



Trabajo Fin de Máster

Aplicación de técnicas estadísticas en la Inteligencia de Clientes

Eduardo Martínez Granda

Máster en Técnicas Estadísticas

Curso 2022-2023

Propuesta de Trabajo Fin de Máster

Título en galego: Aplicación de técnicas estatísticas na Intelixencia de Clientes
Título en español: Aplicación de técnicas estadísticas en la Inteligencia de Clientes
English title: Statistical techniques on Customer Intelligence
Modalidad: Modalidad B
Autor: Eduardo Martínez Granda, Universidade de Santiago de Compostela
Director: Ricardo Cao Abad, Universidade da Coruña
Tutora: Marta Pou Bueno, ABANCA
Breve resumen del trabajo: La presente memoria se adecua a las líneas de trabajo del área de Inteligencia de Clientes de ABANCA. Entre ellas se encuentran: la explotación de bases de datos estructuradas y desestructuradas mediante técnicas de minería de texto; la aplicación de técnicas de aprendizaje supervisado para la clasificación de textos y técnicas de <i>web scrapping</i> para la obtención de información externa; y el desarrollo de modelos de recomendación dirigidos a patrones de consumo.
Recomendaciones: Conocimientos básicos de analítica avanzada. Manejo intermedio de herramientas de modelización estadística (R y/o Python) y tratamiento de datos (SQL).

Don Ricardo Cao Abad, Catedrático de Estadística e Investigación Operativa de la Universidade da Coruña y doña Marta Pou Bueno, Manager de Analítica Avanzada (Inteligencia de Clientes) de ABANCA informan que el Trabajo Fin de Máster titulado

Aplicación de técnicas estadísticas en la Inteligencia de Clientes

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

En A Coruña, a 5 de Junio de 2023.

El director:
Don Ricardo Cao Abad



La tutora:
Doña Marta Pou Bueno

El autor:
Don Eduardo Martínez Granda

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 autor 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

En primer lugar, quería agradecer a quien siempre me decía que con trabajo y esfuerzo el camino sería más fácil. Espero que esté orgulloso de mi trayectoria y del esfuerzo realizado.

Quería agradecer a Marta Pou su asesoramiento a lo largo de estos meses y la confianza depositada. Tampoco me puedo olvidar de mis compañeros del departamento de Inteligencia de Clientes de ABANCA. En especial, quería agradecer a Karina, Inés y Damian toda su ayuda. Gracias también a mi tutor académico, Ricardo Cao, por su imprescindible colaboración y consejos. Debo decir que sin él este trabajo no hubiera podido salir adelante.

Por último agradecer a toda mi familia, mis amigos, los compañeros de Moneyball BC y a María por su apoyo incondicional, paciencia y consejos.

Índice general

Resumen	XI
1. Introducción	1
1.1. Inteligencia de Clientes	1
1.2. Motivación del problema	1
1.3. Marketing Mix Modelling	2
1.4. SQL	3
2. Marketing Mix Models	7
2.1. Objetivos y datos de los MMMs	7
2.2. Formulación del modelo	7
2.3. Efectos de las variables de publicidad	8
2.3.1. Efecto adstock	8
2.3.2. Diminishing returns	11
2.3.3. Combinación del efecto adstock y el diminishing returns	12
2.4. Cálculo del ROI	13
2.5. Técnicas de modelización	14
2.5.1. Modelo de regresión lineal múltiple	14
2.5.2. Modelos de regresión penalizada	15
2.5.3. Modelo semiparamétrico	16
2.5.4. Modelo de regresión lineal utilizando las transformaciones de las variables de publicidad	17
2.5.5. Modelo de regresión penalizada utilizando las transformaciones de las variables de publicidad	18
2.5.6. Modelo Robyn	19
3. Descripción de las variables	23
3.1. Definición de las variables	23
3.1.1. Variable objetivo	23
3.1.2. Variables explicativas	24
3.2. Serie de altas semanales	26
3.3. Variables explicativas	27
3.3.1. Variables explicativas de publicidad	28
3.3.2. Variables de contexto	28
3.4. Relación de la variable objetivo con las variables explicativas	32
4. Resultados estadísticos	35
4.1. Modelo lineal	35
4.2. Modelo de regresión penalizada	38
4.3. Modelo semiparamétrico	39
4.4. Modelo de regresión lineal con transformaciones	41

4.5. Modelos de regresión penalizada con transformaciones	43
4.6. Modelo Robyn	45
4.7. Comparación de los modelos	50
5. Optimización del presupuesto	53
5.1. Algoritmo del lagrangiano aumentado	54
5.2. Resultados	55
6. Conclusiones y líneas futuras	59
6.1. Conclusiones	59
6.2. Líneas de trabajo futuras	60
A. Prophet	63
B. Algoritmos de evolución diferencial	67
B.1. Fase de inicialización	68
B.2. Mutación	68
B.3. Recombinación	69
B.4. Selección	70
B.5. Evolución diferencial en Robyn	70
Bibliografía	73

Resumen

Resumen en español

ABANCA Corporación Bancaria S. A lanzó una campaña publicitaria en el año 2022 en la que se invirtió una gran suma de dinero. Esto se debe en parte al lanzamiento de anuncios televisivos de forma nacional en abril de ese mismo año. El número de nuevos clientes del banco se vió afectado positivamente por esto. Sin embargo, se quiere desarrollar una herramienta que permita a los agentes de marketing detectar cuáles son los medios publicitarios más efectivos a la hora de atraer nuevos clientes. De esta forma se tendría una idea de como repartir a posteriori el presupuesto para maximizar los niveles de captación de clientes

En este proyecto se desarrollan distintos Marketing Mix Models para determinar esto. Estos permiten tratar el problema inicial mediante el ajuste de un modelo de regresión. Posteriormente una vez obtenido el ajuste, son capaces de resolver un problema de optimización utilizando las variables añadidas en el modelo de regresión.

Para el desarrollo de estos modelos de regresión se comenzó con modelos clásicos como el lineal múltiple. Posteriormente se comenzaron a utilizar técnicas más complejas como los modelos semiparamétricos. Llegando a Robyn el MMM creado por los analistas de Facebook. Este tiene en cuenta efectos de estacionalidades, vacacionales y de tendencia a parte de los añadidos originalmente. Posteriormente para resolver el problema de optimización se utilizaron las herramientas aportadas por los creadores del modelo.

English abstract

ABANCA Corporación Bancaria S. A launched an advertising campaign in the year 2022 in which it invested a large sum of money. This is partly due to the launch of advertisements on national television on April of that same year. The number of new bank customers was affected positively because of that investment. However, ABANCA wants to develop a tool that allows it's marketing agents to detect which are the most effective advertising media channels when it comes to attracting new customers. In this way, one would have an idea of how to distribute the budget a posteriori to maximize the number of bank registrations.

In this project different Marketing Mix Models are developed to determine this. These models deal with the initial problem by fitting a regression model. Once the fit is obtained, they are able to solve an optimization problem using the variables added in the regression model.

For the development of these regression models, the most classical regression models where used. As the proyect advanced more complex models where adjusted. At the end, the MMM Robyn created by the Facebook analysts was introduced. This models takes into account seasonality , holiday and

trend effects apart from those originally added. Subsequently, to solve the optimization problem, the tools provided by the creators of the model were used.

Capítulo 1

Introducción

El capítulo inicial se destina a explicar el interés de ABANCA por este proyecto. Además, en él se introducen los modelos estudiados y las variables utilizadas a lo largo del trabajo. En cuanto a su estructura, la Sección 1.1 se corresponde con una breve introducción del equipo en el que se realiza este proyecto. La Sección 1.2 abarca la motivación del problema. En la Sección 1.3 se presentan los Marketing Mix Models, empleados para dar solución a la cuestión. En la Sección 1.4 se introducen las nociones básicas de SQL, el lenguaje utilizado dentro de ABANCA para realizar búsquedas en las bases de datos de la empresa. Finalmente, en la Sección 3.1 se incluye una breve descripción de las variables de estudio.

1.1. Inteligencia de Clientes

Este trabajo se ha realizado en colaboración con *ABANCA Corporación Bancaria S. A.* (de ahora en adelante *ABANCA*). Esta es una institución financiera que se distribuye en distintos departamentos.

La analítica de datos está presente en muchos de los departamentos de la entidad, pero destaca en el equipo de Inteligencia de Clientes. Esta abarca los procesos, la recopilación y el análisis de información con el objetivo de entender las necesidades de los clientes y adelantarse a estas. Este departamento esta formado por cuatro equipos: *Analítica Avanzada*, *CRM Omnicanal y Marketing*, *Investigación Comercial y Conocimiento de Clientes* y *Modelos Predictivos y Prospección*.

En particular, este proyecto se ha realizado en el área de Analítica Avanzada, un campo imprescindible debido al crecimiento exponencial del volumen de datos disponible y a la capacidad de diferentes soportes de Internet de recopilar información detallada de los clientes. Dentro las líneas de trabajo del equipo, destaca el desarrollo y mantenimiento de activos analíticos y que sirva de base para nuevos servicios de monetización del dato, es decir, de la extracción, depuración y análisis de información disponible en la compañía con el propósito de generar beneficios para los clientes.

1.2. Motivación del problema

El presente trabajo se centra en proponer modelos que permitan a la entidad entender cuales son los factores que atraen de una forma más eficiente a nuevos clientes.

ABANCA realizó un gasto bastante elevado en la campaña publicitaria de 2022 debido al comienzo de la inversión en publicidad televisiva en el ámbito nacional. Esto genera un salto cuantitativo en el número de altas que se obtienen las semanas siguientes. Sin embargo, la elevada inversión generó

dudas sobre si esta se podría mejorar. A parte de los anuncios televisivos, *ABANCA* invirtió a su vez en publicidad en medios publicitarios online generando también nuevas altas. Se quieren evaluar estas inversiones y comprobar cual habría sido el reparto óptimo del presupuesto. ¿Hubiera sido mejor invertir algo del dinero utilizado en televisión en algún otro tipo de medio publicitario?

El problema al que se enfrentan estos modelos se puede resumir en con la siguiente frase de John Wanamaker :

Half the money I spend on advertising is wasted; the trouble is I don't know which half.

Con esto se refería a que no conocía cual era el reparto óptimo de su presupuesto en publicidad y por tanto, él consideraba que estaba perdiendo una gran suma de dinero con su inversión. Si se quiere conocer más sobre la vida de John Wanamaker, uno de los padres de la publicidad moderna consultar ver Britannica (2023).

Para estudiar esto se ha dispuesto de información relativa a los medios de comunicación tradicionales, donde se incorpora la publicidad a través de anuncios en periódicos, televisión, letreros, rótulos o pósters, entre otros. No obstante, la sociedad está experimentando multitud de avances debidos a la evolución de los medios digitales, lo que también se ve reflejado en las campañas publicitarias. Más aún, esto ha ampliado las opciones de interacción entre una empresa y sus posibles clientes potenciales. Sin ir más lejos, *ABANCA* ha decidido aprovecharse de esta evolución tecnológica y apostar por la publicidad en medios digitales desde el año 2011. Además, y a diferencia del enfoque tradicional, el uso de herramientas digitales permite recabar información adicional del impacto de un anuncio. Por ejemplo, se contempla la posibilidad de registrar el número de veces que un mensaje ha aparecido en la pantalla de un usuario, el total de visitas a una página web, el tiempo de permanencia en la misma, etc.

