aunque precedentemente ya sacamos conclusiones con respecto a la tendencia (T) y la componente estacional (S) de la serie original, vamos a corroborar que realmente eran acertadas, además de evaluar a mayores el error (E). Para ello, efectuamos una descomposición aditiva ($\mathtt{X} = \mathtt{T} + \mathtt{S} + \mathtt{E}$) y otra multiplicativa ($\mathtt{X} = \mathtt{T} \times \mathtt{S} \times \mathtt{E}$) de la serie a través de la función decompose en R. Las gráficas resultantes para los dos métodos aparecen representadas en la Figura 3.15 y la Figura 3.16, respectivamente.

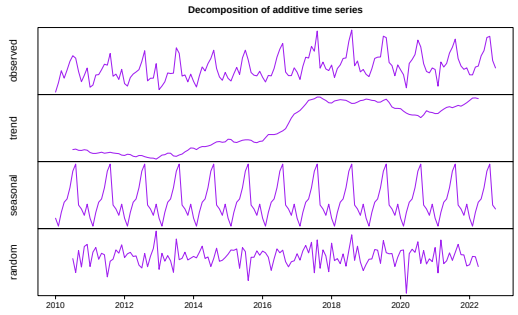

Figura 3.16: Descomposición aditiva.

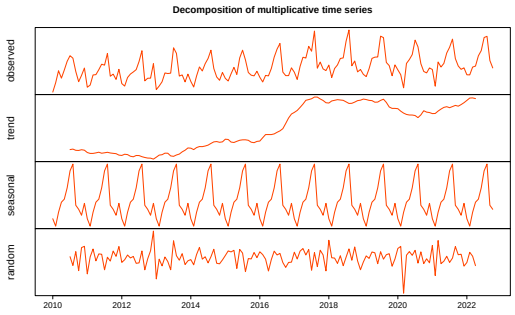

Figura 3.17: Descomposición multiplicativa.

A la vista de ambas gráficas, se constata que la serie cuenta con una intensa estacionalidad mensual y una tendencia creciente en sus inicios. Sin embargo, a partir de 2018, esta comienza a suavizarse y durante el periodo de la pandemia del coronavirus experimenta una disminución, para luego retomar a incrementarse en esta última fase. En lo tocante al error, debemos remarcar su abrupto descenso momentáneo, que también coincide con la etapa del coronavirus. En consecuencia, podemos inferir que el modelo `ETS` a ajustar no poseerá ningún elemento de carácter nulo.

Debido a que aquí no resulta necesario realizar ninguna transformación sobre la serie original, nos enfocamos ya en la elección del modelo más adecuado (del mismo modo que antes, escogeremos aquél que de lugar a un menor `AICc`) y la estimación de sus parámetros. Valiéndonos de la función `ets()` en R, se genera la siguiente salida (Figura 3.18):

```
ETS(M,A,M)
Smoothing parameters:
alpha = 0.1182
beta  = 0.0841
gamma = 1e-04
Initial states:
l = 302348.3802
b = 655.5802
s = 0.9666 0.8129 0.9013 0.9556 1.4926 1.3837
1.1814 1.0299 0.9808 0.8733 0.6671 0.7549
sigma:  0.1144
AIC     AICc      BIC
4049.738 4054.238 4101.366
```

Figura 3.18: Modelo `ETS` ajustado y valores de sus parámetros.

Como puede visualizarse en el código anterior, el `AICc` más reducido se obtiene para el modelo `ETS(M,A,M)`, es decir, un modelo con tendencia aditiva, estacionalidad multiplicativa y errores multiplicativos. Asimismo, las estimaciones de los parámetros son:

$$\hat{\alpha} = 0{,}1182 \qquad \hat{\beta} = 0{,}0841 \qquad \hat{\gamma} = 0{,}0001$$

En cuanto a los valores elementos no observables, se tiene que:

$$l_0 = 302348{,}3802$$
$$b_0 = 655{,}5802$$
$$(s_{12+1}, ..., s_0) = (0{,}9666, ..., 0{,}7549)$$

De acuerdo con lo anotado en el cuadro 3.4, las ecuaciones explícitas para este modelo `ETS(M,A,M)` se plasman tal que así:

$$\begin{aligned}
\ell_t &= (\ell_{t-1} + b_{t-1}) \cdot (1 + 0{,}1182 \cdot a_t) \\
b_t &= b_{t-1} + 0{,}0841 \cdot (\ell_{t-1} + b_{t-1}) \cdot a_t \\
s_t &= s_{t-12} \cdot (1 + 0{,}0001 \cdot a_t) \\
x_t &= (\ell_{t-1} + b_{t-1}) \cdot s_{t-12} \cdot (1 + a_t)
\end{aligned}$$

Finalmente, la Figura 3.19 recoge la representación gráfica de los valores estimados por el modelo para cada observación, incluyendo la tendencia, la componente estacional y el error.

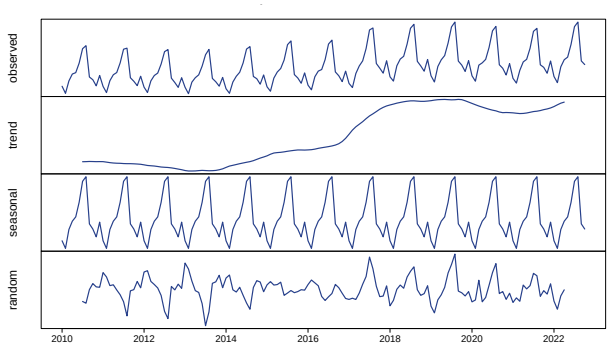

Figura 3.19: Descomposición de los valores de los elementos estimados por el modelo `ETS(M,A,M)`.

Ajustado el modelo, antes de hallar las predicciones puntuales hemos de validar que las innovaciones del modelo son gaussianas, pues si no es el caso, no será posible recurrir a la vía asintótica para generar los pronósticos. Del mismo modo que sucedía en el modelo `SARIMA`, tras atender al gráfico Q-Q (Figura 3.20) y a los resultados de los contrastes de Shapiro-Wilk y Jarque-Bera (p-valores obtenidos de $5{,}773\, e^{-15}$ y 0,000165, respectivamente), se extrae que no se satisface la hipótesis de normalidad de los residuos y que, por consiguiente, las innovaciones no son gaussianas.

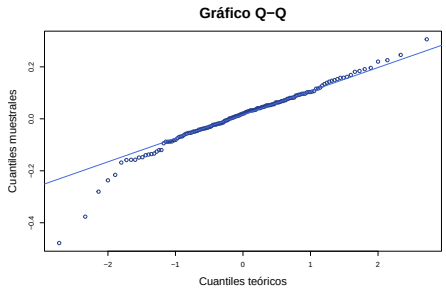

Figura 3.20: Gráfico Q-Q.

Para terminar estimamos las predicciones de los valores futuros de la serie en un $h = 14$ vía bootstrap, basándonos en el modelo `ETS` ajustado. Los pronósticos generados y sus intervalos de predicción (todos se representan alterados por el mismos parámetro λ de la sección anterior) se recopilan en el cuadro 3.5 y la Figura 3.21, a la vez que las medidas de adecuación de estas predicciones aparecen en la Figura 3.22.

Fechas	Predicción	Inf.80	Sup.80	Inf.95	Sup.95
Nov 2022	207718	179112	235000	156161	251051
Dec 2022	248718	213771	281218	190343	298551
Jan 2023	197119	171235	222379	151497	237561
Feb 2023	175929	151002	199020	129153	212346
Mar 2023	228804	196336	258147	172432	276256
Apr 2023	253163	219825	287227	193811	305239
May 2023	264587	229395	300222	201850	321325
Jun 2023	305314	262531	347457	230587	372284
Jul 2023	364732	314080	416908	271230	446560
Aug 2023	378062	325459	430762	290261	460303
Sep 2023	244336	210602	278839	184353	298193
Oct 2023	229848	196474	263282	168887	281208
Nov 2023	207718	176610	238101	153146	255503
Dec 2023	248718	212738	283242	185659	303829

Cuadro 3.5: Predicciones e intervalos de confianza obtenidos (tanto al 80 % como al 95 %) por la vía bootstrap para la serie de ventas de la delegación A.

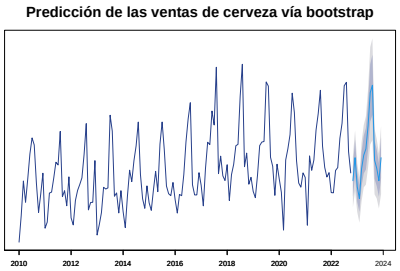

Figura 3.21: Gráfico secuencial de la serie acompañado de las predicciones y los intervalos de confianza estimados (tanto al 80 % como al 95 %) vía bootstrap para los 14 próximos meses.

```
                    ME     RMSE      MAE        MPE     MAPE      MASE       ACF1
Training set 4155.672 37428.67 28239.95 -0.2409337 8.606215 0.7429706 -0.2164038
```

Figura 3.22: Medidas de adecuación de las predicciones halladas.

Resulta llamativo comentar que las medidas de adecuación de las predicciones computadas son muy similares a las obtenidas para el modelo `ARIMA(1,1,2)`$\times$`(0,1,1)`$_{12}$ (Figura 3.15). Por tanto, se deduce que la capacidad predictiva de ambos modelos es semejante. Aún así, si nos basamos en el `MAPE` y el `MAPE`, el modelo `SARIMA` estaría prediciendo infímamente mejor, puesto que se obtienen unos valores ligeramente menores para ambas medidas. No obstante, tenemos que insistir en el hecho de que posiblemente los pronósticos estén infraestimados.

Capítulo 4

Modelos avanzados de series temporales

Tras caracterizar y explorar en detalle los dos modelos más emblemáticos y usuales en el análisis de series temporales, se ha llegado a la conclusión de que ninguno de ellos resulta óptimo para extraer predicciones confiables acerca de los datos de ventas de la empresa. Mas, pese a esto, se ha deducido que, aún recurriendo a modelos más sotisficados, seguiríamos contando con el mismo problema de infraestimación en los pronósticos.

Como bien hemos venido remarcando hasta ahora, en general, durante el periodo de pandemia ha tenido lugar un intenso descenso en las ventas de la mayoría de las delegaciones de la compañía. En otras palabras, nos encontramos con que en el transcurso de esa etapa las ventas han sido considerablemente más reducidas de lo que teóricamente habría correspondido. Puesto que que dicho *shock* es de carácter transitorio en buena parte de las series, parece evidente que estamos frente a un caso claro de presencia de valores atípicos. Es por ello, por lo que lo primero a resolver será el tratamiento de lo que autodenominaremos "efecto covid", pues constituye el elemento que más negativamente ha repercutido en la calidad de las predicciones, al margen de la complejidad del modelo utilizado.

Sin embargo, no ha de obviarse que, una vez solucionado lo previamente mencionado, tenemos que acudir a modelos más avanzados de series temporales, con el objetivo de perfeccionar la precisión de los pronósticos. Por un lado, y siempre que la naturaleza de la serie lo permita, debemos valorar qué factores externos condicionan significativamente las futuras ventas de la compañía. Asimismo, también resulta conveniente evaluar cuál es el nivel de desagregación óptimo para cada canal y tener la capacidad, por ende, de obtener predicciones globales más certeras.