Las inversiones tanto en digital como en televisión nacional causaron un efecto muy positivo en el cómputo de nuevos clientes de *ABANCA*. De todas formas, obtener un buen resultado no supone garantías de óptima rentabilidad y es imprescindible cuantificar como afectó cada medio de comunicación en el número de altas. Efectivamente, modelar las estructuras de dependencia subyacentes entre el gasto en publicidad y la captación de clientes, según el medio empleado, permitirá proponer nuevas estrategias de inversión y optimizar recursos en base a una visión comprensiva y estratégica de las decisiones pasadas.

Existen distintos modelos estadísticos destinados a estudiar la efectividad de los anuncios en Marketing.

1.3. Marketing Mix Modelling

Dentro de las herramientas utilizadas para medir el impacto que generan los anuncios se encuentran distintos modelos estadísticos. A lo largo de los últimos años, se han propuesto metodologías para medir el efecto que tienen los anuncios como, por ejemplo, los *multi-touch attribution models*. Estos modelos intentan darle pesos a los anuncios que llevan a un cliente a realizar una acción deseable. Por otro lado, destacan los *Marketing Mix Models* (MMM) que, utilizando los datos del gasto en publicidad o de la exposición de los anuncios, pretenden explicar cuántas de las altas del banco han sido generadas por los anuncios publicados y además intentan asociar cuál ha sido el anuncio que provocó que ese cliente consumiera el producto.

El presente trabajo hará uso de los MMMs para optimizar el gasto en publicidad de *ABANCA*. Estos aportarán un base estadística sólida que ayudará a explicar la relación entre el número de altas semanales del banco y la inversión en publicidad. Para ello, se utilizan datos relativos al gasto televisivo desde abril hasta diciembre de 2022, ambos incluidos, y de gasto online de 2021 y 2022, en su

totalidad. A pesar de que su nombre pueda sugerir que se tratan de modelos de regresión mixtos, se trata de modelos de regresión de efectos fijos. La segunda M de los MMMs se refiere al mix de medios que emplea una empresa. Es decir, a todas aquellas actividades publicitarias de esta.

A continuación se describen brevemente las variables de los modelos. La variable respuesta de estos modelos son KPIs (Key Performance Indicator) de interés. Un KPI es un indicador de calidad. Cada empresa puede fijar los suyos propios y se utilizan para evaluar el éxito que tienen ciertas acciones o decisiones que se toman. En la sección 3.1 se han introducido los distintos KPIs de interés que hay en la captación de clientes.

Por otro lado, las variables explicativas del modelo se diferencian en dos grupos:

- Variables explicativas de publicidad o marketing: son aquellas variables que tienen que ver con la actividad publicitaria de la empresa. Destacar que se tiene decisión sobre el valor es estas.
- Variables explicativas de contexto: Aquellas que no tienen relación con las decisiones de marketing de la entidad, sirven para explicar la situación del sector. Al no depender de la actividad publicitaria de la empresa no tenemos posibilidad de acción sobre ellas. Estas variables se espera que también afecten de una u otra manera al número de altas del banco. Por ejemplo, la evolución de factores macroeconómicos como el *PIB* (Producto Interior Bruto) o la actividad de la competencia.

Utilizando estas variables y modelos de regresión se estima el KPI elegido. Una vez se ha realizado el ajuste del modelo es de gran interés utilizar la fórmula de este para optimizar el reparto. De esta forma se puede dar recomendaciones a los agentes de marketing de como se puede repartir el presupuesto en publicidad.

1.4. SQL

Los datos utilizados en este problema se extraen de TERADATA. Este es un gestor de base de datos relacionales que administra las bases de datos internas de ABANCA. Emplea el lenguaje de consultas SQL (Structured Query Language) para la extracción de datos. La información está distribuida en tablas de nombre único, compuestas por records (filas) y campos (columnas). Son bases de datos estructuradas, ya que las tablas pueden estar relacionadas.

SQL permite especificar qué campos de una tabla se desean seleccionar y qué requisitos deben cumplir los records. Por ejemplo, suponiendo como disponible una tabla de gastos en publicidad web, se pueden consultar aquellos campos presentes en la tabla empleando consultas como la siguiente:

```
select *  
from inversion_ficticia  
limit 4
```

En esta consulta se utilizan 3 palabras clave importantes. Por una parte, con *select* se seleccionan los campos. El carácter *** se permite visualizar todos los campos. Seguidamente, en *from* se especifica el nombre de la tabla donde están almacenados los datos y *limit* restringe el número de records a mostrar por pantalla. La Tabla 1.1 recoge un ejemplo de este tipo de consultas.

Fecha	Soporte	Inversion	Tipo_Anuncio	Clicks	Impresiones
2023-02-01	Google	1500	Ventana emergente	1000	2000
2023-02-01	Facebook	2000	Red social	700	4000
2023-02-03	Amazon	1200	Ventana emergente	1000	8000
2023-02-07	Instagram	500	Red social	600	3000

Tabla 1.1: Tabla ficticia de inversión en publicidad.

Se pueden agrupar los registros según algunas de sus características y aprovechar para hacer operaciones sobre esos grupos, para ello se utiliza la función `group.by`. Una vez agrupados los records, se pueden hacer distintas operaciones por columnas:

- `sum()` y `prod()`: para sumar y multiplicar todos los elementos, respectivamente,
- `count()`: para contar elementos sean iguales o no,
- `min()` y `max()`: para obtener mínimos y máximos, respectivamente,
- `avg()`: para calcular medias.

Por ejemplo, para consultar la inversión realizada según el tipo de anuncio, es necesario agrupar todos aquellos que sean de un mismo tipo. La siguiente consulta permite realizar esto. En la Tabla 1.2 se recoge el resultado.

```
select tipo_Anuncio, sum(Inversion)
from inversion_ficticia
group by tipo_Anuncio
```

tipo_anuncio	inversion
Display	2700
Social Network	2500

Tabla 1.2: Suma de la inversión según el tipo de anuncio.

A continuación, se muestran algunas de las funciones que se han utilizado para extraer las series de tiempo semanales de gasto en publicidad. Con la siguiente consulta se obtienen los datos de la Tabla 1.3 agregados por lunes de la semana a la que pertenece la columna *fecha*. La función `Next.day` es la utilizada para conseguir agregar las fechas tal y como se puede observar en la siguiente consulta

```
select Next_day(fecha-6,'Monday'), sum(Inversion)
from inversion_ficticia
where Inversion>0
group by 1
order by 1
limit 2
```


fecha	inversion
2023-01-30	4700
2023-02-06	500

Tabla 1.3: Inversión las dos primeras semanas de febrero.

Al trabajar con bases de datos estructuradas y relacionales, SQL permite obtener consultas combinadas de tablas. Para ello se muestra primero una tabla donde se representa el número de altas en las mismas semanas que aparecen en la tabla 1.4.

fecha	altas
2023-01-30	575
2023-02-06	394

Tabla 1.4: Número de altas las dos primeras semanas de febrero.

Se puede combinar estas tablas de forma que obtengan en una misma salida las altas obtenidas en esa semana junto con la inversión en publicidad online. Destacar que hay distintos tipos de *join* con los que se pueden hacer estos cruces de tablas. La consulta con la que se unen las tablas utiliza la función `left join`.

```
select a.fecha, a.inversion, b.altas
from (select Next_day(fecha-6,'Monday'),
sum(inversiones) as inv
from inversion_ficticia
where inv>0
group by 1
order by 1
limit 2) a
left join
(select Next_day(fecha-6,'Monday'), altas
from altas_ficticias
) b
on a.fecha=b.fecha
```

Con esta consulta se obtendría una tabla igual a la Tabla 1.5.

fecha	altas	inversion
2023-01-30	575	4700
2023-02-06	394	500

Tabla 1.5: Número de altas las dos primeras semanas de febrero junto con la inversión.

Para finalizar esta sección comentar que se utilizó SQL para crear, modificar y eliminar tablas. Se organizaron todos los datos de forma que los datos se pudieran obtener con unas pocas consultas.

Se utilizaron los comandos `DROP TABLE` y `CREATE TABLE` para eliminar y crear tablas. Después de crear una tabla se deben insertar los nuevos datos. Para ello, empleando el comando `INSERT INTO` indicando los datos que se quieren añadir a la tabla. Para más información sobre este lenguaje de computación para trabajar con bases de datos consultar Nield (2016).

Capítulo 2

Marketing Mix Models

En este Capítulo se ha definido las bases teóricas de los Marketing Mix Models y se han presentado los ajustes que hemos realizado a lo largo del trabajo.

La primera vez que se introdujo el concepto de Marketing Mix fue Borden (1964). En él, define a un publicista como un "mezclador de ingredientes". La referencia bibliográfica considerada como base para la formulación teórica y exposición de los conceptos incorporados en este capítulo y, en concreto sobre los MMMs, son Wang et al. (2017) y Jin et al. (2017).

2.1. Objetivo y datos de los MMMs

Los MMMs son utilizados para dos tareas principales:

- Entender el efecto de los distintos tipos de publicidad sobre el KPI elegido como variable objetivo del problema. Generalmente, se utilizan datos de ventas o visitas a la web.
- Optimizar el reparto del presupuesto en publicidad con el objetivo de maximizar el KPI.

Los datos utilizados para ajustar estos modelos son mayoritariamente semanales, de entre dos y cinco años. De todas formas se pueden utilizar datos mensuales, de entre cuatro o seis años. De esta forma se obtienen entre 50 y 250 observaciones para ajustar el modelo. En Wang et al. (2017) sugieren que incluso si se dispone datos más lejanos en el tiempo, no es recomendable utilizarlos. De utilizarse datos de gran longitud en el tiempo, el entorno que envuelve al marketing puede haber cambiado a lo largo del periodo de estudio. Esto provoca que estos modelos tengan un volumen bajo de datos, teniendo en cuenta que para ajustar estos modelos se pueden llegar a utilizar una gran cantidad de variables.

Estos modelos ayudan a encontrar las diferencias entre los distintos canales publicitarios incluidos en el modelo. Por ejemplo, se desconoce si son más efectivos los anuncios tipo Display o en anuncios de videos-online a la hora de atraer nuevos clientes.

Una vez obtenido el ajuste del modelo, este se utiliza para optimizar el reparto del presupuesto en marketing de la empresa. Para ello, se tienen en cuenta únicamente las variables de marketing añadidas al modelo.

2.2. Formulación del modelo

Para el momento $t = 1, \dots, T$, se denota por y_t al valor de la serie objetivo en la semana t . Por otro lado, $x_{t,m}$ se corresponde con el valor de la variable de publicidad m en la semana t del estudio con

$m = 1, \dots, M$. Por último se define $z_{t,c}$ como el valor de la variable de control c en el momento t siendo $c = 1, \dots, C$. Las variables de control se deben de elegir dependiendo de la variable objetivo utilizada.

La función de regresión se puede escribir de la forma:

$$\Phi(y_t) = \beta_0 + \sum_{m=1}^M \beta_m f_m(x_{t-L+1,m}, \dots, x_{t,m}) + \sum_{c=1}^C \delta_c z_{t,c} + \varepsilon_t, \quad (2.1)$$

donde β_m es el coeficiente de regresión asignado a un canal publicitario, δ_c es el coeficiente asociado a una variable de control. $f_m()$ es una función que captura la no linealidad de las variables de marketing m , sobre la variable objetivo. Por último, destacar que L es el parámetro que aporta el efecto temporal en la fórmula. $\Phi()$ es la función que transforma la variable respuesta, generalmente se utiliza la identidad o la función logarítmica y ε_t es el error que captura la variabilidad de y_t que no pueden explicar las variables de publicidad y de contexto. Estos además, deben de ser ruido blanco, se asume que no presentan correlación con el resto de variables del problema y tienen varianza constante.

Definición 2.2.1 (Ruido blanco). Ruido blanco es una colección de variables aleatorias incorreladas, con media nula y varianza finita σ^2

Una vez presentada la fórmula, cabe destacar que las variables de contexto se ajustan con un efecto lineal mientras que las de marketing capturan efectos no lineales.

2.3. Efectos de las variables de publicidad

Se pueden encontrar algunos patrones al trabajar con variables de marketing (ver Tellis (2006)). Existen siete efectos a destacar dentro de las variables de publicidad. De estos se han presentado el efecto *adstock* o *carryover* y el efecto de *diminishing returns* o saturación.

2.3.1. Efecto adstock

Este efecto permite medir la cantidad de gasto de la semana t que se ve reflejado en las semanas posteriores, puede ocurrir por diversos motivos como por ejemplo, la publicación tardía de un anuncio o un cliente que tras ver el anuncio tarda una semana o más en convertirse en alta. Típicamente este efecto es de corta duración. El impacto total del gasto en publicidad se puede medir como la suma de la inversión y del efecto carryover estimado. Existen distintas funciones con las que obtener la inversión en publicidad utilizando este efecto. Entre ellas se han estudiado las siguientes:

- En Jin Y et al. (2017) utilizan la función *adstock* definida en (2.2):

$$adstock(x_{t-L+1,m}, \dots, x_{t,m}, \omega_m, L) = \frac{\sum_{l=0}^{L-1} \omega_{l,m} x_{t-l,m}}{\sum_{l=0}^{L-1} \omega_{l,m}} \quad (2.2)$$

dónde $m = 1, \dots, M$ se corresponde con el índice de la variable de marketing y ω_m con un vector de pesos no negativos. La transformación *adstock* es un promedio de los datos de la semana t con las anteriores $L - 1$ semanas. La elección de L es importante para ajustar el modelo, es normal utilizar valores altos de L obteniendo así una aproximación al infinito, haciendo que los pesos finales $\omega_{l,m}$ sean cada vez más próximos a 0 a medida que l se acerca a L .

En Jin Y et al. (2017) proponen dos formas distintas de obtener el vector de pesos ω_m . La primera de ellas es la función de decaimiento geométrico de forma que se define ω_m de la siguiente forma:

$$\omega_m^g(l, \theta_m) = \theta_m^l, \quad \theta_m \in (0, 1) \quad (2.3)$$

dónde θ_m es el ratio de caída de la función de decaimiento geométrico. Utilizar esta fórmula para el vector ω_m provoca que la inversión que tiene mayor influencia en la semana t es la de esa misma fecha. Este no es siempre el caso, a veces la inversión se ve reflejada con algo de retraso por motivos como la publicación tardía de un anuncio. Para ello, se propone utilizar una función similar a (2.3) pero añadiendo un parámetro que provoque el retardo del efecto de la inversión:

$$\omega_m^d(l, \theta_m, \delta_m) = \theta_m^{(l-\delta_m)^2}, \quad l = 0, \dots, L-1, \quad \theta_m \in (0, 1), \quad \delta_m \in [0, L-1] \quad (2.4)$$

donde δ_m es el retraso del pico de la publicidad m . Comentar que Jin et al. (2017) también propone funciones como la distribución binomial negativa para modelar el efecto retardado del pico de la media.

- Zhou et al. (2023) desarrollan el MMM Robyn y para el efecto adstock proponen la siguiente función para la transformación adstock

$$adstock(x_{1,m}, \dots, x_{t,m}, \omega_m) = \sum_{l=0}^{t-1} \omega_{l,m} x_{t-l,m}, \quad (2.5)$$

es una función muy similar a (2.2). Sin embargo, no se tiene que elegir un L debido a que se tienen en cuenta todos los datos previos para hacer esta transformación. Además, no se divide el sumatorio por la suma de los pesos. Para el vector de pesos ω_m proponen la función de decaimiento geométrico y las funciones de distribución y supervivencia de la Weibull.

Esta distribución es una de las más utilizadas en los modelos del análisis de supervivencia y tiene aplicaciones en la durabilidad de objetos, como motores o maquinarias industriales. Su función de densidad viene dada por:

$$f(t) = \theta \gamma (\theta t)^{\gamma-1} \exp(-(\theta t)^\gamma),$$

mientras que la función de supervivencia viene formulada como

$$S(t) = \exp(-(\theta t)^\gamma).$$

Para más información de esta distribución consultar Lawless (2011), dónde se describen diversas funciones utilizadas en el análisis de supervivencia. Entre ellas algunas que se podrían utilizar para el ajuste del efecto adstock, como la función de supervivencia de la distribución exponencial.

Los pesos en el caso de utilizar la distribución Weibull serían

$$w_{l,m} = f_m(l),$$

si se utiliza la función de distribución. Por otro lado, si se utilizase la de supervivencia se toma

$$w_{l,m} = S_m(l),$$

donde f_m y S_m son las funciones de la densidad y supervivencia Weibull, respectivamente, con parámetros θ_m y γ_m . Recalcar que l toma únicamente valores enteros.

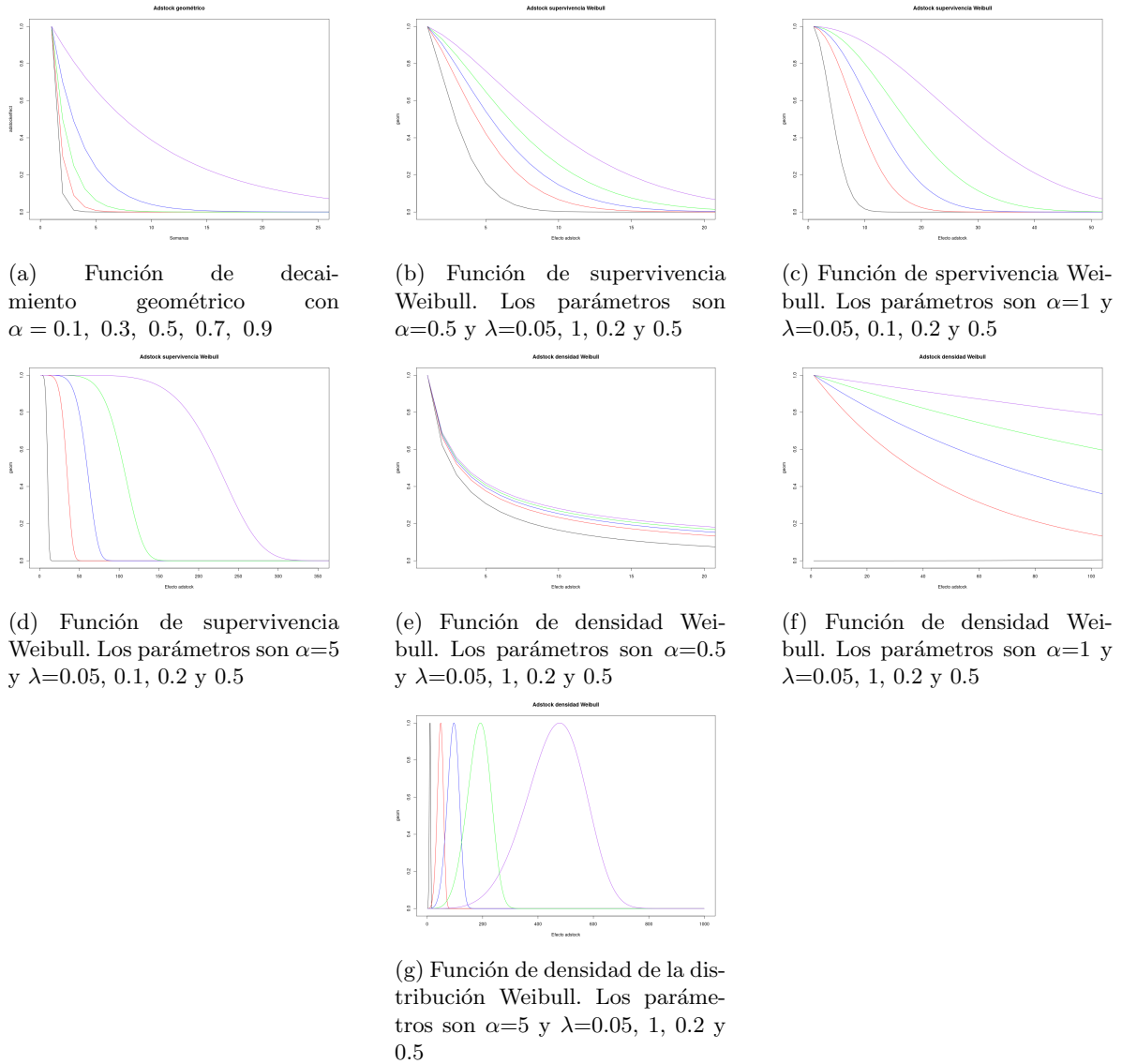


Figura 2.1: Gráficas de las funciones propuestas para ajustar el efecto adstock de los anuncios.

En la Figura 2.1 se pueden observar las funciones de decaimiento geométrico, la función de supervivencia y densidad de la distribución Weibull para distintas combinaciones de los hiperparámetros de estas. La función más simple de estas es la de decaimiento geométrico debido a que sólo tiene un único parámetro. Sin embargo, las funciones de densidad y supervivencia de la distribución Weibull permiten dar más flexibilidad a este efecto. Con la combinación de parámetros adecuada se puede ajustar de forma que el pico no sea inmediato, de forma similar a (2.4).

Destacar que en este trabajo se ha utilizado la función propuesta en Zhou et al. (2023) con la transformación de decaimiento geométrico. Además de que los anuncios de publicidad online son más inmediatos que los anuncios offline, por lo que añadir un efecto retardado no tendría demasiado sentido.

Por último, de ahora en adelante se denota por $x_{t,m}^{ad}$ a la variable de marketing m en el momento

t una vez realizada la transformación adstock.

2.3.2. Diminishing returns

A la hora de invertir en publicidad, el efecto sobre las ventas o las altas de empresa no es lineal. Una vez se ha invertido una gran suma de dinero, el número de clientes potencial a los que genera interés se ve disminuido y por tanto no se obtienen unos resultados tan cuantiosos como antes, provocando así un efecto de saturación.