4.1. Variaciones temporales: Efecto covid

Primeramente, cabe subrayar que a lo largo de esta sección incidiremos en el manejo del "efecto covid", el aspecto que más complicaciones ha causado en las estimaciones de la corporación Hijos de Rivera en estos últimos tiempos. Por consiguiente, es de vital importancia para la empresa lograr implementar un método que pueda estimar las ventas futuras a medio plazo, sin que estas se vean muy influidas por el impacto de la pandemia. Con este fin, nos respaldaremos fundamentalmente en lo recogido en Matarise *et al.* (2012).

4.1.1. Marco teórico

Antes de adentrarnos específicamente en los cambios temporales, abordaremos el concepto de valor atípico de manera amplia. Ergo, podemos definir **valor atípico** como aquella observación o conjunto de observaciones que se alejan sustancialmente del patrón general de la serie. Fruto de esta tesitura, todas estas observaciones pueden tener una notable influencia dentro del análisis de la serie y en la posterior capacidad predictiva del modelo ajustado como generador de la misma. Dado entonces su intenso potencial distorsionador, es elemental ser capaces de detectarlos adecuadamente para mejorar el rendimiento de los modelos y pronósticos.

Según el número de observaciones a las que afecten y su efecto sobre los siguientes valores de la serie, se distinguen varios tipos de *outliers*, tales como los atípicos innovativos (`IO`), los atípicos aditivos (`AO`), los atípicos de cambio de nivel (`LS`), etc. No obstante, en lo que respecta a nuestro caso, nos focalizaremos exclusivamente en los atípicos de cambio transitorio, o también conocidos como variaciones temporales (`TC`).

Con **variaciones temporales** nos referimos a las alteraciones que ocurren en un espacio temporal como resultado de un determinado fenómeno pasajero. En consecuencia, es conveniente resaltar que los *temporary changes* no reflejan una situación permanente en el tiempo, sino que hace alusión a las que se presentan de forma momentánea y luego desaparecen. Como resultado de lo anterior, su impacto sobre la serie va a ir disminuyendo exponencialmente con el paso del tiempo, hasta que esta acabe retomando a su nivel original.

En notación matemática, los efectos transitorios ocasionados por la intervención se modelizan a través de una variable ficticia, denominada variable impulso ($I_t^{(h)}$), la cual se expresaría así:

$$I_t^{(h)} = \begin{cases} 0, \text{ si } t \neq h \\ 1, \text{ si } t = h \end{cases}$$

A partir de la variable impulso que hemos denotado, se puede establecer el efecto de la intervención, que consiste en variar los valores de la serie desde el instante $t + h$ hasta el instante $t = h + m$. El modelo resultante que quedaría es el siguiente:

$$Y_t = \omega_0 \cdot I_t^{(h)} + \omega_1 \cdot I_t^{(h+1)} + ... + \omega_m \cdot I_t^{(h+m)} + X_t$$

donde ω_j plasma el número de unidades en los que varía la serie en una determinada observación, X_t el proceso `ARIMA` que generó la serie original y Y_t el proceso que generó la serie intervenida.

Equivalemente, en su versión extendida, dicho modelo se representaría como:

$$Y_t = \begin{cases} X_t, & \text{si } t < h \\ \omega_0 + X_t, & \text{si } t = h \\ \omega_1 + X_t, & \text{si } t = h + 1 \\ \vdots & \vdots \\ \omega_m + X_t, & \text{si } t = h + m \\ X_t, & \text{si } t > h + m \end{cases}$$

El proceso de detección de las variaciones temporales se efectuará en R mediante la función tso() del paquete tsoutliers. Para ello, además será necesario indicar en sus argumentos que solamente identifique los valores atípicos de tipo `TC`, ya que si no lo puntualizamos, la función localizará por defecto toda clase de valor de atípico que pueda hallarse en la serie. Por tanto, debemos explicitar el argumento types="TC" en la propia función para hacer constatar esto. Asimismo, cualquier salto temporal detectado será reemplazado por los valores interpolados linealmente en relación a las observaciones vecinas.

Como se puede apreciar, este paquete aporta un método fuertemente automatizado, que especialmente será de gran utilidad cuando se precise localizar valores atípicos de cambio transitorio en una gran infinidad de series. Puede consultarse información más detallada sobre el funcionamiento de la función tso() en López de Lacalle (2019).

4.1.2. Detección de los TC en los datos de Hijos de Rivera

Pormenorizada la base teórica relativa a las variaciones temporales, llegó la hora de implementar el procedimiento descrito de corrección de series temporales que cuentan con TC. En prácticamente todos los escenarios se modificarán los valores correspondientes al periodo de la COVID-19, lapso comprendido entre marzo de 2020 y mediados del año 2021. No obstante, hemos de recalcar que, dependiendo de como haya sido la incidencia de la pandemia en cada región, en algunas delegaciones de ventas esta etapa puede extenderse hasta ya entrados en el 2022. Además, aunque son minoritarios, también nos topamos con otros TC sin motivación específica debido a la influencia del coronavirus. Por lo general, esta coyuntura acostumbra a darse en áreas donde la marca no está completamente implantada.

Tras este análisis global de la transcendencia de los TC sobre los datos de las distintas delegaciones de ventas de Hijos de Rivera, ejemplificaremos cómo y en qué cuantía se modifica una serie de ventas afectada negativamente por la COVID-19. Al igual que anteriores capítulos, tampoco podremos clarificar a cuál delegación particular son acordes esas ventas; mas, la metodología a seguir, tanto en este caso como en los otros, será siempre la misma.

Para ilustrar el proceso y también mostrar otras series de ventas de la compañía (no ceñirnos siempre a la misma), tomaremos como punto de partida la serie de ventas de la delegación B de Hijos de Rivera, cuyo gráfico secuencial se ilustra en la Figura 4.1. El denominado "efecto covid" resulta muy patente, pues las ventas experimentan una abrupta reducción en marzo de 2020 y no recuperan sus niveles pre-covid hasta inicios del 2022. Tampoco hay lugar a dudas de que mencionada etapa concuerda claramente con un TC, dada la transitoriedad del fenómeno.

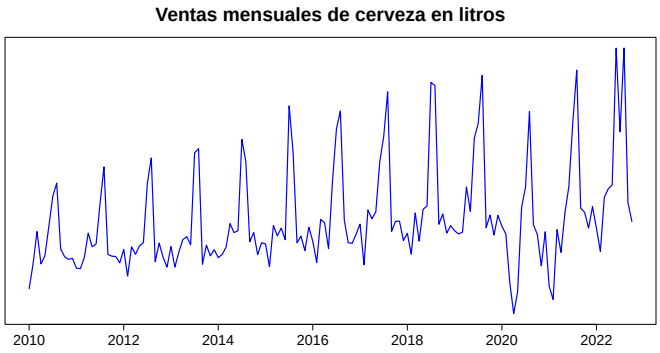

Figura 4.1: Gráfico secuencial de la serie original de ventas de la delegación B.

Tal y como hemos venido comentando, este episodio, además de perjudicar al cumplimiento de la hipótesis de gaussianidad en las innovaciones, incide adversamente en la exactitud de los pronósticos, tendiendo estos últimos a la infraestimación. Ante esta coyuntura, hemos de ejecutar la función tso() para que sea detectados los TC y "se arregle así la serie contaminada por la pandemia". A continuación, graficamos la salida generada, plasmada en la Figura 4.2.

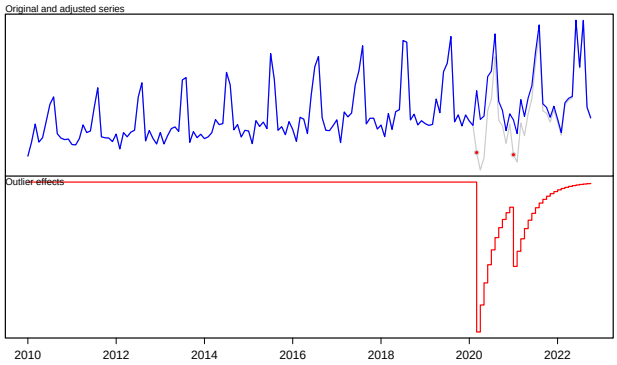

Figura 4.2: Representación del impacto de los valores atípicos TC y corrección del efecto covid.

En la gráfica superior, la serie original se representa en color gris, mientras que la serie corregida tras el ajuste del "efecto covid" se traza en color azul. Por otro lado, en la gráfica inferior, se muestra en color rojo las cuantía del impacto estimado de los TC sobre las ventas mensuales. Esta cantidad refleja el valor de disminución en las ventas que se da mes a mes y que se le atribuye a la pandemia (por razones de confidencialidad no estamos capacitados para desvelar los valores concretos de dicho impacto). A lo largo de este periodo temporal, se distinguen dos grandes puntos críticos: uno en marzo de 2020 y otro en enero de 2021, en los cuales las ventas sufren pronunciados descensos.

A la vista de ambas gráficas, se advierte que la serie obtenida estima presumiblemente bien los valores que teóricamente habría mostrado la serie original desde marzo de 2020 hasta entrado el 2022, si no se hubiese producido el "efecto covid". Por consiguiente, optamos por trabajar con esta nueva serie semificticia durante las etapas de ajuste del modelo y de cálculo de las predicciones, pues, a priori, nos aportará pronósticos más fiables.

A pesar de que hemos evidenciado el enfoque metodológico seguido a través de una única serie de tiempo, el procedimiento expuesto resulta extrapolable al resto de series. Justo aquí debajo se registra el código a utilizar en R para llevar a cabo el procedimiento anterior:

```
library(tsoutliers)
serie<-ts(datos,frequency=12,start=c(2010,1),end=c(2022,10))
fit_serie<-tso(serie,types="TC")
serie_corregida<-fit_serie$yadj
```

Listing 4.1: Código de R para detectar las variaciones temporales debidas al COVID-19 en una serie de ventas

Hemos de poner de relieve que todas las series fueron sometidas a la función tso() con el propósito de adecuar todas sus observaciones a lo que, en ausencia de variaciones temporales, debería haber sucedido. Debido a que mayoritariamente el coronavirus afectó negativamente a las ventas de las diversas delegaciones, durante el periodo de la COVID-19 las series tendieron a enmendarse hacia arriba al aplicárseles la función tso(). Sin embargo, existen honrosas excepciones donde el COVID-19 desencadenó el efecto contrario, es decir, las ventas se dispararon artificialmente hacia cifras excepcionalmente elevadas, como por ejemplo aconteció con el *ecommerce*.

Explayado y solventado el problema primordial que degradaba todos nuestros pronósticos, ahora debemos descubrir y analizar los modelos que mejor se ajustan a las características de nuestros datos (en los futuros análisis ya partiremos siempre desde las series de tiempo corregidas).

4.2. Modelos de regresión dinámica

Tanto los modelos Box-Jenkins como los de suavizado exponencial del capítulo anterior se centran en modelizar y obtener predicciones en base a las observaciones históricas de la serie en cuestión, pero ninguno de ellos admite la inclusión de información externa que pueda afectar a la propia variable de interés. Este hecho puede limitar enormemente la capacidad predictiva de estos modelos, ya que la variable objeto de estudio es posible que se encuentre marcada e influenciada por los valores que toman otras variables ajenas. Así pues, dado que la variable "ventas de cerveza" también puede estar correlacionada con potenciales variables exógenas, hemos de pasar a desarrollar una nueva clase de modelos que sí incorporen variables regresoras: los modelos de regresión dinámica. Para su caracterización nos sustentaremos en lo recogido en Hyndman y Athanasopoulos (2018), mayormente, y Pankratz (2012), en menor medida.