La función más popular en la literatura para modelar esta curvatura es la función de Hill. Esta función se utilizaba originalmente en el mundo de la farmacología y fue introducida por Gesztelyi et al. (2012). Viene definida de la siguiente forma:

$$Hill(x_{t,m}; \gamma_m, \alpha_m) = \frac{x_{t,m}^{\alpha_m}}{x_{t,m}^{\alpha_m} + \gamma_m^{\alpha_m}}.$$

Esta función toma valores entre el 0 y el 1. Destacar que el parámetro $\alpha_m > 0$ le da forma a la función, cuanto más pequeño sea este parámetro, menos va a tardar la función en llegar a un punto próximo al 1. Por otro lado, el parámetro $\gamma_m > 0$ es conocido como el punto medio de saturación debido a que cuando $x_{i,j} = \gamma_m$ se cumple que $Hill(\gamma_m, \gamma_m, \alpha_m) = 1/2$ para cualquier valor de γ_m y α_m . La función de Hill converge a 1 cuando $x_{t,m}$ tiende a infinito.

$$\lim_{x_{t,m} \rightarrow \infty} Hill(x_{t,m}; \gamma_m, \alpha_m) = \lim_{x_{t,m} \rightarrow \infty} \frac{x_{t,m}^{\alpha_m}}{x_{t,m}^{\alpha_m} + \gamma_m^{\alpha_m}} = \lim_{x_{t,m} \rightarrow \infty} \frac{1}{1 + \frac{\gamma_m^{\alpha_m}}{x_{t,m}^{\alpha_m}}} = 1$$

Por último, destacar que en Jin et al. (2017) mencionan que esta función se puede reescribir cómo sigue:

$$Hill(x_{t,m}, \gamma_m, \alpha_m) = 1 - \frac{\gamma_m^{\alpha_m}}{x_{t,m}^{\alpha_m} + \gamma_m^{\alpha_m}}.$$

Multiplicando esta función por el coeficiente de regresión correspondiente β_m se obtiene

$$\beta_m Hill(x, \gamma_m, \alpha_m) = \beta_m - \frac{\beta_m \gamma_m^{\alpha_m}}{x^{\alpha_m} + \gamma_m^{\alpha_m}}$$

que tiene una formulación muy similar a la utilizada en Jin et al. (2012) para estudiar la relación entre el alcance de los anuncios de televisión con los GRPs. Esto nos indica que esta función nos va a aportar un buen ajuste a la hora de utilizarla para modelar la curva $\beta_m f_m(x_{t-L+1,m}, \dots, x_{t,m})$.

A continuación vamos a estudiar la naturaleza de esta función, por lo que lo primero que vamos a hacer es la derivada de la función de Hill en función de la variable de media:

$$\frac{\partial Hill(x_{t,m}, \gamma_m, \alpha_m)}{\partial x_{t,m}} = \frac{\alpha_m x_{t,m}^{\alpha_m-1} \gamma_m^{\alpha_m}}{(x_{t,m}^{\alpha_m} + \gamma_m^{\alpha_m})^2} > 0 \quad \forall x_{t,m} > 0$$

al ser positiva para todo $x_{t,m} > 0$ la función es creciente, es decir cuanta más inversión mayor será el valor de la función de Hill.

Destacar también que esta función cumple que $x_{t,m}^{\alpha_m} = \gamma_m^{\alpha_m}$ es un punto de inflexión de la función de Hill $\forall \alpha_m > 0$.

$$\frac{\partial^2}{\partial x^2} Hill(x, \gamma_m, \alpha_m)|_{x=\gamma_m} = \frac{\alpha_m \gamma_m^{\alpha_m} ((\alpha_m - 1) \gamma_m^{\alpha_m} x^{\alpha_m-2} - (\alpha_m + 1) x^{2\alpha_m-2})}{(\gamma_m^{\alpha_m} + x^{\alpha_m})^3} \Big|_{x=\gamma_m} = 0$$

Se ha comprobado que la inversión en publicidad siempre va a afectar positivamente al *KPI* de interés, sin embargo no siempre de la misma manera, hay canales publicitarios que invirtiendo en ellos al principio no hay apenas efecto en el *KPI* pero posteriormente hay una subida más elevada. La función de Hill nos ayuda con ambos casos pero para eso la selección de los parámetros de esta va a ser esencial. Podemos observar en la Figura 2.2 como, dependiendo de la elección que los parámetros de la función, vamos a obtener funciones que se saturan en distintos momentos. También se han observado casos donde la función toma una forma de *S*.

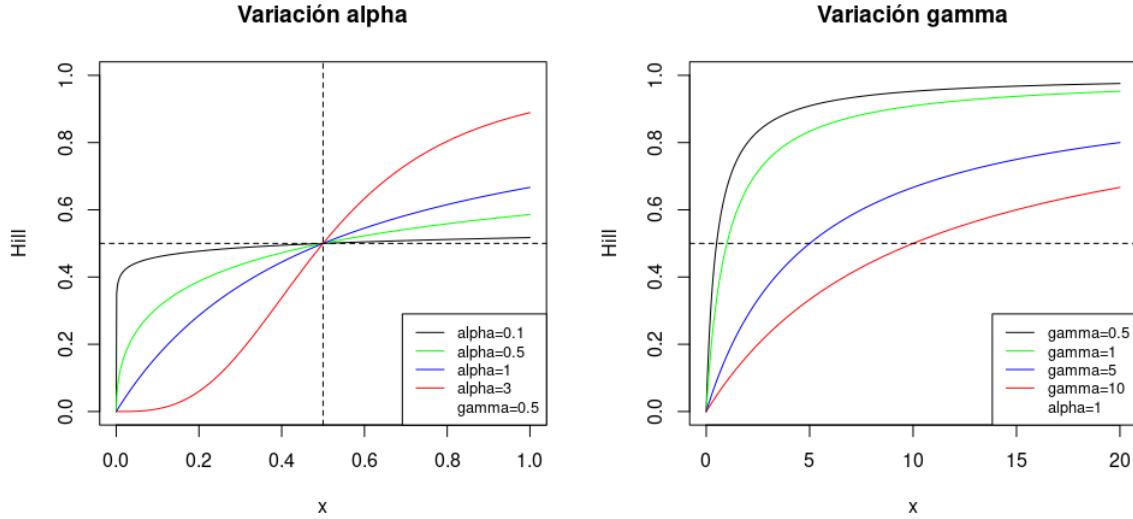


Figura 2.2: Gráfico con distintas funciones de Hil

En la Figura 2.2 se pueden observar dos gráficos. En el izquierdo, se pueden observar varias funciones de Hill con parámetros $\gamma = 0.5$ y α tomando valores 0.1, 0.5, 1 y 3, mientras que en el derecho se muestran funciones de Hill con parámetros $\alpha = 1$ y γ tomando valores 0.5, 1, 5 y 10.

Destacar que según la magnitud de la variable a transformar, un mismo valor de γ puede ser muy grande o muy pequeño. De ahora en adelante se denota por γ_m^H al hiperparámetro utilizado en la función de Hill y $\gamma_m \in [0, 1]$ lo define como:

$$\gamma_m^H = \min(X_m)(1 - \gamma_m) + \max(X_m)\gamma_m.$$

De esta forma γ_m^H se define según los datos de la variable X_m . Esto se debe a que de esta manera se pueden definir de igual forma las funciones de Hill para los distintos canales de marketing.

2.3.3. Combinación del efecto adstock y el diminishing returns

Estos efectos están presentes siempre a la hora de hacer publicidad. Por lo tanto, se deben añadir ambos dentro del modelo. Existen dos opciones para ello:

- Añadir primero el efecto adstock y posteriormente el efecto de saturación. Para ello se debe utilizar la función compuesta $(Hill \circ adstock)(x_{t,m}, \dots, x_{1,m}; \alpha_m, \gamma_m, \theta_m)$
- Añadir primero el efecto de saturación y posteriormente el efecto de adstock. Esta transformación sería de la siguiente manera $(adstock \circ Hill)(x_{t,m}, \dots, x_{1,m}; \alpha_m, \gamma_m, \theta_m)$

En Jin et al. (2017) se discuten bajo qué circunstancias se utiliza cada uno de estos caminos. Si se dispone de datos muy concentrados en periodos de tiempo, se debe utilizar primero el efecto saturación y posteriormente el adstock. Por otro lado, si hay cambios relativamente pequeños en la inversión, se obtiene un mejor ajuste utilizando la otra ruta.

Se ha optado optado por ejercer primero el efecto adstock y sobre el hacer la transformación utilizando la función de Hill. Obteniendo así la siguiente fórmula para el modelo:

$$y_t = \beta_0 + \sum_{m=1}^M \beta_m \text{Hill}(x_{t,m}^{ad}; \alpha_m, \gamma_m) + \sum_{c=1}^C \delta_c z_{t,c} c + \varepsilon_t \quad (2.6)$$

donde, como se ha comentado anteriormente, se denota por $x_{t,m}^{ad}$ al valor que toma la variable de media tras aplicarle la transformación adstock. Se puede interpretar el parámetro β_0 como las altas base, es decir, las altas que se generan por puro conocimiento de la marca.

Se puede destacar que al combinar estas transformaciones han sido añadidos hiperparámetros al modelo. Por cada variable de publicidad se habrán añadido 3 hiperparámetros si se ha tomado la función ad-stock para la transformación adstock. Sin embargo, si se toma alguna de las funciones Weibull comentadas en esta sección, esto implica que cada variable de marketing implica añadir cuatro hiperparámetros al modelo.

Destacar que esta es la fórmula más popular para ajustar Marketing Mix Models. Aparte de Facebook, hay otras grandes empresas como Google, que han optado por esta fórmula para ajustar sus modelos. La diferencia entre el modelo de Facebook y el de Google (Lightweight) está en el ajuste, tanto de los hiperparámetros, como de los parámetros.

Por último, aclarar que se denota como $x_{t,m}^{sat}$ a la variable x_m en el momento t tras haber hecho la transformación de *carryover* y de *diminishing returns*. Por otro lado, se denota a

$$g_m(x) = \beta_m \text{Hill}(x; \alpha_m, \gamma_m), \quad (2.7)$$

como la curva respuesta de la variable de marketing m .

2.4. Cálculo del ROI

Como se ha comentado en la Sección 1.3 una de las medidas más interesantes de estos modelos es el ROI (Return of Investment). Esta nos permite medir la efectividad que tiene el gasto en publicidad a la hora de captar clientes. El ROI se obtiene por cada tipo de variable publicitaria, por lo que el modelo tendrán tantas estimaciones del ROI como variables de marketing. En Wang et al. (2017) se estima de la siguiente forma:

$$ROI_m = \frac{\sum_{t=1}^T \hat{y}_t(X_{t,m} = x_{t,m}^{sat}) - \hat{y}_t(X_{t,m} = 0)}{\sum_{t=1}^T x_{t,m}} = \frac{\sum_{t=1}^T \beta_m x_{t,m}^{sat}}{\sum_{t=1}^T x_{t,m}}, \quad (2.8)$$

de esta forma el ROI no es más que el cociente entre las altas que ha provocado la variable de marketing m entre el la inversión que se hizo en ese tipo de publicidad a lo largo del periodo de estudio.

Esta medida será imprescindible para que los publicistas tomen decisiones. Los canales que tengan una gran inversión se van a ver afectados negativamente si no han obtenido un gran número de altas. Esta medida nos ayuda a medir la efectividad de la inversión a la hora de atraer nuevos clientes. Cabe destacar que la inversión está medida en euros y como se ha observado en el Capítulo 3 tiene una magnitud muy superior a la del número de altas. Esto implica que se obtengan valores muy pequeños

de al estimación del ROI para los distintos canales añadidos al modelo.

El ROI ha servido a los publicistas durante años para apoyarse en él a la hora de tomar decisiones. Las variables publicitarias con valores de ROI más altos son aquellas en las que interesa invertir. De la misma manera, las variables con valores de ROI más bajos son las que nos aportan menos retorno por cada euro invertido, es decir, en las que se debería invertir menos dinero.

2.5. Técnicas de modelización

A lo largo de este trabajo se han propuesto varios modelos. Se ha comenzado estudiando los modelos de regresión más sencillos. Posteriormente, debido a que los MMMs trabajan con variables con coeficientes de correlación altos, se han introducido los modelos de regresión penalizada. A continuación, se han comenzado a añadir efectos no lineales sobre las variables de marketing, empezando con los modelos parcialmente lineales y posteriormente utilizando las transformaciones de la Sección 2.3 a las variables de marketing.

2.5.1. Modelo de regresión lineal múltiple

El modelo de regresión lineal múltiple permite estudiar las relaciones lineales entre una variable respuesta y unas variables explicativas. Este modelo considera una combinación lineal de las variables cómo se expone en Crujeiras Casais y Conde Amboage (2019)

$$Y_t = \beta_0 + \beta_1 X_{t,1} + \cdots + \beta_M X_{t,M} + \delta_1 Z_{t,1} + \cdots + \delta_C Z_{t,C} + \varepsilon_t \quad (2.9)$$

siendo Y_t el valor de la variable respuesta en la t -ésima semana, $X_{t,1}, X_{t,2}, \dots, X_{t,M}$ son los valores que toman las variables de marketing en la semana t del estudio y $Z_{t,1}, \dots, Z_{t,C}$ los valores que toman las variables de contexto, ε_t se corresponde con los errores del modelo de regresión.

El modelo de regresión lineal múltiple puede ser reformulado utilizando la formulación matricial. Para ello primero vamos a reescribirlo de la siguiente forma:

$$\begin{pmatrix} Y_1 \\ \vdots \\ Y_t \\ \vdots \\ Y_t \end{pmatrix} = \begin{pmatrix} 1 & X_{1,1} & \cdots & X_{1,M} & Z_{1,1} & \cdots & Z_{1,C} \\ \vdots & \vdots & \ddots & \vdots & \vdots & \ddots & \vdots \\ 1 & X_{t,1} & \cdots & X_{t,M} & Z_{t,1} & \cdots & Z_{t,C} \end{pmatrix} \cdot \begin{pmatrix} \beta_0 \\ \beta_1 \\ \vdots \\ \beta_M \\ \delta_1 \\ \vdots \\ \delta_C \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_t \end{pmatrix}.$$

y de esta forma se puede escribir utilizando la siguiente notación:

$$Y = X\Omega + \varepsilon$$

dónde la matriz X contiene todos los valores de las variables explicativas y se denomina matriz de diseño, $\Omega = (\beta, \delta) \in \mathbb{R}^{M+C}$ es un vector desconocido. Este toma los valores para los que se cumpla:

$$\min_{\beta} \sum_{i=1}^n (Y_i - X_i \Omega)^2 \quad (2.10)$$

siendo x_i la i -ésima fila de la matriz X también denominada matriz de diseño y Ω el vector dónde están contenidos los coeficientes de regresión. En notación matricial se puede reescribir la ecuación (2.10) como:

$$\min_{\beta} (Y - X\Omega)'(Y - X\Omega),$$

de esta forma el estimador por mínimos cuadrados es

$$\hat{\Omega} = (X'X)^{-1}X'Y$$

y para que esté bien definido es necesario que la matriz $X'X$ sea no singular.

Por otro lado la simplicidad, de este modelo puede dar problemas. No se tiene certeza de que la relación entre las variables de media y la respuesta tengan un efecto positivo. Además, para que los resultados de estos modelos sean válidos, se deben cumplir ciertas condiciones:

- Linealidad: debe de existir una relación lineal entre las variables explicativas y la variable respuesta.
- Homocedasticidad: La varianza de los errores debe ser constante, es decir:

$$Var(\varepsilon|X = x) = \sigma^2$$

- Normalidad: los errores del modelo deben de ser normales.

$$\varepsilon \sim N(0, \sigma^2)$$

- Independencia: las variables aleatorias que representan los errores $\varepsilon_1, \dots, \varepsilon_t$, son mutuamente independientes, siendo t el número de observaciones de la muestra.

Además de estas, al estar trabajando con series de tiempo, se comprueba si los residuos siguen teniendo dependencia temporal. Para ello se estudia si la serie de residuos del modelo se puede considerar ruido blanco. De cumplir estas hipótesis, el modelo sería válido para obtener conclusiones.

Destacar que este modelo no suele realizar buenos ajustes cuando las variables predictoras presentan coeficientes de correlación elevados entre ellas.

2.5.2. Modelos de regresión penalizada

Como se comentó en la sección anterior, estos modelos normalmente presentan una correlación elevada entre las variables que utilizan. Este es un problema cuando se ajustan modelos de regresión debido que pueden dar estimaciones con mucha varianza. La solución para esto está en añadir complejidad al modelo, a la par que se reduce la varianza. Para tratar este problema son de gran utilidad los modelos de regresión penalizada. Estos tratan la colinealidad mediante la regularización (penalización) en la estimación de los parámetros. Esto va a provocar que muchos de los parámetros β se estimen con valores muy próximos al 0 o que sean directamente cero.

Los tres modelos más conocidos de regresión penalizada son:

- Ridge: Se penaliza utilizando la norma euclídea del vector de parámetros β . La estimación de β para este modelo es aquella que minimiza:

$$\min_{\Omega} \sum_{t=1}^n \left(Y_t - \beta_0 - \sum_{m=1}^M X_{t,m} \beta_m + \sum_{c=1}^C \delta_c Z_{t,c} \right)^2 + \lambda \sum_{j=1}^{M+C} \Omega_j^2$$

dónde λ es el hiperparámetro de penalización. Cuanto mayor sea λ más pequeñas serán las estimaciones del vector de parámetros Ω . El estimador $\hat{\Omega} = (\hat{\beta}, \hat{\delta})$ de la regresión ridge viene dado por:

$$\hat{\Omega}_{RR} = (X'X + \lambda \cdot I)^{-1} X'Y$$

Para más información de este modelo consultar Hoerl y Kennard (1970).

- Lasso: Se penaliza utilizando la norma 1 del vector de parámetros β . Es decir, se busca los coeficientes β que cumplan:

$$\min_{\Omega} \sum_{t=1}^T \left(Y_t - \beta_0 - \sum_{m=1}^M X_{t,m} \beta_m + \sum_{c=1}^c \delta_c Z_{t,c} \right)^2 + \lambda \sum_{j=1}^{M+C} |\Omega_j|.$$

Este modelo hace una selección de variables a la par que estima los coeficientes Ω del modelo de regresión. Además, λ vuelve a ser el parámetro de penalización y, cuanto más grande sea, más coeficientes serán nulos. Si se tiene más interés en el estimador lasso y sus propiedades consultar Tibshirani (1996).

- Elastic Net: Se combina la penalización de la regresión ridge y lasso. De forma que se minimiza:

$$\min_{\Omega} \left(Y_t - \beta_0 - \sum_{m=1}^M X_{t,m} \beta_m + \sum_{c=1}^c \delta_c Z_{t,c} \right)^2 + \lambda \left(\frac{1-\alpha}{2} \sum_{j=1}^{M+C} \Omega_j^2 + \alpha \sum_{j=1}^{M+C} |\Omega_j| \right).$$

Este modelo mantiene la selección de variables que lleva a cabo la regresión Lasso a la vez que la penalización que se utiliza en el modelo de regresión ridge. Para más información sobre la estimación de los parámetros de este modelo y sus propiedades consultar Zou y Hastie (2005).

2.5.3. Modelo semiparamétrico

Los modelos semiparamétricos de regresión intentan resolver el problema del desastre de la multidimensionalidad reduciendo la dimensión de los no paramétricos. Esto aporta mucha flexibilidad al modelo a la vez que se mantiene la interpretabilidad del efecto causado por las variables explicativas.

Dentro de los modelos semiparamétricos se encuentran los modelos parcialmente lineales, los modelos aditivos y los modelos aditivos generalizados. Los dos primeros se utilizan cuando la variable respuesta es una variable continua. Los modelos aditivos generalizados se aplican a una variable respuesta que no es continua.

Los modelos parcialmente lineales buscan ajustar unas variables con efectos suaves mientras que otras se estiman con efectos lineales.

$$Y = \beta_0 + \beta_1 Z_1 + \cdots + \beta_C Z_C + s_1(X_1) + \cdots + s_M(X_M) + \varepsilon$$

dónde el efecto de las variables Z_1, \dots, Z_C es lineal y el efecto de las variables X_1, \dots, X_M no paramétrico, es decir, se estiman utilizando las funciones suaves s_1, \dots, s_m . Por último, ε representa el error de regresión que sigue una distribución normal de media nula y varianza constante.

Por otro lado, en los modelos aditivos se suaviza en varias variables explicativas de forma univariante,

$$Y = s_1(X_1) + \cdots + s_M(X_M) + \varepsilon$$

dónde s_i son las funciones suaves ajustadas para las variables X_i y ε sigue una distribución normal de media nula y varianza σ . Nótese que ni en el modelo aditivo ni en el parcialmente lineal hay interacción entre las variables estimadas con efectos suaves.

Estos modelos presentan un problema de identificación. Si s_1, \dots, s_M es una solución para el modelo de regresión aditivo, entonces también lo sería $s_1 + c, \dots, s_M - c$. Para evitar esto, las funciones suaves deben cumplir

$$\sum_{t=1}^T s_1(X_{t,1}) = \dots = \sum_{t=1}^T s_M(X_{t,M}) = 0.$$

Además el modelo de regresión aditivo se debe reformular como:

$$Y = \beta_0 + s_1(X_1) + \dots + s_M(X_M) + \varepsilon$$

donde β_0 es el intercepto de la formula de regresión.

Las funciones suaves de estos modelo se pueden estimar de distintas formas, entre ellas se encuentran los splines de suavizado. Para más información de las estimaciones de estas funciones suaves consultar Wood (2017).

Hay múltiples algoritmos utilizados para ajustar estos modelos, entre ellos se encuentran:

- algoritmo Backfitting
- algoritmos basados en la verosimilitud penalizada
- algoritmos REML
- procedimientos bayesianos

Por último, comentar que los modelos aditivos generalizados se utilizan cuando la variable respuesta no sigue una distribución normal. Esto puede ocurrir cuando la variable de estudio es discreta o dicotómica. Si es discreta, puede que esta provenga de un recuento y sigue una distribución de Poisson. En estos casos se utiliza una función link para transformar la variable respuesta y ajustar el modelo aditivo o parcialmente lineal. Para más información sobre estos modelos y los distintos tipos de funciones link existentes consultar Wood (2017).

2.5.4. Modelo de regresión lineal utilizando las transformaciones de las variables de publicidad

Es un modelo muy similar al propuesto por Jin et al (2017). En este modelo se utilizan las transformaciones de las variables de marketing propuestas en la sección 2.3. En esta referencia se ajusta el modelo utilizando inferencia bayesiana y forzando a los coeficientes β de las variables de marketing a tomar valores positivos. En nuestro caso, se crea una malla con posibles valores de los hiperparámetros del modelo. Posteriormente, se ajustan modelos de regresión penalizada, tomando las combinaciones presentes en la malla.

No se conocen los valores exactos que pueden tomar los hiperparámetros de las transformaciones del efecto carryover y shape effect. Sin embargo, en Zhou et al (2023) proponen cotas para los valores que pueden tomar estos según el tipo de variable de marketing al que estén asignados.

- Variables de marketing online: estos anuncios tienen un efecto más inmediato que los anuncios offline. Los clientes potenciales están a un click de acceder a la página web de la empresa. Esto

nos indica que el hiperparámetro θ va a ser pequeño. Este parámetro se encuentra entre 0 y 0.3. Por otro lado, dentro de los parámetros de la función de Hill, se podrían modificar si se tuviera conocimiento previo de la curva respuesta. ABANCA no ha hecho un estudio previamente, por lo que los límites de los hiperparámetros α_m y γ_m no estarán tan restringidos como θ . Se ha supuesto que $\alpha_m \in [0.5, 3]$ y $\gamma_m \in [0.3, 1]$.