4.2.1. Fundamentos teóricos

Los modelos de regresión dinámica constituyen una extensión de los modelos Box-Jenkins, en los cuales, además, se tienen en cuenta otras covariables. Por ende, se tratan de estructuras que surgen fruto de la combinación de los modelos de regresión tradicionales con los modelos `ARIMA`. En general, acostumbran a utilizarse cuando se disponen de razones empíricas para considerar que lo que se pretende modelar no solo depende de su historial, sino que también de otras variables que no forman parte de la serie.

A diferencia de los modelos `ARIMA` clásicos, este conjunto de modelos permiten que los errores de la regresión estén autocorrelados, ajustándose dicha autocorrelación por medio de un modelo `ARIMA` o `SARIMA`, según sea el caso. Sea una variable X que viene determinada por las variables explicativas $X_1, ..., X_k$, podemos establecer el siguiente modelo de regresión dinámica:

$$\begin{aligned} Y_t &= \beta_0 + \beta_1 X_{1,t} + ... + \beta_k X_{k,t} + \eta_t \\ \text{donde} \quad \eta_t &\sim \texttt{ARIMA(p,d,q)} \times \texttt{(P,D,Q)}_s \end{aligned}$$

Analógamente, en su versión extendida, matemáticamente se denotaría así:

$$\phi_p(B)\,\Phi_P(B^s)\,\nabla^d\,\nabla_s^D Y_t = c + \delta(B)X_t + \theta_q(B)\,\Theta_Q(B^s)\,a_t$$

Tal y como se observa, en los modelos de regresión dinámica el comportamiento de la variable dependiente, Y_t, se explica a través del propio pasado de Y_t, y a través del pasado de una o más variables regresoras, $X_{k,t}$. Asimismo, el modelo contiene dos términos de error: los errores del modelo de regresión, especificados por η_t, y los errores del modelo `ARIMA`, designados por a_t. No obstante, los únicos que han de verificar la hipótesis de ruido blanco son los segundos.

Si exclusivamente tenemos una variable exógena, puede darse la circunstancia de que el efecto de la misma se manifieste en la variable endógena con cierto nivel de retardo, r, puesto que esta viene expresada en formato de serie temporal. Para ese caso, el modelo se representaría de esta forma:

$$Y_t = \beta_0 + \gamma_0\,X_t + \gamma_1\,X_{t-1} + ... + \gamma_r\,X_{t-r} + \eta_t$$

Mas, cuando se presenta más de una variable regresora, la inclusión de retardos puede volverse problemática debido a que aumenta el riesgo de la multicolinealidad y que, por consiguiente, pueda darse una correlación más elevada entre los retardos de las diferentes variables. Por estos motivos, siempre que contemos dos o más variables explicativas, se recomienda delimitar un modelo sin o con muy pocos retardos para alguna de las variables.

En lo que respecta a la estimación de los parámetros del modelo, debemos aclarar que lo que se ha de minimizar es la suma de los valores a_t al cuadrado. Por el contrario, si optásemos por minimizar la suma de los valores η_t al cuadrado, se desencadenarían una serie de problemáticas:

- Los coeficientes estimados $\hat{\beta}_0, ..., \hat{\beta}_k$ dejarían de ser las mejores estimaciones, al ignorarse una porción de la información en el proceso de cálculo.
- Cualquier contraste estadístico relativo al análisis de los residuos del modelo sería improcedente.
- Los criterios de selección de modelos (`AIC`, `AICc` o `BIC`) ya no serían fiables para seleccionar el modelo más adecuado para la fase de predicción.
- En la mayoría de los casos, podría acaecer un fenómeno de regresión espuria[1].

Debido a todos estos motivos, siempre tenemos que minimizar la suma de los valores a_t al cuadrado con el objetivo eludir estos contratiempos.

Por otro lado, a la hora de estimar un modelo de regresión con errores `ARIMA` (connotación con la que también se hace referencia a los modelos de regresión dinámica) se precisa que la serie Y_t y todos las series predictoras $(X_{1,t}, ..., X_{k,t})$ sean estacionarias. Por tanto, esto mismo es lo primero a examinar inicialmente. En caso de que tan solo una serie no cumpla la hipótesis de estacionariedad, los coeficientes estimados no serán consistentes y, en consecuencia, tampoco serán significativos. Empero, para la eventualidad excepcional de que las variables no estacionarias estén cointegradas[2], no se necesitará efectuar ninguna diferenciación, ya que los coeficientes serán igualmente consistentes.

Habitualmente, lo común es diferenciar todas las variables si una necesita ser alterada. En ese escenario al modelo resultante se le denominará "modelo en diferencias", en vez de "modelo en niveles", que es el que se consigue con los datos originales sin sufrir transformación alguna.

Aunque el procedimiento propuesto pueda dar la impresión de ser muy engorroso, en R se encuentra totalmente automatizado en la función auto.arima(), ya previamente abordada. Para ajustarla a un modelo de regresión dinámica se empleará el argumento xreg para añadir las variables regresoras. Sin embargo, de añadirse dos o más variables predictoras, estas han de ser anexionadas a través de la función cbind() ("unión por columnas"). Si se requiere diferenciación, todas las variables se diferenciarán por defecto durante el proceso de estimación aplicando la función auto.arima() (también puede personalizarse de forma manual señalando d=1 y D=1 en la propia función).

Finalmente, resulta conveniente resaltar que en la etapa de predicción será imprescindible pronosticar tanto la parte de regresión como la parte `ARIMA` del modelo, y luego combinar ambos resultados. En particular, primero debemos predecir las variables explicativas. Los pronósticos de estas últimas se pueden estimar por separado, o bien asumiendo una serie de futuros valores para ellas.

Tras haber discutido y expuesto los fundamentos teóricos que respaldan esta clase de modelos, es momento de llevar a cabo un análisis práctico para evaluar el desempeño de los modelos de regresión dinámica a partir de una serie de datos de ventas de Hijos de Rivera.

[1] La regresión espuria es una situación en la cual se halla una aparente relación o asociación estadística entre dos variables, cuando realmente dicho vínculo es causado por una tercera variable. Habitúa a darse en contexto donde las variables en estudio están correlacionadas indirectamente mediante una variable de confusión no controlada. Como consecuencia de esto, los modelos contarán con variables predictoras que asemejan ser importantes, pese a que verdaderamente no lo son.

[2] La cointegración de dos variables tiene lugar cuando existe una relación de largo plazo entre ellas. Este hecho implica que las variables van a compartir una tendencia común y que sus desviaciones temporales se mantendrán alrededor de esta tendencia a lo largo plazo. De este modo, cualquier desequilibrio a corto plazo entre ambas, se subsanará en el largo plazo.

4.2.2. Aplicación de los modelos de regresión dinámica a los datos de Hijos de Rivera

Como bien hemos ido subrayando hasta ahora, en virtud de las políticas de confidencialidad de Hijos de Rivera, no podemos revelar los valores históricos de ventas de cerveza ni la ubicación exacta de la delegación de ventas a la que nos referimos. Es por ello, por lo que también estamos incapacitados para registrar a qué zonas geográficas corresponden las distintas covariables que hayan sido elegidas para incluir en los modelos. Mas, en cambio, sí que tenemos permiso para recoger el nombre del canal en el que fueron extraídos los datos de nuestra serie, una información que será de gran utilidad en este caso.

Concretamente, los datos temporales que modelaremos pertenecen a una delegación de ventas C, que se sitúa encuadrada dentro del canal HORECA. Aunque en la presente memoria solamente ejemplificaremos el ajuste del modelo de regresión dinámica a una serie, hemos de destacar que, en realidad, fue aplicado a las 30 delegaciones del canal HORECA de las que disponemos[3]. No obstante, no fue implementado para ninguna de las delegaciones de los otros canales. La justificación detrás de esto es la siguiente: mientras que en el canal HORECA se documenta la comercialización de productos y la fecha de entrega al propio establecimiento minorista donde serán vendidos, en el resto de canales exclusivamente se contabilizan las ventas al proveedor local o mayorista. Fruto de esta coyuntura, en los segundos puede acontecer un lapso de tiempo demasiado prolongado desde que los artículos salen de los almacenes de Hijos de Rivera hasta su venta al por menor para el consumo final. Por el contrario, en el canal HORECA este intervalo temporal se acorta bastante, lo que posibilita medir de manera casi inmediata el posible impacto que otras variables puedan ocasionar sobre las ventas de cerveza.

Exhibido el canal donde se ejecutarán los modelos de regresión dinámica, debemos esclarecer las variables exógenas que se incorporarán como regresoras en esos modelos. Después de un estudio exhaustivo, donde se cotejó la incidencia y la correlación de cada una de las potenciales variables regresoras con la variable principal, se llegó a la conclusión de que las únicas que se requerían integrar son: "temperatura media" y "días de lluvia". Tanto "pernoctaciones hoteleras" como "días laborables" fueron desechadas por presentar unos niveles muy bajos de significación en la mayoría de los modelos ajustados para cada delegación. Por otra banda, en la disyuntiva de escoger entre "días de lluvia" y "precipitaciones acumuladas", tras probar a añadir cada una de ellas al modelo por separado junto a "temperatura media", se obtuvo que "días de lluvia" era excluible del modelo un menor número de veces que "precipitaciones acumuladas". Asimismo, también se atisbaba una correlación superior de "dias de lluvia" con "ventas de cerveza" para la mayoría de delegaciones.

Remarcado lo anterior, pasamos a ilustrar cómo se efectuaría el proceso de ajuste de un modelo de regresión para una determinada serie. Del mismo modo que para el resto de modelos, empezamos construyendo el gráfico secuencial de la serie de ventas. Recordemos que los valores que padecieron el "efecto covid" ya han sido corregidos para enmendar los problemas de incumplimiento de la normalidad en los residuos y de infraestimación en las predicciones. Así pues, la serie de partida aparece reflejada en la Figura 4.2.

A continuación, tomamos en consideración las dos nuevas series temporales con las que trabajaremos. Para visualizar su evolución hemos creado gráficos secuenciales para cada una de ellas junto a el gráfico asociado a la serie logarítmica de ventas. De acuerdo con lo evidenciado en la Figura 4.3, las series de temperatura media y de días de lluvia no necesitan ser transformadas logarítmicamente, ya que se considera que cumplen la hipótesis de homocedasticidad. Sin embargo, resulta muy palpable que tienen que ser sometidas a una diferenciación estacional, al igual que la serie principal. Además, también sufrirán una diferenciación regular, puesto que la serie de ventas así lo precisa (recordemos que con que una sola serie demande ser diferenciada, lo más idóneo es diferenciar absolutamente todas).

[3]Siempre se modelaron las series de ventas modificadas por la función `tso()`, nunca las originales

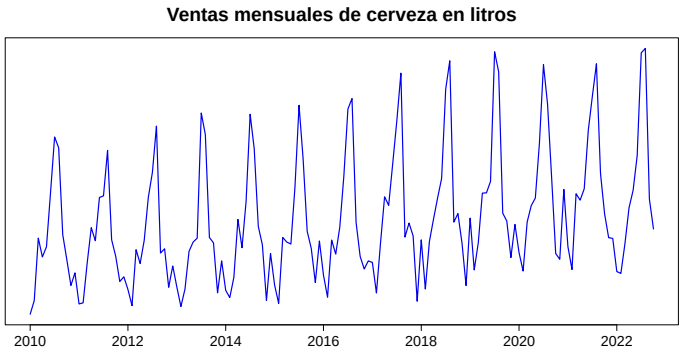

Figura 4.3: Gráfico secuencial de la serie corregida de ventas de la delegación C.