- Variables de marketing offline: Una vez publicados estos anuncios, el efecto que tienen sobre la serie de altas es más retardado. Esto se debe a que estos anuncios no tienen una forma de conectar directamente con la empresa, como si pasa en los anuncios online. Esto indica que el hiperparámetro θ asociado a estas variables va a ser más grande. Este parámetro se encuentra entre 0.3 y 0.8. Por otro lado, con los parámetros de la función de Hill ocurre lo mismo que con los anuncios online. No se tiene conocimiento de la curva respuesta, por lo que los hiperparámetros asociados a la función de Hill están menos restringidos. Se ha supuesto que $\alpha_m \in [0.5, 3]$ y $\gamma_m \in [0.3, 1]$.

Se generan secuencias dentro de los límites propuestos y con los valores de las secuencias se crea una malla de posibles valores. Se ajustan modelos de regresión lineal restringida para las combinaciones presentes en la malla y se escoge como modelo final el que presente mejor ajuste.

Cuando los coeficientes asociados a las variables de marketing, β , toman valores negativos, las estimaciones del ROI también pueden serlo y esto es algo poco probable cuando se invierte en marketing. Para solucionar esto se ha optado por utilizar un modelo de regresión lineal restringido cuya fórmula viene dada por:

$$Y_t = \beta_0 + \sum_{m=1}^M \beta_m X_{t,m}^{sat} + \sum_{c=1}^C \delta_c Z_{t,c} + \varepsilon_t \quad (2.11)$$

sujeto a $\beta_m \geq 0 \quad m = 1, \dots, M$

dónde los residuos deben ser de media nula e varianza constante. Los estimadores del modelo de regresión lineal con restricciones son los que dan solución al problema (2.12)

$$\min_{\beta, \delta} \sum_{t=1}^T \left(Y_t - \left(\beta_0 + \sum_{m=1}^M \beta_m X_{t,m}^{sat} + \sum_{c=1}^C \delta_c Z_{t,c} \right) \right)^2, \quad (2.12)$$

sujeto a $\beta_m \geq 0 \quad \forall m = 1, \dots, M$.

Este problema fue planteado por Judge y Takayama (1966) y el estimador viene dado por:

$$\hat{\Omega}_{LR} = \hat{\Omega}_{LS} + (X'X)^{-1}A'(A(X'X)^{-1}A')^{-1}(c - A\hat{\Omega}_{LS}),$$

dónde X es la matriz de diseño del problema, $\hat{\Omega}_{LS}$ es el estimador de mínimos cuadrados, el vector c en nuestro caso es nulo y A se puede observar definida por bloques como:

$$A = \left(\begin{array}{c|c} I_M & 0_{M \times C} \\ \hline 0_{C \times M} & 0_{C \times C} \end{array} \right) \quad (2.13)$$

Si se tiene interés en ver más características de este estimador consultar Judge y Takayama (1996).

2.5.5. Modelo de regresión penalizada utilizando las transformaciones de las variables de publicidad

A lo largo del capítulo, se ha comentado en varias ocasiones que las variables explicativas de estos modelos generalmente presentan coeficientes de correlación bastante elevados. Desafortunadamente,

esto puede ocasionar problemas de colinealidad y sobreajuste. Como posible solución, se han empleado modelos de regresión ridge. Se han añadido las transformaciones de las variables de marketing presentadas en la Sección 2.3. Además, se ha restringido algunos de los coeficientes de regresión de forma que sean positivos.

Para la selección de los hiperparámetros de las transformaciones se ha seguido el procedimiento descrito en 2.5.4. Se genera una malla de posibles valores y posteriormente se ajustan los modelos de regresión ridge restringida para para estos valores. De todos los modelos se escoge el que presente mejor ajuste.

La formulación es la misma que en (2.11). La diferencia con el modelo anterior es que a la hora de estimar las coeficientes de regresión se penaliza la función de mínimos cuadrados:

$$\min_{\beta, \gamma} \sum_{t=1}^T \left(Y_t - \left(\beta_0 + \sum_{m=1}^M \beta_m X_{t,m}^{sat} + \sum_{c=1}^C \delta_c Z_{t,c} \right) \right)^2 + \lambda \left(\sum_{m=1}^M \beta_m + \sum_{c=1}^C \delta_c \right), \quad (2.14)$$

sujeto a $\beta_m \geq 0 \quad \forall m = 1, \dots, M,$

Este problema fue planteado por Toker et al. (2013) y el estimador propuesto viene dado por:

$$\hat{\Omega}_{RR} = \hat{\Omega}_R + (X'X + \lambda I)^{-1} A' (A(X'X + \lambda I)^{-1} A')^{-1} (c - A\hat{\Omega}_R),$$

dónde λ es el parámetro de penalización de la regresión ridge y en nuestro caso, A viene dada por la matriz (2.13) y c es un vector nulo. Si se desea tener más información de las propiedades de este estimador consultar Toker et al. (2013).

2.5.6. Modelo Robyn

El modelo Robyn fue propuesto por primera vez en Zhou et al. (2021). La idea de este es ajustar un Marketing Mix Model seleccionando los hiperparámetros con un algoritmo de optimización.

Este modelo es similar al modelo presentado en la Sección 2.5.5 pero en el se encuentran algunas diferencias. La fórmula propuesta para Robyn es la siguiente:

$$y_t = \beta_0 + \beta_g x_{t,g} + \beta_s x_{t,s} + \beta_h x_{t,h} + \sum_{m=1}^M \beta_m x_{t,m}^{sat} + \sum_{c=1}^C \theta_c z_{t,c} + \varepsilon_t. \quad (2.15)$$

$$(2.16)$$

Esta es muy similar a la fórmula (2.6) pero se pueden identificar nuevas componentes como la variable de tendencia $X_{t,g}$, la de estacionalidad $X_{t,s}$ y la de festivos $x_{t,h}$ y los coeficientes β asociados a estas variables. Se han obtenido utilizando la metodología Prophet presente en el Apéndice A.

El ajuste del modelo de regresión es el mismo que en la Sección 2.5.5, sin embargo la estimación de los hiperparámetros del modelo es diferente. Esta se lleva a cabo utilizando un algoritmo de optimización perteneciente a la familia de algoritmos de evolución diferencial, presentado en el Apéndice B. Esta es una familia de algoritmos iterativos basados en la teoría de evolución de Darwin.

Este algoritmo utilizado minimiza los siguientes criterios de error:

- Normalized Root Mean Square Error (NRMSE), es un criterio de error que se centra en la bondad de ajuste. Este viene definido como

$$NRMSE = \frac{\sqrt{\sum_{i=1}^T (Y_i - \hat{Y}_i)^2 / T}}{\max(Y) - \min(Y)},$$

dónde \hat{Y}_i es el valor predicho por el modelo para la i -ésima observación. Cuánto más pequeño sea el valor del NRMSE mejor se ajustarán los valores ajustados por el modelo a las observaciones originales de la serie de altas.

- DECOMP.RSSD, centrado en el negocio. Este criterio de error provoca que se obtengan resultados demasiado irreales. Antes de introducirlo se deben definir algunos conceptos. Entre estos se encuentra *spend_share*. Este es el vector que contiene los porcentajes de gasto en cada uno de los canales. Viene dado por:

$$spend_share_j = \frac{\sum_{i=1}^n X_{i,j}}{\sum_{i=1}^n \sum_{m=1}^M X_{i,m}}$$

donde $x_{i,j}$ es la inversión en el canal j en la i -ésima semana. Por otro lado, se define *effect_share* como:

$$effect_share_j = \frac{\sum_{i=1}^n \beta_m x_{i,j}^{sat}}{\sum_{m=1}^M \sum_{i=1}^n \beta_m x_{i,m}^{sat}}.$$

Este estima la proporción del efecto de la variable de marketing j sobre el efecto total de los canales publicitarios añadidos al modelo. Una vez definidos estos dos conceptos se define *DECOMP.RSSD* como:

$$DECOMP.RSSD = \sqrt{\sum_{m=1}^M (effect_share_m - spend_share_m)^2}$$

es decir, la distancia euclídea entre los vectores *effect_share* y *spend_share*. Este criterio provoca que se descarten las soluciones más extremas, por ejemplo aquellas en las que haya muchos coeficientes β nulos.

Las variables decisoras utilizadas para este problema de optimización son los hiperparámetros de las funciones de Hill y la transformación carryover. Además se añade como variable decisoras el parámetro λ de la regresión ridge. Estas variables del problema van a tener restricciones, estas son las cotas en la Sección 2.5.4. Por otro lado, el parámetro λ de la regresión ridge tomará como valor mínimo el 0.

Los algoritmos evolutivos toman, de forma aleatoria, las soluciones de las primeras iteraciones. Por lo tanto, para obtener más variedad de resultados, se toman distintos inicios. Cada uno de estos se denomina trial y todos tienen el mismo número de iteraciones. En Zhou et al. (2023) se recomienda tener al menos 5 trials y 2000 iteraciones. Esto implica qel ajuste 10000 modelos de regresión ridge con restricciones.

De todos estos debe escoger un modelo. Para ello se sigue el siguiente procedimiento:

1. Se calcula las fronteras de Pareto, este concepto ha sido introducido en el Apéndice B. Esta contendrá los modelos que tengan mejores valores para el *NRMSE* y el *DECOMP.RSSD*. Normalmente contiene entre 40 y 60 modelos.
2. Se calcula el ROI de los modelos pertenecientes a la frontera de Pareto y se hace un clustering de $k - means$ para agrupar los modelos con estimaciones de ROI similares. Antes se debe fijar el número de clusters que se quieren obtener. Para ello se hace el clustering con los modelos de la frontera de Pareto obteniendo de 1 hasta 30 grupos y se toma el número de clústers que no haya reducido en más de un 5% su valor de *Within Group Sum of Squares*, definido como:

$$WSS = \sum_{k=1}^K \sum_{t=1}^{T_k} (X - \bar{X}_k),$$

dónde T_k es el número de elementos pertenecientes al grupo k y \bar{X}_k es el valor medio del ROI para los modelos .

3. Una vez obtenido los clusters. Se obtiene un modelo representante para cada grupo. Este será el modelo con menos distancia al centroide. Obteniendo así tantos modelos candidatos cómo grupos se hayan obtenido.
4. Tras tener los modelos candidatos se debe escoger uno de ellos. Para ello se pueden tener en cuenta la opinión del departamento de Marketing de la empresa, es imprescindible que opinen de esos modelos cual se ajusta más a la realidad. Para ello deben evaluar las curvas respuesta obtenidas para las variables de marketing y las estimaciones del vector de ROIs. Por otro lado, si no conocen cuál es el modelo con mejor ajuste a la realidad, se puede tomar el modelo con mejor NRMSE.

Una vez hecho esto, para que las hipótesis del modelo sean válidas se debería estudiar si los residuos son de media nula, varianza constante y además si se ha eliminado la dependencia temporal

Capítulo 3

Descripción de las variables

En este capítulo se presentan las variables utilizadas para este problema. Además se harán distintos estudios sobre las mismas. En la Sección 3.1 se definen los aspectos principales de las variables utilizadas en los MMMs de ABANCA. En la Sección 3.2 se ha ilustrado la serie de las altas semanales, además de haber hecho un estudio de sus características estadísticas. Más adelante, en la Sección 3.3, se han analizado individualmente las variables explicativas del problema y para finalizar este capítulo se han estudiado las relaciones que tienen con la variable respuesta y entre ellas.