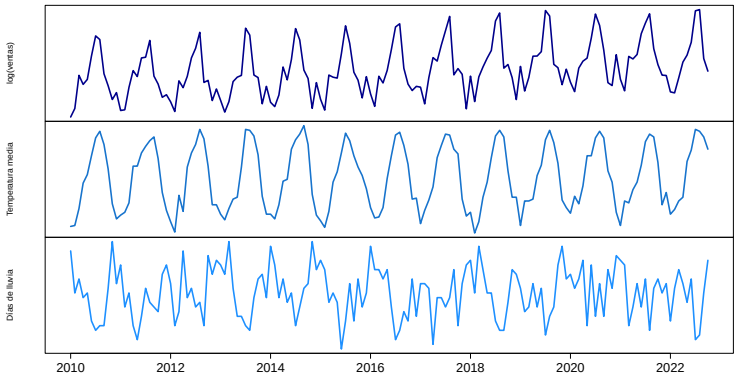

Figura 4.4: Gráficos secuenciales de la serie logarítmica de ventas, la serie de temperatura media y la serie de días de lluvia.

Para ratificar que ambas series climatológicas están correlacionadas con la serie de ventas, plasmamos los diagramas de dispersión relativos a cada una ("ventas de cerveza" frente a "temperatura media" y "ventas de cerveza" frente a "días de lluvia"), así como sus respectivas rectas de regresión lineal[4] (Figura 4.5). Como cabría esperar, advertimos que la relación existente entre las ventas y la temperatura es positiva, mientras que la de las ventas y días de lluvia es negativa. Adicionalmente, también cabe remarcar que se da un mayor grado de relación lineal para el caso de "ventas de cerveza" versus "temperatura media", pues los puntos de su diagrama de dispersión se sitúan más cercanos a la recta de regresión lineal. De todas maneras, ambas gráficas exhiben un significativo grado de correlación entre las variables.

Finalizada la puesta en escena, nos centramos en lo fundamental: ajustar el modelo de regresión dinámica más adecuado. Tal y como se ha mencionado antes, utilizando la función `auto.arima()`, el proceso de elección y estimación de los parámetros del modelo es automático. La salida que se genera se encuentra reflejada en la Figura 4.6.

[4]Resulta conveniente aclarar que para este caso, las rectas de regresión lineal solamente se agregan de forma ilustrativa, pues, como bien hemos recalcado previamente, en esta clase de modelos no se satisface la hipótesis de independencia de los residuos de la regresión.

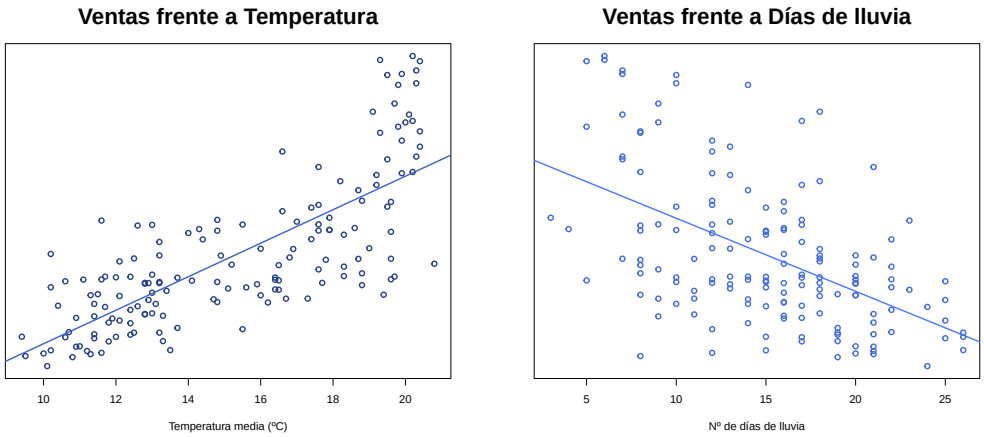

Figura 4.5: Diagramas de dispersión y rectas de regresión lineal de las ventas frente a la temperatura y de las ventas frente a los días de lluvia.

```
Series: serie
Regression with ARIMA(2,1,3)(0,1,1)[12] errors
Box Cox transformation: lambda= 0

Coefficients:
           ar1      ar2      ma1     ma2      ma3     sma1  lluvia  temperatura
      -0.9495  -0.8349  -0.2835  0.3143  -0.5595  -0.8489  -0.003       0.0090
s.e.   0.0958   0.0768   0.1312  0.1503   0.1037   0.1193   0.001       0.0048

sigma^2 = 0.003463:  log likelihood = 195.29
AIC=-372.58   AICc=-371.21   BIC=-346.04
```

Figura 4.6: Valores estimados de los parámetros del modelo de regresión dinámica ajustado.

Una vez interpretada la salida obtenida, lo primero que hacemos es nomenclaturar las variables que forman parte del modelo a estimar:

- Y_t: "logaritmo de las ventas de cerveza"
- T_t: "temperatura media"
- L_t: "días de lluvia"

Posteriormente, expresamos el modelo de `regresión con errores ARIMA(2,1,3)`$\times$`(0,1,1)`$_{12}$ ajustado en notación matématica (sin reemplazar los coeficientes, debido a que todavía no hemos comprobado si alguno puede ser suprimido del modelo):

$$Y_t = \beta_1\, T_t + \beta_2\, L_t + \eta_n\,;$$
$$\eta_n = (1+\phi_1)\, Y_{t-1} - (\phi_1 - \phi_2)\, Y_{t-2} - \phi_2\, Y_{t-3} + Y_{t-12} - (1+\phi_1+\phi_2)\, Y_{t-13} + a_t + \theta_1\, a_{t-1} + \theta_2\, a_{t-2} +$$
$$+ \theta_3\, a_{t-3} + \Theta_1\, a_{t-12} + \theta_1\, \Theta_1\, a_{t-13} + \theta_2\, \Theta_1\, a_{t-14} + \theta_3\, \Theta_1\, a_{t-15}$$

Tras acreditar que efectivamente todos los parámetros incluidos en el modelo resultan significativamente distintos de 0 a un nivel de significación del 5 % (no es necesario prescindir de ninguno), sustituimos dichas estimaciones en el modelo precedente:

$$
\begin{aligned}
Y_t &= 0{,}009\, T_t - 0{,}003\, L_t + \eta_n\,; \\
\eta_n &= 0{,}0505\, Y_{t-1} + 0{,}1146\, Y_{t-2} + 0{,}8349\, Y_{t-3} + Y_{t-12} + 0{,}7844\, Y_{t-13} + a_t - 0{,}2825\, a_{t-1} + 0{,}3143\, a_{t-2} + \\
&\quad - 0{,}5595\, a_{t-3} - 0{,}8489\, a_{t-12} + 0{,}2407\, a_{t-13} - 0{,}2668\, a_{t-14} + 0{,}4796\, a_{t-15}
\end{aligned}
$$

Para que el modelo sea válido es suficiente con que se compruebe que los residuos del modelo `SARIMA`, a_t, siguen un proceso ruido blanco, puesto que se permite que los residuos de la regresión, η_t, contengan autocorrelación. Atendiendo a los resultados del contraste de Ljung-Box (para el cual se obtuvo un p-valor = 0,564) y de los gráficos de la Figura 4.7 confirmamos que se corroboran las condiciones de independencia e incorrelación de los residuos de la parte `ARIMA`. De forma análoga, constatamos que no se rechaza la hipótesis de media nula, por lo que ya podemos garantizar la validez del modelo de `regresión con errores ARIMA(2,1,3)`$\times$`(0,1,1)`$_{12}$.

Seguidamente, contrastamos la hipótesis de normalidad de los residuos a_t mediante las pruebas de Shapiro-Wilk y Jarque-Bera. Para ambos test se obtienen p-valores por encima de 0.05, de forma que también se verifica la normalidad de los residuos de la parte `ARIMA` y, por consiguiente, la gaussianidad de sus innovaciones.

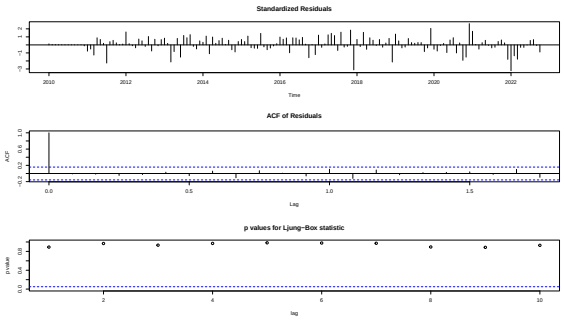

Figura 4.7: Comprobación de la hipótesis de independencia de los residuos de la parte del modelo `ARIMA(2,1,3)`$\times$`(0,1,1)`$_{12}$.

Por último, tenemos que estimar las predicciones de los futuros valores de la serie a partir de la parte `ARIMA` del modelo y de la parte de regresión del modelo. Teniendo en cuenta que no se divisa una tendencia o patrón en el historial de temperaturas y días de lluvia (Figura 4.4), para realizar los pronósticos de la serie de ventas, primero estratificamos por mes los valores medios de temperatura y de días de lluvia habidos en cada uno de los meses de los últimos 13 años. Por tanto, obtendremos un conjunto de valores medios de temperatura y días de lluvia para cada mes del año, asumiendo que serán los que se darán en el año 2023. Con este método, al utilizar los valores medios de temperatura y días de lluvia estratificados mes a mes como base para los pronósticos de ventas en 2023, hallaremos predicciones más precisas si se presentan valores "habituales" de temperatura y de días de lluvias en cada mes de 2023.

A diferencia de lo que sucedió en los modelos previamente aplicados, ahora sí tenemos normalidad en las innovaciones, por lo que construimos los intervalos de predicción por la vía asintótica (computacionalmente más eficiente). En el cuadro 4.1 y la Figura 4.8 se pueden visualizar los pronósticos e intervalos de predicción estimados a partir del modelo anterior. Además, también se recogen las principales medidas de adecuación de esas predicciones en la Figura 4.9.

Fechas	Predicción	Inf.80	Sup.80	Inf.95	Sup.95
Nov 2022	338000	313384	364550	301087	379438
Dec 2022	343159	317526	370860	304741	386419
Jan 2023	328201	303218	355243	290773	370448
Feb 2023	305227	281321	331164	269434	345775
Mar 2023	361396	332805	392443	318597	409944
Apr 2023	396770	364905	431418	349086	450968
May 2023	411442	377845	448026	361185	468692
Jun 2023	461310	423506	502488	404764	525755
Jul 2023	571157	523778	622821	500310	652036
Aug 2023	578526	530024	631465	506018	661424
Sep 2023	403243	369406	440179	352659	461082
Oct 2023	373568	341906	408162	326247	427753
Nov 2023	338186	308632	370570	294047	388950
Dec 2023	350939	320247	384573	305102	403663

Cuadro 4.1: Predicciones e intervalos de confianza obtenidos (tanto al 80 % como al 95 %) por la vía asintótica para la serie de ventas modificada de la delegación C.

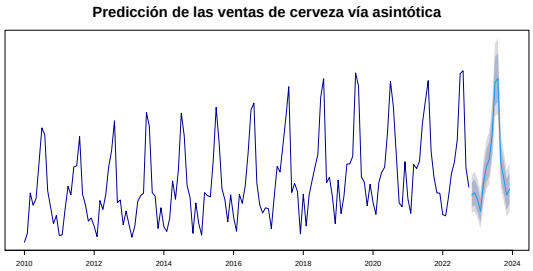

Figura 4.8: Gráfico secuencial de la serie acompañado de las predicciones y los intervalos de confianza estimados (tanto al 80 % como al 95 %) vía asintótica para los 14 próximos meses.

```
                    ME      RMSE      MAE        MPE     MAPE      MASE         ACF1
Training set 3403.86 30747.15 23096.9 0.2677207 4.093331 0.6724368 -0.02464499
```

Figura 4.9: Medidas de adecuación de las predicciones halladas.

A la luz de estas medidas de adecuación de las predicciones computadas, se concluye que la capacidad predictiva del modelo ajustado es notable. Si prestamos atención al `MAPE` (una medida adimensional), vemos que este adquiere un valor inferior al de un 5 %. Así pues, tras la corrección del "efecto covid" y el manejo de modelos más complejos que dan la posibilidad de incorporar variables regresoras, aparentemente hemos logrado obtener unas predicciones más fiables, además de superar los problemas pasados de infraestimación en los pronósticos.

4.3. Modelos de series temporales jerárquicas

Pese a que las predicciones halladas a través de los modelos de regresión dinámica demostraron ser bastante precisas, no tenemos la posibilidad de extender el uso de esta clase de modelos al resto de canales de ventas de la empresa por los motivos ya señalados. En consecuencia, surge aquí la necesidad de gozar de nuevos modelos que sean adecuados para modelizar los otros canales.

Dado que la series de las distintas delegaciones de cada canal comparten un patrón de comportamiento más o menos similar, puede resultar oportuno mantener una estructura anidada entre las distintas series y buscar un método de predicción común para todos los canales. Es por ello, por lo que apostamos por recurrir a los modelos de series jerárquicas, los cuales detallaremos seguidamente, fundamentándonos en la literatura elaborada por Hyndman junto a otros autores (Hyndman *et al.* (2011) y Hyndman *et al.* (2016)).

4.3.1. Base teórica

Dentro de las principales fortalezas de los modelos de series jerárquicas, indiscutiblemente la más diferencial es su capacidad para abordar las interrelaciones y dependencias que se dan entre los diversos niveles de agregación de una serie temporal. Por ende, estos modelos son utilizados cuando se pretende capturar las peculiaridades de los subgrupos individuales, pero manteniendo al mismo tiempo la coherencia con las tendencias generales de la serie en su conjunto. Asimismo, este hecho propiciará la consecución de unos pronósticos más fiables para los diferentes niveles de granularidad existentes en los datos.

Con la finalidad de definir y determinar cómo es la disposición de una serie temporal jerárquica, tomaremos como ejemplo la Figura 4.10, en la que se muestra una estructura desagregada en $K = 2$ niveles. En la parte superior de la jerarquía, la denominada nivel 0, se encuentra siempre el "Total", es decir, el nivel más agregado del conjunto de datos. Posteriormente, este "Total", se desagrega en tres series, que a su vez, se subdividirán en cuatro, dos y tres series, respectivamente, en el nivel inferior de la jerarquía.

Matemáticamente, denotaremos por x_t a la observación del nivel $K = 0$. Continuando con la notación anterior, para las de debajo (niveles $K = 1$ y $K = 2$) emplearemos $x_{i,t}$ para designar a la observación del nodo i en el instante t. Asimismo, simbolizaremos con la letra n al número total de series de la jerarquía ($n = 13$ en nuestro ejemplo) y con la letra m al conjunto de series del nivel inferior ($m = 9$ para nuestro caso).

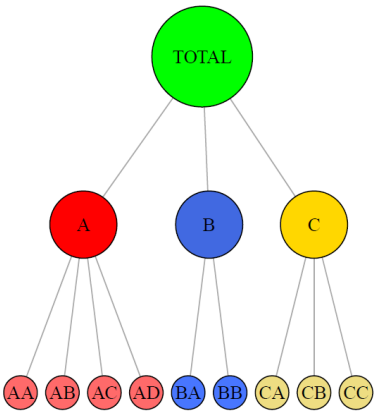

Figura 4.10: Diagrama de árbol con $K = 2$ niveles.

Fruto de la distribución anidada en la que se encuadran todas las observaciones, para todo instante t, las observaciones del nivel superior serán el resultado del sumatorio de las observaciones habidas por niveles:

$$x_t = x_{A,t} + x_{B,t} + x_{C,t} x_t = (x_{AA,t} + x_{AB,t} + x_{AC,t} + x_{AD,t}) + (x_{BA,t} + x_{BB,t}) + (x_{CA,t} + x_{CB,t} + x_{CC,t})$$

Equivalentemente, si consideramos a estas ecuaciones como restricciones de agregación o igualdades de suma, las relaciones anteriores se expresarían matricialmente así:

$$\begin{pmatrix} x_t \\ x_{A,t} \\ x_{B,t} \\ x_{C,t} \\ x_{AA,t} \\ x_{AB,t} \\ x_{AC,t} \\ x_{AD,t} \\ x_{BA,t} \\ x_{BB,t} \\ x_{CA,t} \\ x_{CB,t} \\ x_{CC,t} \end{pmatrix} = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} x_{AA,t} \\ x_{AB,t} \\ x_{AC,t} \\ x_{AD,t} \\ x_{BA,t} \\ x_{BB,t} \\ x_{CA,t} \\ x_{CB,t} \\ x_{CC,t} \end{pmatrix}$$

Análogamente, en notación más compacta:

$$x_t = S \cdot b_t$$

donde x_t es el vector n-dimensional de todas las observaciones jerárquicas en el instante t, S la matriz $n \times m$ que define la relación entre cada una de las series de los diferentes niveles y b_t el vector m-dimensional de todas las observaciones en el nivel inferior de la jerarquía en el instante t.

A diferencia de las series temporales jerárquicas, la series agrupadas no imponen una estructura de agrupación única, sino que se pueden desagregar en más de una vía. Esto es debido a que en ellas la estructura no se desagrega de manera natural en una única forma jerárquica, permaneciendo, a menudo, los factores de desagregación anidados y cruzados.

Una ilustración de una serie agrupada es la que aparece capturada en la Figura 4.10, donde se plasma una estructura de $K = 2$ niveles.

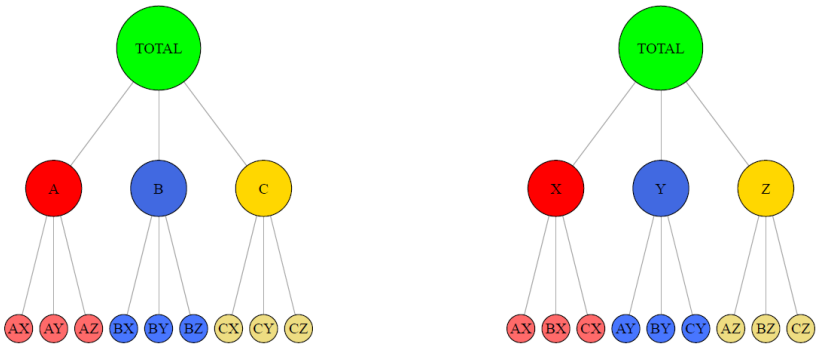

Figura 4.11: Representaciones alternativas de una estructura agrupada de $K = 2$ niveles.

En la cima vuelve a ubicarse el "Total", nuevamente representado por x_t. No obstante, ahora este último se puede desagregar por los atributos (A, B, C) (diagrama de árbol de la izquierda), o bien por los atributos (X, Y, Z) (diagrama de árbol de la derecha). Finalmente, en el nivel inferior, los datos se sitúan desagregados por ambos atributos.

Del mismo modo que con la estructura jerárquica, las observaciones del nivel superior para cualquier momento t serán producto de la agregación de las del nivel más bajo:

$$x_t = x_{AX,t} + x_{AY,t} + x_{AZ,t} + x_{BX,t} + x_{BY,t} + x_{BZ,t} + x_{CX,t} + x_{CY,t} + x_{CZ,t}$$

Sin embargo, para el primer nivel de la estructura tendremos que:

$$\begin{aligned} x_{A,t} &= x_{AX,t} + x_{AY,t} + x_{AZ,t} \\ x_{B,t} &= x_{BX,t} + x_{BY,t} + x_{BZ,t} \\ x_{C,t} &= x_{CX,t} + x_{CY,t} + x_{CZ,t} \end{aligned}$$

pero alternativamente, también nos encontraremos con que:

$$\begin{aligned} x_{X,t} &= x_{AX,t} + x_{BX,t} + x_{CX,t} \\ x_{Y,t} &= x_{AY,t} + x_{BY,t} + x_{CY,t} \\ x_{Z,t} &= x_{AZ,t} + x_{BZ,t} + x_{CZ,t} \end{aligned}$$

Así pues, se constata la existencia de más de una ruta de agregación válida para estructuras agrupadas. Por otro lado, si las planteamos en forma matricial, se obtendría lo siguiente:

$$\begin{pmatrix} x_t \\ x_{A,t} \\ x_{B,t} \\ x_{C,t} \\ x_{X,t} \\ x_{Y,t} \\ x_{Z,t} \\ x_{AX,t} \\ x_{AY,t} \\ x_{AZ,t} \\ x_{BX,t} \\ x_{BY,t} \\ x_{BZ,t} \\ x_{CX,t} \\ x_{CY,t} \\ x_{CZ,t} \end{pmatrix} = \begin{pmatrix} 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \\ 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 & 0 & 0 & 1 \\ 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 \end{pmatrix} \cdot \begin{pmatrix} x_{AX,t} \\ x_{AY,t} \\ x_{AZ,t} \\ x_{BX,t} \\ x_{BY,t} \\ x_{BZ,t} \\ x_{CX,t} \\ x_{CY,t} \\ x_{CZ,t} \end{pmatrix}$$

Dentro de R, tanto para crear series jerárquicas como agrupadas, acudiremos a la función hts() de su paquete homónimo (Hyndman *et al.* (2021)). Con el objetivo de asegurar su correcto funcionamiento, se precisará la colección de series que conforman el nivel inferior. Asimismo, también se deberá delimitar la estructura de agregación de nuestras series a través de los argumentos nodes o characters que incorpora la propia función.

4.3.2. Clasificación de los enfoques de predicción de series jerárquicas

Expuesta la estructura de las series con las que operaremos, pasamos a desglosar los enfoques esenciales para estimar predicciones congruentes sobre las mismas. Por tanto, aquí el reto más importante será recabar unos pronósticos que se sumen de manera consistente en la estructura de agregación de las series.

Al margen de los diferentes métodos de pronóstico vigentes, es relevante establecer el modelo que mejor se adecua a los datos. Dicho procedimiento se automatiza en R aplicando la función forecast.gts(). En concreto, se requerirá ajustar el argumento fmethod, el cual nos permite seleccionar tres clases de modelos. No obstante, dada la tipología de nuestras series objeto de estudio, únicamente nos serviremos de dos de ellos. Eston son: los modelos Box-Jenkins (fmethod="arima") y los modelos de suavizado exponencial (fmethod="ets"). Ambos ya han sido singularizados con anterioridad.