Además, se han definido una serie de conceptos relevantes de la teoría básica de las series de tiempo, que serán fundamentales para el desarrollo de este capítulo.

3.1. Definición de las variables

En esta sección se presentan los KPIs de interés para ABANCA y posteriormente las variables que se espera que guarden relación con estos.

3.1.1. Variable objetivo

Recordar que se denotan como KPIs a los indicadores de calidad de las empresas. Este trabajo se centra en los KPIs de captación de clientes. Estos son aquellos que están relacionados con el proceso de obtener nuevos clientes.

Dentro del proceso de captación de clientes existen distintas fases:

- Lead: es un cliente potencial. Ha mostrado interés en entablar una relación con la entidad. Son usuarios que han entregado sus datos a la empresa y consecuentemente se han convertido en un registro en la base de datos de esta. Dentro de los leads se encuentran los contactos, que son aquellos que han comenzado un formulario de la web de la empresa o han realizado una llamada telefónica.
- Solicitudes: contacto que muestra un interés relevante sobre el producto a través de una solicitud. Se pueden diferenciar entre solicitudes iniciadas y finalizadas.
- Expediente: solicitud finalizada que inicia un proceso de contratación.
- Formalizaciones: expediente que se ha convertido en una relación contractual con la empresa.
- Alta anticipada: son formalizaciones que abren una cuenta o un ex-cliente que lleva 24 meses sin movimientos y abre una cuenta. A lo largo del trabajo lo denotaremos como alta únicamente.

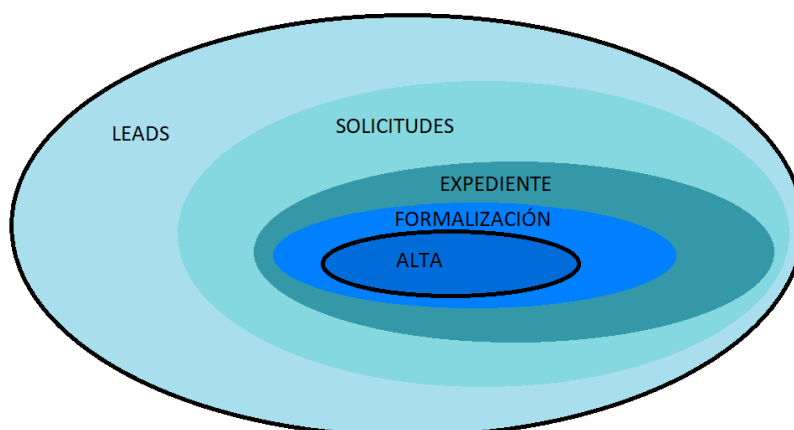


Figura 3.1: Ilustración de la estructura de los niveles de captación.

En vista de lo mencionado, la organización de los niveles de captación es jerárquica, esto es, el proceso de evolución de un cliente respecto a su relación con la empresa es progresivo y exige la completitud de todas las etapas intermedias. Por ejemplo, nadie puede ser un alta anticipada sin ser antes un lead, solicitud, expediente o una formalización. Por tanto el volumen de leads debe ser mayor o igual que el de volumen de solicitudes, el de solicitudes mayor que el de expedientes y así de forma sucesiva. En la Figura 3.1 se puede apreciar una ilustración de la estructura de los niveles de captación.

Cabe destacar que cuanto más se avanza en los niveles de captación de clientes, las diferencias entre los grupos se van minimizando. Es decir, la diferencia de leads y de solicitudes ha sido mucho mayor que la de formalizaciones y altas donde apenas se pierden clientes potenciales.

Se ha estudiado el número de nuevas altas del banco de forma semanal para ello se han utilizado las distintas actividades publicitarias de la entidad. Destacar que el número de altas no sólo se ve afectado por las variables de marketing del banco, sino que también por otras variables relacionadas con el contexto en el que se han desarrollado estas actividades publicitarias.

3.1.2. Variables explicativas

Como se define en la Sección 1.3, las variables explicativas se diferencian en dos grupos, las de marketing y las de contexto.

Dentro de las variables de marketing se disponen de variables de marketing online y offline. Las primeras son todas aquellas variables relacionadas con las actividades publicitarias en soportes digitales. En el marketing online se pueden utilizar distintas variables, entre las que se encuentran:

- Inversión: dinero gastado en ese tipo de anuncio a lo largo de una semana.
- Impresiones: impacto publicitario ocasionado en el usuario al ver un anuncio en línea, ya sea en vídeo, audio o cualquier otro formato digital.

- Clicks: cada vez que un usuario hace click en un formato publicitario accediendo a su contenido digital. Es una métrica basada en la interacción con el usuario.

Contenido eliminado por temas de confidencialidad

Por otro lado se encuentran los anuncios offline. En ABANCA se han recogidos los datos de los anuncios de televisivos. Como ocurría con las variables de marketing online, esta variable también tiene una medida de exposición, los GRPs (Gross Rating Points). Estos son indicadores del número total de impactos conseguidos por cada cien personas del público objetivo a lo largo de un periodo de tiempo determinado. Es una medida de gran importancia en cuanto a anuncios de televisión debido a que muchas veces los precios de anuncios de televisión dependen directamente de los GRPs. Entre los factores que afectan a su valor podemos encontrar: el canal televisivo en el que se publica el anuncio, la duración, la posición en la que se encuentra dentro del parón publicitario (el primero y el último anuncio generalmente son los más caros), programa en el que se publicita, etc...

Contenido eliminado por temas de confidencialidad

Por último, se introducen brevemente las variables de contexto. Estas permiten entender en qué entorno se encuentra el cliente cuando ve el anuncio.

En los Marketing Mix Models por lo general se utilizan variables macroeconómicas como variables de contexto. Pero también es recomendable utilizar otras como la actividad de la competencia o en el caso de trabajar con ventas se podrían añadir ofertas o descuentos. Las más interesantes para este proyecto son:

- PIB (Producto Interior Bruto): cantidad que mide el valor monetario total de los bienes y servicios finales producidos para el mercado, dentro de las fronteras de un país, en un año dado.
- IPC (Índice de Precios de Consumo): indicador que mide la variación de los precios de una cesta de bienes y servicios en un lugar concreto durante un determinado periodo de tiempo.
- Número de trabajadores en alta en la seguridad social.
- Población en edad laboral residente en España: población española en el rango de edad de 25 a 64 años.
- Euribor a 12 meses: tipo de interés a la que un grupo de bancos europeos prestan dinero entre sí con vencimiento a los 12 meses.

Comentar que todas las variables de contexto se han obtenido de la página web del Ministerio de Economía (2023).

Contenido eliminado por temas de confidencialidad

3.2. Serie de altas semanales

ABANCA está muy interesado en captar más leads y posteriormente convertirlos en altas. Se ha optado por ajustar MMMs que modelen el número de altas de la empresa.

Contenido eliminado por temas de confidencialidad

A continuación, se estudian las propiedades estadísticas de esta serie de tiempo. Pero antes, se definen los conceptos de proceso estocástico y serie de tiempo.

Definición 3.2.1 (Proceso estocástico). Un proceso estocástico es un conjunto de variables aleatorias $\{X_t\}_{t \in C}$ definidas sobre el mismo espacio de probabilidad

Nos centraremos en un proceso estocástico $\{X_t\}_{t \in \mathbb{N}}$, siendo \mathbb{N} el conjunto de números naturales. Donde t representa el instante temporal.

Definición 3.2.2 (Serie de tiempo). Una serie temporal es un proceso estocástico con conjunto de índices \mathbb{N} y que toma valores en el conjunto de números reales

Para estudiar las características de las series de tiempo, es imprescindible estudiar sus funciones de autocovarianzas.

Definición 3.2.3 (Función de autocovarianzas). La función de autocovarianzas se define como:

$$\gamma_x(s, t) \text{Cov}(X_s, X_t)$$

para todo $s, t \in T$, siendo T el número de observaciones de la serie de tiempo.

Esta función da lugar a las funciones de autocorrelaciones simples y parciales.

Definición 3.2.4 (Función de autocorrelaciones simples (AFC)). Se define la función de autocorrelaciones simples como:

$$\rho(s, t) = \frac{\gamma(s, t)}{\sigma_s \sigma_t}, \quad (3.1)$$

dónde $\gamma(s, t)$ es la covarianza entre el retardo s y el retardo t , mientras que σ_s y σ_t son las desviaciones típicas en los retardos s y t , respectivamente.

Definición 3.2.5 (Función de autocorrelaciones parciales (PACF)). La función de autocorrelaciones parciales se define como:

$$\alpha(s, t) = \frac{\text{Cov}(X_s - \hat{X}_s^{(s,t)}, X_t - \hat{X}_t^{(s,t)})}{\sqrt{\text{Var}(X_s - \hat{X}_s^{(s,t)}) \text{Var}(X_t - \hat{X}_t^{(s,t)})}} \quad (3.2)$$

donde $\hat{X}_j^{(s,t)}$ denota el mejor predictor lineal de X_j construido a partir de las variables medidas en los instantes comprendidos entre s y t . Esta función mide el grado de dependencia lineal entre X_s y X_t .

Estas funciones permiten medir la serie en distintos instantes de tiempo en función del retardo. Cabe destacar que estas funciones son complicadas de estimar, pero cuando la serie es estacionaria, esta tarea es mucho más sencilla.

Una vez hecho esto, se puede definir el concepto de estacionaridad de una serie de tiempo.

Definición 3.2.6 (Proceso estacionario). Un proceso estocástico es débilmente estacionario cuando cumple las siguientes condiciones:

- $\mathbb{E}[X_t] = \mu \forall t = 1, \dots, T$
- $Var(X_t) = \sigma^2 \forall t = 1, \dots, T$
- $Cov(X_t, X_{t+k}) = \gamma_k \forall t, k \in T$

Esto quiere decir que una serie de tiempo es estacionaria cuando tiene estabilidad en la media, la varianza y las autocovarianzas. Las estimaciones de las ACF y las PACF para series de tiempo estacionarias se pueden encontrar en Shumway y Stoffer (2017).

Observando las funciones de autocorrelaciones simples y parciales podemos determinar si la serie estudiada presenta tendencia temporal. Si hay muchos valores de la función de autocorrelación simple con valores grandes, se puede determinar que la serie no es estacional.

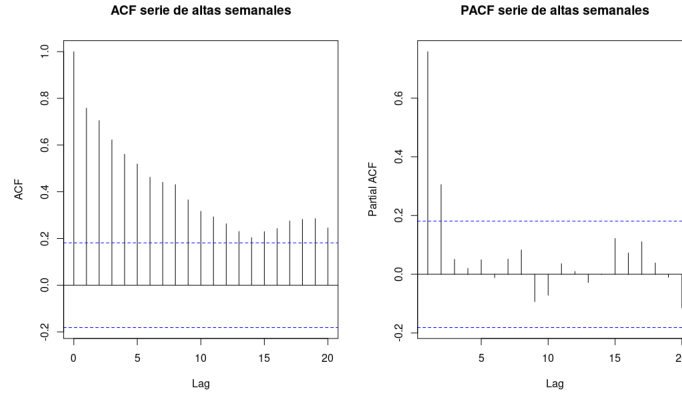


Figura 3.2: Funciones de autocorrelaciones simples y parciales muestrales para la serie de tiempo de las altas semanales.

En la Figura 3.2 se puede observar la estimación de las ACF y PACF de la serie de altas semanales. La primera presenta muchos valores no nulos, lo que implica que existe una tendencia en la serie de tiempo. Esto es algo que ya se observaba en la Figura ?? y que provoca que nuestra serie de tiempo sea no estacionaria.