- **Enfoque *bottom-up*:**

El enfoque *bottom-up* ("enfoque de abajo hacia arriba"), tal y como su nombre indica, se basa en generar predicciones para las series del nivel más bajo y luego ir añadiéndolas en los niveles superiores para hallar pronósticos de todas las series de la estructura jerárquica (Hyndman *et al.* (2017)). Se trata del método más asiduo en la realidad, ya que al centrarse en el modelaje de las series más específicas, se detectan mejor las características intrínsecas a cada serie y no se pierde tanta información durante la agregación. Lógicamente, por medio de este enfoque se conseguirán pronósticos más precisos para los niveles inferiores; mas, puede ofrecer dificultades si el historial de los datos es limitado y si el espectro temporal de predicción es muy a largo plazo. Añadido a esto, el enfoque *bottom-up* también restringe la elección de modelos de *forecast*.

Tomando como base la estructura jerárquica que hemos dibujado en la Figura 4.11, estimaremos las predicciones de las observaciones con origen T y horizonte h: $\hat{x}_{AA,T}$, $\hat{x}_{AB,T}$, $\hat{x}_{AC,T}$, $\hat{x}_{AD,T}$, $\hat{x}_{BA,T}$, $\hat{x}_{BB,T}$, $\hat{x}_{CA,T}$, $\hat{x}_{CB,T}$ y $\hat{x}_{CC,T}$. De este modo, para el nivel $K = 1$ tendríamos:

$$\begin{aligned}
\hat{x}_{A,T}(h) &= \hat{x}_{AA,T} + \hat{x}_{AB,T} + \hat{x}_{AC,T} + \hat{x}_{AD,T} \\
\hat{x}_{B,T}(h) &= \hat{x}_{BA,T} + \hat{x}_{BB,T} \\
\hat{x}_{C,T}(h) &= \hat{x}_{CA,T} + \hat{x}_{CB,T} + \hat{x}_{CC,T}
\end{aligned}$$

Análogamente, en forma matricial:

$$\begin{pmatrix}
\hat{x}_T(h) \\ \hat{x}_{A,T}(h) \\ \hat{x}_{B,T}(h) \\ \hat{x}_{C,T}(h) \\ \hat{x}_{AA,T}(h) \\ \hat{x}_{AB,T}(h) \\ \hat{x}_{AC,T}(h) \\ \hat{x}_{AD,T}(h) \\ \hat{x}_{BA,T}(h) \\ \hat{x}_{BB,T}(h) \\ \hat{x}_{CA,T}(h) \\ \hat{x}_{CB,T}(h) \\ \hat{x}_{CC,T}(h)
\end{pmatrix} =
\begin{pmatrix}
1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \\
1 & 1 & 1 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 1 & 1 \\
1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 1 & 0 \\
0 & 0 & 0 & 0 & 0 & 0 & 0 & 0 & 1
\end{pmatrix} \cdot
\begin{pmatrix}
\hat{x}_{AA,T}(h) \\ \hat{x}_{AB,T}(h) \\ \hat{x}_{AC,T}(h) \\ \hat{x}_{AD,T}(h) \\ \hat{x}_{BA,T}(h) \\ \hat{x}_{BB,T}(h) \\ \hat{x}_{CA,T}(h) \\ \hat{x}_{CB,T}(h) \\ \hat{x}_{CC,T}(h)
\end{pmatrix}$$

o también escrito como

$$\hat{x}_T(h) = S \cdot \hat{b}_T(h)$$

donde $\hat{x}_T(h)$ es el vector n-dimensional de predicción para un horizonte h y $\hat{b}_T(h)$ es el vector m-dimensional de predicción de las series del nivel inferior para un horizonte h.

En R para implementar este enfoque explicitaremos dentro de la función forecast.gts() el argumento method="bu".

- **Enfoque *top-down*:**

En contraposición al enfoque *bottom-up*, se coloca el enfoque *top-down* ("enfoque de arriba hacia abajo"), cimentado en efectuar un único pronóstico para la serie del nivel superior, el cual posteriormente se va desagregando en los niveles inferiores (Gross y Sohl (1990)). Dada su metodología de aplicación, este enfoque solamente puede ser llevado a la práctica cuando la estructura de las series es estrictamente jerárquica, quedando excluidas, así pues, las estructuras agrupadas. Tanto las virtudes como las debilidades de este enfoque son totalmente opuestas a las del *bottom-up*. Evidentemente la confiabilidad de los pronósticos desciende conforme descendamos de nivel. Su mayor hándicap se da cuando la series de niveles inferiores no conservan el mismo patrón de comportamiento o tendencia que la serie del nivel $K = 0$, puesto que, para este caso, estaríamos asumiendo suposiciones que realmente no son ciertas.

Sea $p = (p_1, .., p_m)$ el vector de proporciones de desagregación que dicta cómo se repartirán los pronósticos en el nivel inferior; de nuevo en relación a la Figura 4.10, postulamos que:

$$\begin{aligned}
\hat{x}_{AA,T}(h) &= p_1\hat{x}_T(h)\\
\hat{x}_{AB,T}(h) &= p_2\hat{x}_T(h)\\
\hat{x}_{AC,T}(h) &= p_3\hat{x}_T(h)\\
\hat{x}_{AD,T}(h) &= p_4\hat{x}_T(h)\\
\hat{x}_{BA,T}(h) &= p_5\hat{x}_T(h)\\
\hat{x}_{BB,T}(h) &= p_6\hat{x}_T(h)\\
\hat{x}_{CA,T}(h) &= p_7\hat{x}_T(h)\\
\hat{x}_{CB,T}(h) &= p_8\hat{x}_T(h)\\
\hat{x}_{CC,T}(h) &= p_9\hat{x}_T(h)
\end{aligned}$$

En notación matricial se reduciría a:

$$\hat{x}_T(h) = \ S \cdot p \cdot \hat{x}_T(h)$$

Empero, cabe enfatizar que existen tres técnicas distintas para fijar los pesos que se le asignan a cada serie. Estos son:

  - **Proporciones históricas medias:**

Constituye el método más simple de los tres. Consiste en calcular cuán proporción del total corresponde a cada una de las series en cada instante t:

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

para $i = 1, ..., m$ y $t = 1, ..., T$.

En R este método se configura dentro de la función forecast.gts() con el argumento method="tdgsa".

– **Proporciones de las medias históricas:**

Se calcula la media histórica de cada serie, observándose qué proporción supone sobre la media histórica de la serie total.

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

para $j = 1, ..., m$

En R este método se configura dentro de la función forecast.gts() con el argumento method="tdgsf".

– **Proporciones estimadas:**

Por causa de que las dos técnicas anteriores no evalúan cómo esas proporciones puedes transmutar con el tiempo, estas tienden a producir predicciones menos precisas en los niveles inferiores de la jerarquía que las de los enfoques *bottom-up*. Como resultado de esta situación, se propone un método basado en los pronósticos en lugar de los datos históricos.

La explicación de mencionado método es sencilla. Primeramente, se calculan las predicciones con horizonte h para todas las series del primer nivel de la jerarquía, pero individualmente una a una, puesto que su agregación no sería coherente. Con estos pronósticos, denominados "predicciones iniciales", averiguamos la proporción que suponen sobre el total agregado de las predicciones iniciales de ese nivel ("proporciones estimadas"). En última instancia, manejamos estas proporciones para desagregar las estimaciones iniciales en sus respectivos subniveles y obtener así predicciones para la integridad de la jerarquía.

Podemos definir la proporción estimada de la serie j-ésima del nivel K como

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

donde $j = 1, ..., m$, $\hat{x}_{j,h}^{(\ell)}$ es la predicción inicial con horizonte h de la serie que está ℓ niveles por encima de j y $\hat{x}_{j,h}^{(\ell)}$ 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.

En R este método se configura dentro de la función forecast.gts() con el argumento method="tdfp".

- **Enfoque *middle-out*:**

El enfoque *middle-out* combina técnicas de los dos enfoques previos. En él se comienza escogiendo un "nivel medio", sobre el que se calculan las predicciones iniciales de sus series. Inmediatamente después, se hallan predicciones mediante el enfoque *bottom-up* para las series de niveles superiores y mediante el enfoque *top-down* para las series de niveles inferiores (Hyndman *et al.* (2009)).

En R este método se configura dentro de la función forecast.gts() con el argumento method="mo', pudiéndose elegir el nivel medio deseado a través del argumento level. Aún así, si no se especifica el level, la función empleará por defecto el método de proporciones estimadas bajo el enfoque *top-down*.

Tras haber presentado y desmenuzado minuciosamente los fundamentos teóricos por los que se rigen los modelos de series jerárquicas, es el momento de poner en acción todo lo estudiado sobre nuestras series de datos reales.

4.4. Aplicación de los modelos de series jerárquicas a los datos de Hijos de Rivera

Inicialmente, recordemos que para las diferentes delegaciones de ventas del canal HORECA ya fueron predichos los futuros valores de cada una de las series utilizando los modelos de regresión dinámica. Es por ello, por lo que nos vamos a focalizar en el análisis de las series del resto de canales de venta de Hijos de Rivera.

Igualmente que en los otros capítulos y secciones, se ocultarán los verdaderos valores de los pronósticos obtenidos. Mas, adicionalmente para este caso, tampoco estamos autorizados a mencionar explícitamente el nombre del canal al que hace referencia cada gráfico secuencial, debido a que ahora sí será ilustrada gráficamente la serie global de cada canal (con la salvedad de HORECA, por supuesto) y no solo la de una determinada delegación de ventas. Asimismo, es preciso subrayar que las series de partida serán las alteradas por la función `tso()`, con la finalidad de que el factor coronavirus no distorsione ni afecte en demasía a nuestros pronósticos. Por otra parte, el lapso temporal que comprenden las series va desde enero de 2014 hasta a octubre de 2022 (fue necesario acotar más la amplitud de los intervalos al no disponerse de datos anteriores a 2014 para algunos canales).

Esclarecido esto, comenzamos delimitando la estructura jerárquica bajo la que se dispondrán las series de tiempo. En particular, la estructura que hemos implementado cuenta con $K = 2$ niveles:

- Nivel $K = 0$: Es el nivel máximo de agregación, en el cual Figuran los litros totales de cerveza que vende la compañía, exceptuando lo comercializado en el canal HORECA.
- Nivel $K = 1$: En el primer nivel de la jerarquía se sitúan los litros de cerveza vendidos por los distintos canales de ventas de la compañía. Estos son un total de 5.
- Nivel $K = 2$: En el segundo nivel se recogen los litros vendidos por cada una de las delegaciones de ventas de cada canal. Hay un conjunto de 17 delegaciones.

La representación gráfica de las estructura descrita se puede visualizar en el diagrama de árbol reflejado en la Figura 4.12:

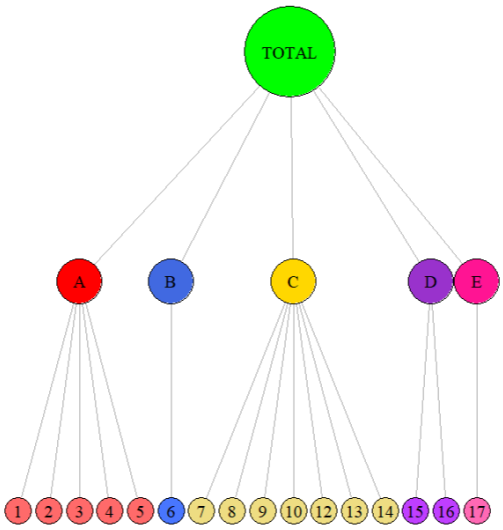

Figura 4.12: Diagrama de árbol de las ventas de cerveza en relación al canal y la delegación de ventas correspondiente.

A la vista del gráfico, se evidencia como las ventas han sido divididas primeramente por canal y luego subdivididas por delegación de ventas. Cada uno de los canales se denota siguiendo un orden alfabético, mientras que cada una de las delegaciones mediante un orden numérico de números cardinales. También se atestigua que la estructura representada es la única agrupación posible, confirmándose que estamos ante una estructura jerárquica pura.

Enmarcada la manera en la que se distribuyen las series temporales que conforman la estructura, hemos de seleccionar tanto la clase de modelos que mejor se ajustan a las observaciones como el enfoque predictivo más pertinente.

Tras cotejar el rendimiento de los modelos Box-Jenkins frente a los modelos de suavizado exponencial en varios enfoques, se percibe que los modelos `ETS` tienden a ofrecer, a priori, pronósticos infraestimados para los meses que conglomeran un mayor número de ventas. Este patrón de comportamiento choca frontalmente con la tendencia creciente que se observa en muchas de las series, especialmente a partir de 2020. Dado que desde la empresa se espera que las ventas prosigan aumentando durante 2023, podemos inferir que los modelos `SARIMA` nos brindan unas predicciones más fiables y creíbles.

En lo tocante a los enfoques de predicción, tenemos que destacar que, en general, no hay grandes divergencias en los resultados derivados de uno u otro enfoque. Al no tener la posibilidad de acceder a las medidas de adecuación de las predicciones (la función `forecast.gts` no las computa), resulta muy complicado ser capaces de determinar el enfoque más idóneo para nuestras series. Con la pretensión de extraer respuestas más concluyentes, graficamos los pronósticos alcanzados a través de algunos métodos estudiados en las Figuras 4.13, 4.14 y 4.15, respectivamente. En ellas, se muestran las predicciones para todos los niveles de agregación prefijados.

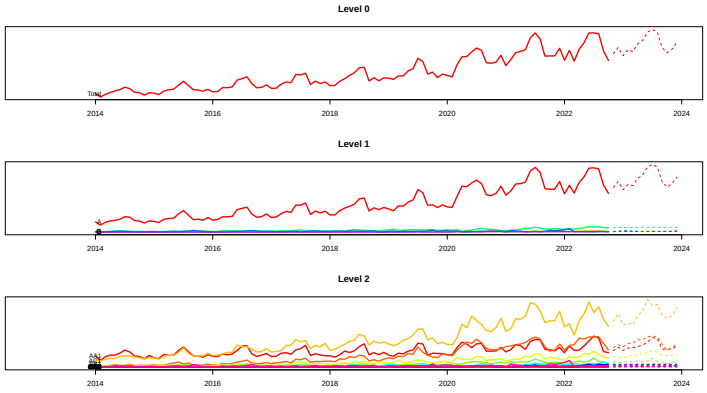

Figura 4.13: Predicciones jerárquicas halladas a través del método *bottom-up*.

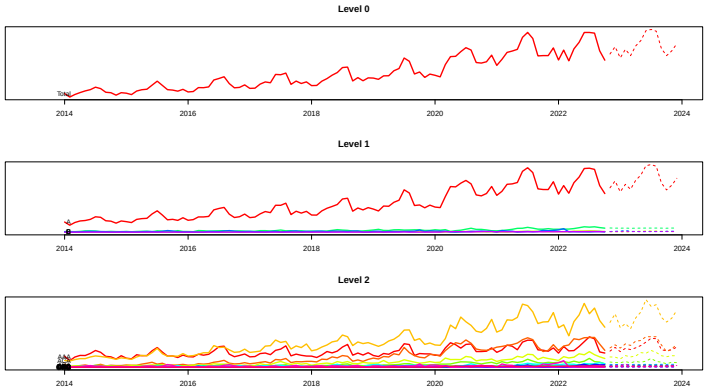

Figura 4.14: Predicciones jerárquicas halladas a través del método *top-down*.

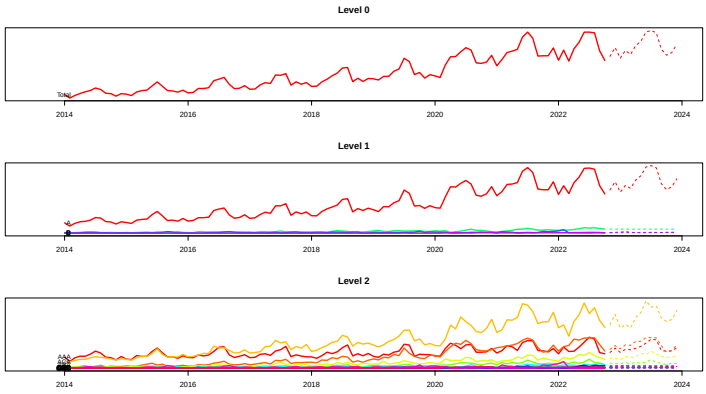

Figura 4.15: Predicciones jerárquicas halladas a través del método *middle-out*.

Después de una exhaustiva reflexión, hemos optado por utilizar el enfoque *bottom-up* en nuestra estructura jerárquica. Esta elección se respalda en la habilidad de este enfoque para adaptarse a las particularidades de cada una de las delegaciones, lo que nos permitirá obtener pronósticos más precisos en los niveles inferiores.

Capítulo 5

Implementación y conclusiones

Puesto que a lo largo de los dos capítulos previos ya hemos ido comentando los resultados específicos que se estimaban para alguna delegación, ahora únicamente nos dedicaremos a extender las conclusiones inducidas para el resto de ellas.

El procedimiento inicial a seguir será igual para todas las delegaciones, pues primeramente todas serán sometidas a la función `tso()` con el objetivo de mitigar la influencia del "efecto covid". En la mayoría de ocasiones, las series verán modificados algunos de sus valores más recientes, tendiendo estos generalmente al alza. Por tanto, pasaremos a trabajar con nuevas series, distintas a las que originalmente teníamos.

Una vez aquí, emplearemos dos clases de modelos para ajustar dichas series: para las correspondientes al canal HORECA manejaremos los modelos de regresión dinámica, mientras que para el resto de canales los modelos de series jerárquicas.

En lo que respecta a los primeros, hemos de destacar que, en general, a través de ellos se lograron conseguir unos pronósticos notablemente precisos, a la vista de las medidas de adecuación de las predicciones computadas. Si atendemos a lo recogido en el apéndice A, observamos que en el aproximadamente 50 % de los modelos ajustados, tanto la variable "temperatura media" como la variable "días de lluvia" resultaron significativas. Sin embargo, también se percibe que el impacto del factor temperatura sobre las ventas de cerveza es sustancialmente superior al del factor días de lluvia, ya que "temperatura media" únicamente no es significativa en 4 modelos, a la vez que "días de lluvia" alcanza a no serlo en 13 modelos. A pesar de lo anterior, ambas variables demuestran estar correlacionadas con las ventas y, generalmente, incrementan la capacidad predictiva de los modelos.

En lo referente a los segundos, volver a comentar que finalmente se apostó por usar el enfoque `bottom-up`, debido a que era el que menos tendía a la infraestimación. Para este caso, quizás los pronósticos obtenidos no sean tan fiables de acuerdo a lo esperado por la empresa, ya que algunos de los canales habían visto aumentar fuertemente sus ventas en estos últimos tres años. Por ende, están constituidos por delegaciones con ventas todavía escasamente asentadas, que además pueden verse fuertemente afectadas por valores atípicos. Asimismo, a esto se le añade la imposibilidad de disponer en estos canales de potenciales variables regresoras.

Dentro de R se implementaron los modelos detallados siguiéndose los pasos descritos. Una vez averiguados los pronósticos para cada delegación, estos se agregaron para obtener los relativos a cada canal (en el caso de los modelos de regresión dinámica de forma manual, mientras que en los otros automatizadamente). Por último, volvieron a ser agrupados para hallar las estimaciones de los litros de cerveza que serán vendidos durante cada mes de este 2023 y también el total agregado anual.

Valorando en perspectiva la aportación del presente trabajo frente a otros *forecasts* ya realizados, tenemos que poner en liza, por encima de todo, el método desarrollado para el ajuste de los modelos al impacto del factor covid. Debido a que la pandemia de la COVID-19 constituye un suceso todavía relativamente reciente, apenas se disponía de bibliografía acerca de cómo debía ser tomada en consideración a la hora de plantear los distintos modelos. Así pues, sabedores del carácter transitorio del mencionado fenómeno, se optó por aplicar una metodología donde se estimase el valor de las observaciones en ese periodo, bajo la suposición de que la pandemia no hubiese tenido lugar. Los resultados obtenidos tras su ejecución fueron bastante satisfactorios, sobre todo dentro del canal HORECA.

Por otro lado, bien es cierto que los modelos utilizados no son manifiestamente novedosos, ya que abundan las fuentes documentales relativas a ellos. Mas, dada la naturaleza de nuestros datos, y una vez ajustado el "efecto covid", resultan sumamente aptos para modelizar las series con las que contamos, a excepción de las concernientes a canales que viven todavía un proceso de expansión en sus ventas. Es quizás ahí, donde pueda estar una de las líneas a mejorar en próximos proyectos, una vez se vayan asentando sus ventas.

Apéndice A

Modelos de regresión dinámica ajustados para HORECA

Cada uno de los 30 modelos ajustados corresponden a una determinada delegación de ventas del canal HORECA. En este anexo se incluyen los valores estimados para los coeficientes β_1 y β_2. Asimismo, también se adjunta la estimación del modelo $\texttt{ARIMA(p,d,q)} \times \texttt{(P,D,Q)}_{12}$ que siguen los residuos (η_n) del modelo de regresión dinámica. β_1 hace referencia a la variable "temperatura media", mientras que β_2 hace alusión a la variable "días de lluvia". Con el asterisco "*" se designan todos los casos en los que los coeficientes β_1 y β_2 no resultaban significativamente distintos de 0 a un nivel de significación del 5 %, es decir, aquellas situaciones donde se podría prescindir de alguno de los coeficientes (o bien de ambos). A su vez, el hecho de que un coeficiente pueda ser excluido del modelo, denotará que la variable en cuestión (sea "temperatura media", "días de lluvia" o las dos) no es relevante para la estimación de las futuras ventas de cerveza de una delegación.

Delegación	$\hat{\beta}_1$	$\hat{\beta}_2$	**Modelo SARIMA de** η_n
1	0,0346	−0,0120*	$\texttt{ARIMA(1,1,2)}\times\texttt{(0,1,1)}_{12}$
2	0,0147	−0,0045	$\texttt{ARIMA(1,0,2)}\times\texttt{(1,1,1)}_{12}$
3	0,0090	−0,0030	$\texttt{ARIMA(2,1,3)}\times\texttt{(0,1,1)}_{12}$
4	0,0318	−0,0055	$\texttt{ARIMA(0,1,2)}\times\texttt{(0,1,2)}_{12}$
5	0,0079	−0,0035*	$\texttt{ARIMA(1,0,0)}\times\texttt{(1,1,1)}_{12}$
6	0,0279	−0,0038	$\texttt{ARIMA(2,1,1)}\times\texttt{(0,0,2)}_{12}$
7	0,0130	−0,0027*	$\texttt{ARIMA(3,0,1)}\times\texttt{(0,1,1)}_{12}$
8	0,0003*	−0,0057	$\texttt{ARIMA(3,1,1)}\times\texttt{(0,1,1)}_{12}$
9	0,0312	−0,0036	$\texttt{ARIMA(0,1,1)}\times\texttt{(0,0,2)}_{12}$
10	0,0157	−0,0004*	$\texttt{ARIMA(0,1,1)}\times\texttt{(0,0,1)}_{12}$
11	0,0276	−0,0082	$\texttt{ARIMA(1,1,1)}\times\texttt{(1,1,0)}_{12}$
12	0,0133	−0,0017*	$\texttt{ARIMA(0,1,2)}\times\texttt{(0,1,1)}_{12}$
13	0,0294	−0,0004*	$\texttt{ARIMA(0,1,4)}\times\texttt{(1,0,1)}_{12}$

Delegación	$\hat{\beta}_1$	$\hat{\beta}_2$	**Modelo SARIMA de** η_n
14	0,0206	−0,0020*	$\texttt{ARIMA(0,1,2)}\times\texttt{(0,1,1)}_{12}$
15	0,0317	−0,0041	$\texttt{ARIMA(0,1,2)}\times\texttt{(0,0,2)}_{12}$
16	0,0259	−0,0035*	$\texttt{ARIMA(2,1,1)}\times\texttt{(0,1,2)}_{12}$
17	0,0244	−0,0018*	$\texttt{ARIMA(1,1,1)}\times\texttt{(2,0,0)}_{12}$
18	0,0080*	−0,0008*	$\texttt{ARIMA(1,1,1)}\times\texttt{(2,1,0)}_{12}$
19	0,0173	−0,0054	$\texttt{ARIMA(2,1,1)}\times\texttt{(1,1,1)}_{12}$
20	0,0377	−0,0018*	$\texttt{ARIMA(3,1,2)}\times\texttt{(1,0,0)}_{12}$
21	0,0042*	−0,0011*	$\texttt{ARIMA(0,1,2)}\times\texttt{(1,0,0)}_{12}$
22	0,0217	−0,0114	$\texttt{ARIMA(3,1,2)}\times\texttt{(2,1,1)}_{12}$
23	0,0318	−0,0038	$\texttt{ARIMA(1,1,1)}\times\texttt{(2,0,1)}_{12}$
24	0,0200	−0,0006*	$\texttt{ARIMA(2,1,2)}\times\texttt{(2,1,0)}_{12}$
25	0,0103*	−0,0027	$\texttt{ARIMA(0,1,1)}\times\texttt{(2,1,0)}_{12}$
26	0,0423	−0,0084	$\texttt{ARIMA(0,1,1)}\times\texttt{(1,0,0)}_{12}$
27	0,0340	−0,0040	$\texttt{ARIMA(0,1,1)}\times\texttt{(0,1,2)}_{12}$
28	0,0314	−0,0077	$\texttt{ARIMA(1,1,2)}\times\texttt{(2,1,0)}_{12}$
29	0,0244	−0,0062	$\texttt{ARIMA(0,1,1)}\times\texttt{(1,0,0)}_{12}$
30	0,0311	−0,0055	$\texttt{ARIMA(0,1,1)}\times\texttt{(0,1,1)}_{12}$

Cuadro A.1: Resumen de los modelos de regresión dinámica ajustados para cada delegación de ventas.

Bibliografía

Bengtsson, H. (2022). Various programming utilities. R package version 2.12.2. https://cran.r-project.org/web/packages/R.utils/R.utils.pdf. Accedido el 28 de abril de 2023.

Box, G. E. y Cox, D. R. (1964). An analysis of transformations. *Journal of the Royal Statistical Society: Series B (Methodological)*, 26(2):211–243.

Box, G. E., Jenkins, G., Reinsel, G. C., y Ljung, G. M. (1994). *Time series Analysis - Forecasting and Control*, capítulo 1, pp. 19–54. Prentice Hall, Nueva Jersey, 3ª edición.

Brockwell, P. J. y Davis, R. A. (2002). *Introduction to time series and forecasting*, capítulo 1, pp. 9–15. Springer, Nueva York, 2ª edición.

Burnham, K. P. y Anderson, D. R. (2004). Multimodel inference: understanding AIC and BIC in model selection. *Sociological methods & research*, 33(2):261–304.

Cerveceros, A. (2022). Informe Socioeconómico del Sector de la Cerveza en España 2022. Technical Report 003230533, Ministerio de Agricultura, Pesca y Alimentación, Madrid. Catálogo de Publicaciones de la Administración General del Estado.

Chatfield, C. (1978). The holt-winters forecasting procedure. *Journal of the Royal Statistical Society: Series C (Applied Statistics)*, 27(3):264–279.

Cowpertwait, P. S. y Metcalfe, A. V. (2009). *Introductory time series with R*, capítulo 1, pp. 1–25. Springer, Nueva York.

Cryer, J. D. y Kung-Sik, C. (2008). *Time Series Analysis With Applications in R*, capítulo 1, p. 1. Springer, Nueva York.

Galián, R. (1983). La función de autocorrelación extendida: su utilización en la construcción de modelos para series temporales económicas. Technical Report 8314, Banco de España, Madrid. Documentos de Trabajo del Banco de España.

Gardner Jr, E. S. (1985). Exponential smoothing: The state of the art. *Journal of forecasting*, 4(1):1–28.

Gardner Jr, E. S. y McKenzie, E. (1985). Forecasting trends in time series. *Management science*, 31(10):1237–1246.

Groeneveld, R. A. y Meeden, G. (1984). Measuring skewness and kurtosis. *Journal of the Royal Statistical Society Series D: The Statistician*, 33(4):391–399.

Gross, C. W. y Sohl, J. E. (1990). Disaggregation methods to expedite product line forecasting. *Journal of forecasting*, 9(3):233–254.

Holt, C. C. (1957). Forecasting seasonals and trends by exponentially weighted moving averages. *International Journal of Forecasting*, 20(1):5–10.

Hyndman, R. y Athanasopoulos, G. (2023). Forecasting functions for time series and linear models. R package version 8.21. https://cran.r-project.org/web/packages/forecast/forecast.pdf. Accedido el 14 de mayo de 2023.

Hyndman, R., Athanasopoulos, G. J., Kourentzes, N., y Petropoulos, F. (2017). Forecasting with temporal hierarchies. *European Journal of Operational Research*, 262(1):60–74.

Hyndman, R., Koehler, A. B., Ord, J. K., y Snyder, R. D. (2008). *Forecasting with exponential smoothing: the state space approach*, capítulo 3, pp. 33–51. Springer Science & Business Media, Nueva Jersey.

Hyndman, R., Lee, A., Wang, E., y Wickramasuriya, S. (2021). Hierarchical and grouped time series. R package version 6.0.2. https://cran.r-project.org/web/packages/hts/hts.pdf. Accedido el 29 de mayo de 2023.

Hyndman, R. J., Ahmed, R. A., Athanasopoulos, G., y Shang, H. L. (2011). Optimal combination forecasts for hierarchical time series. *Computational statistics & data analysis*, 55(9):2579–2589.

Hyndman, R. J. y Athanasopoulos, G. (2018). *Forecasting: principles and practice*, capítulo 9. OTexts, Melbourne, 2ª edición.

Hyndman, R. J., Athanasopoulos, G., y Ahmed, R. A. (2009). Hierarchical forecasts for Australian domestic tourism. *International Journal of Forecasting*, 25(1):146–166.

Hyndman, R. J. y 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., Snyder, R. D., y Grose, S. (2002). A state space framework for automatic forecasting using exponential smoothing methods. *International Journal of forecasting*, 18(3):439–454.

Hyndman, R. J., Lee, A. J., y Wang, E. (2016). Fast computation of reconciled forecasts for hierarchical and grouped time series. *Computational statistics & data analysis*, 97:16–32.

Jarque, C. M. y Bera, A. K. (1980). Efficient tests for normality, homoscedasticity and serial independence of regression residuals. *Economics letters*, 6(3):255–259.

Konishi, S. y Kitagawa, G. (2008). *Information criteria and statistical modeling*, capítulo 3, pp. 29–74. Springer, Nueva York.

Ljung, G. M. y Box, G. E. P. (1978). On a measure of lack of fit in time series models. *Biometrika*, 65(2):297–303.

López de Lacalle, J. (2019). Detection of outliers in time series. R package version 0.6-8. https://cran.r-project.org/web/packages/tsoutliers/tsoutliers.pdf. Accedido el 29 de mayo de 2023.

Makridakis, S., Andersen, A., Carbone, R., Fildes, R., Hibon, M., y Lewandowski, R. (1984). *The Forecasting Accuracy of Major Time Series Methods*, capítulo 5. John Wiley, Nueva York.

Makridakis, S. y Hibon, M. (1997). ARMA models and the Box-Jenkins methodology. *International journal of forecasting*, 16(3):147–163.

Matarise, F., Dhliwayo, L., Nduna, I. S., Maposa, I., y Siziba, L. (2012). Detecting level shifts, temporary changes and innovational outliers in intervention analysis. *International Journal of Statistics and Systems*, 7(3):241–254.

Oons, J., Lang, D. T., y Hilaiel, L. (2023). A simple and robust json parser and generator for R. R package version 1.8.7. https://cran.r-project.org/web/packages/jsonlite/jsonlite.pdf. Accedido el 28 de abril de 2023.

Orus, A. (2022). La industria de la cerveza en España - Datos estadísticos. https://es.statista.com/temas/5410/la-industria-de-la-cerveza-en-espana/. Accedido el 16 de abril de 2023.

Ostertagova, E. y Ostertag, O. (2012). Forecasting using simple exponential smoothing method. *Acta Electrotechnica et Informatica*, 12(3):62.

Pankratz, A. (2012). *Forecasting with dynamic regression models*, capítulo 5. John Wiley & Sons, Nueva Jersey.

Peña, D. (2005). *Análisis de series temporales*, capítulo 1. Alianza Editorial, Madrid.

Phillips, P. C. y Perron, P. (1988). Testing for a unit root in time series regression. *Biometrika*, 75(2):335–346.

R Core Team (2023). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Viena. https://www.R-project.org/.

Shapiro, S. S. y Wilk, M. B. (1965). An analysis of variance test for normality (complete samples). *Biometrika*, 52(3/4):591–611.

Shumway, R. H., Stoffer, D. S., y Stoffer, D. S. (2000). *Time Series Analysis and Its Applications*, capítulo 3, pp. 111–123. Springer, Nueva York.

Trapletti, A., Hornik, K., y LeBaron, B. (2023). Time series analysis and computational finance. R package version 0.10. https://cran.r-project.org/web/packages/tseries/tseries.pdf. Accedido el 14 de mayo de 2023.

Velasco, M. y del Puerto, I. (2009). *Series temporales*, capítulo 5. Servicio de Publicaciones. Universidad de Extremadura, Badajoz.

Winters, P. R. (1960). Forecasting sales by exponentially weighted moving averages. *Management science*, 6(3):324–342.