3.3. Variables explicativas

Las variables explicativas en los MMMs se distinguen en dos grupos. Por un lado, se tienen las variables explicativas de marketing o variables publicitarias. Estas se corresponden con todos aquellos datos que tienen relación con las actividades publicitarias de ABANCA. Se caracterizan porque se han tomando decisiones sobre los valores que han tomado cada semana y son las protagonistas de la fase de optimización del presupuesto.

Por otro lado, se encuentran las variables explicativas de contexto. Estas son aquellas sobre las que no se ha tenido decisión. Entre ellas se encuentran variables macroeconómicas como el PIB o el IPC. Además de estas, se ha utilizado una variable que mida la actividad de la competencia en medios publicitarios.

3.3.1. Variables explicativas de publicidad

Contenido eliminado por temas de confidencialidad

3.3.2. Variables de contexto

Estas variables son aquellas sobre las que ABANCA no tiene decisión. Como se comentó en la sección 3.1, la gran mayoría son factores macroeconómicos, aunque también se pueden encontrar factores como la estacionalidad o la actividad de la competencia.

Estos se han obtenido en la base de datos del Ministerio de Economía. De esta web se han tomado los datos del PIB, el IPC, el número de trabajadores afiliados a la seguridad social, la población en edad laboral y el euribor.

Muchas de estas series presentan una frecuencia temporal distinta a la serie de altas. Por ejemplo, los datos del PIB son trimestrales y los del IPC mensuales. Todas las variables añadidas al modelo deben presentar la misma frecuencia temporal. Para trabajar con todas las series de tiempo, deben de tener la misma frecuencia temporal. Para obtener estas series de tiempo en la misma frecuencia se utiliza la función `td` de la librería `tsdiag`.

Esta función realiza una regresión en la que la variable respuesta es la serie de baja frecuencia (la de contexto en nuestro caso) y las variables explicativas son indicadores, que en nuestro caso son inexistentes. La información de esta función se puede encontrar en Sax y Steiner (2013).

Permite desagregar la serie de tiempo en distintas frecuencias temporales. Lamentablemente, entre las opciones que nos aporta la función no se encuentra la opción semanal. Como solución a este problema se han desagregado las series a datos diarios y posteriormente se ha obtenido la media de los datos pertenecientes a cada semana. Obteniendo así datos semanales.

Existen distintas opciones para el método de desagregación, se ha seleccionado la opción **average**. De esta forma, la media de los valores desagregados correspondientes a cierto periodo de tiempo se corresponden con el valor en ese periodo. Por ejemplo, el primer mes del IPC tiene un valor de 97.583 y su desagregación es la siguiente:

```
[1] 97.71366 97.72103 97.72675 97.73080 97.73318 97.73390 97.73296
[8] 97.73035 97.72607 97.72014 97.71254 97.70327 97.69234 97.67975
[15] 97.66549 97.64957 97.63198 97.61273 97.59182 97.56924 97.54499
[22] 97.51909 97.49152 97.46228 97.43138 97.39882 97.36459 97.32870
[29] 97.29114 97.25192 97.21103
```

y la media de estos días es 97.583 que es coincidente con el valor original de la serie en ese mes.

Destacar que al tomar la media de los valores pertenecientes a cada semana se pierde esta característica de la función `td`. Esto se debe a que existen semanas que contienen días de meses distintos. Por ejemplo, la semana del 29 de mayo y el 1 de junio de 2023 es la misma pero son días que pertenecen a meses diferentes.

En la Figura 3.3 se compara la serie de tiempo del IPC desagregada y la serie de tiempo con datos originales. Se observa que la serie de tiempo con los datos desagregados es una con subidas y bajadas menos pronunciadas que la original. Esta serie presenta una tendencia ascendente hasta el mes de junio

de 2022, donde parece que estabilizarse.

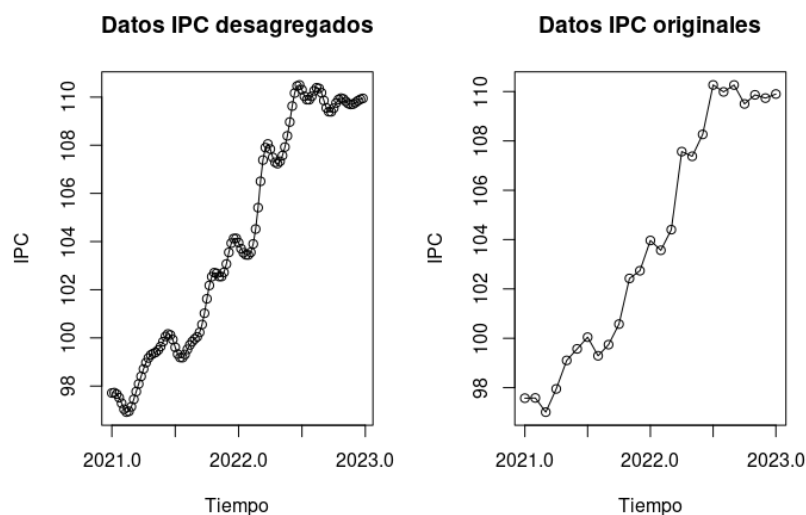


Figura 3.3: Desagregación temporal de la serie mensual del IPC.

El proceso de desagregación temporal se ha repetido para la serie de tiempo del número de trabajadores afiliados a la seguridad social. Esta serie de tiempo también tiene frecuencia mensual, por lo que el proceso es exactamente igual al realizado para la serie temporal del IPC.

Una vez desagregados las series mensuales se ha seguido un procedimiento similar con las series trimestrales. Esto implica que cada cuatrimestre se vea desagregado en 91 días, que posteriormente se agrupan en 13 semanas con un procedimiento similar al realizado para los datos trimestrales. En la Figura se puede observar la desagregación de la serie de tiempo de los trabajadores afiliados a la seguridad social, la del PIB y la de la población entre 24 y 65 años. Las series de número de trabajadores afiliados a la seguridad social y del PIB parecen tener una tendencia ascendente. Por otro lado, la serie de la población tiene un bajón claro a comienzos de la serie y a finales de esta parece volver a aumentar considerablemente.

Se han presentado todas las variables macroeconómicas menos la del euribor. Destacar que existen muchos tipos de Euribor que van medidos según vencimiento, en este caso se ha tomado el euribor a 7 días. Estos datos son diarios, por lo que se puede prescindir de la función td . Se toma la media de los días pertenecientes de cada semana como valor de la serie. Esta se puede observar en la Figura 3.5

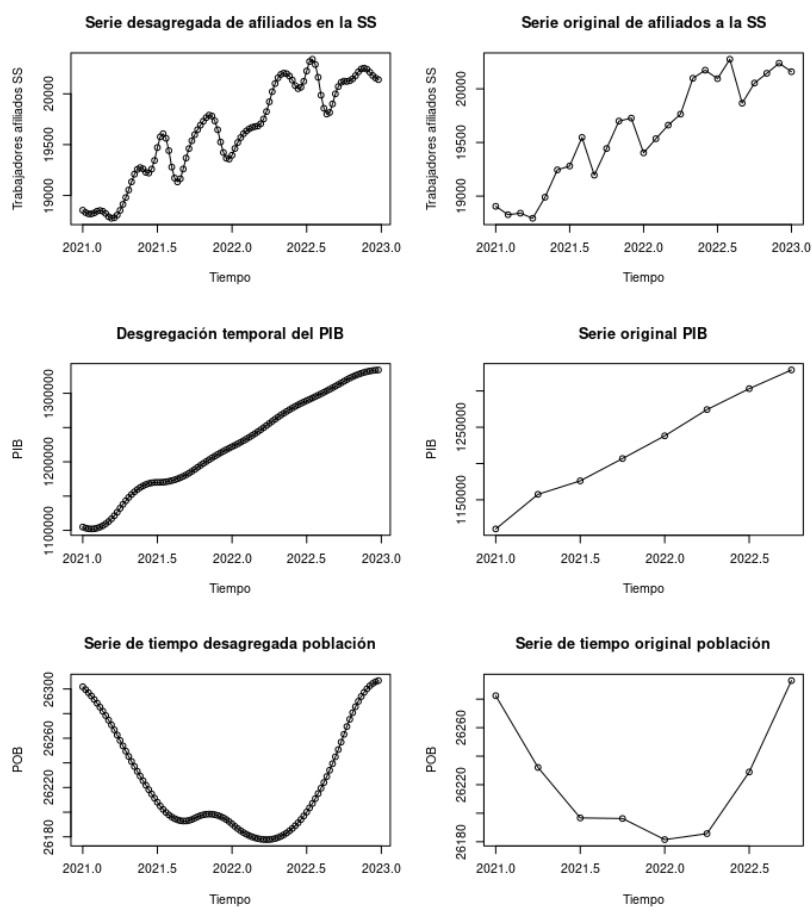


Figura 3.4: Desagregación temporal número de trabajadores afiliados en alta en la seguridad social, el PIB y la población entre 25 y 64 años. Las unidades de estas gráficas son miles en el caso de los trabajadores en alta en la seguridad socual y en la población. Por otro lado la unidad en la serie del PIB son los millones de euros. En las gráficas de la izquierda podemos observar la serie de tiempo desagregada y en la de la derecha las series originales.

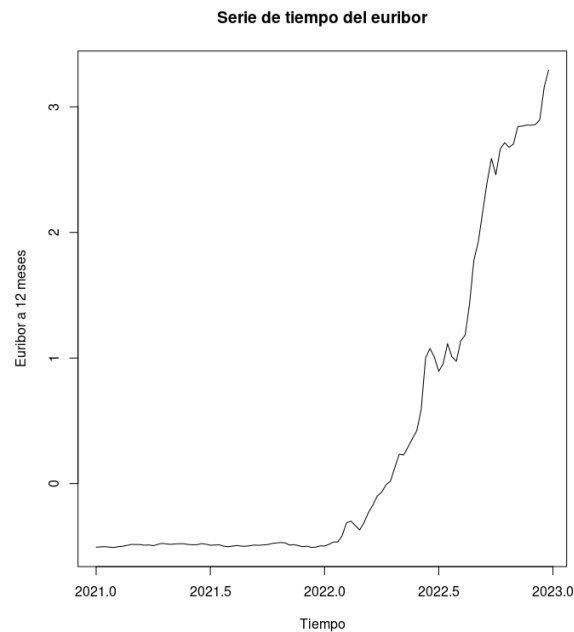


Figura 3.5: Serie de tiempo de los datos del euribor.

Destacar que la serie del euribor era estable hasta poco después del comienzo de 2022, donde debido a la guerra de Ucrania se dispara, haciendo que cada vez sea más elevado el valor de esta serie. A finales de 2022 alcanza su valor máximo en el periodo de estudio con un 3.2925 %.

Contenido eliminado por temas de confidencialidad

Para acabar esta sección se busca la relación que tienen entre sí las variables de contexto, pues es de esperar que algunos de los factores macroeconómicos que se han presentado tengan una correlación bastante elevada. Para ello se muestra en la Figura ?? la matriz de correlaciones de las variables de contexto. Se denota por PIB a la serie del Producti Interior Bruto, IPC a Índice Precio de Consumo, AFIL_SEG al número de trabajadores afiliados a la seguridad social, POB a la población entre 24 y 65 años, EURIBOR al euribor con vencimiento en 7 días y COMP_TOTAL a los GRPs emitidos por la competencia.

Las correlaciones entre las variables de contexto son muy elevadas. Hay siete coeficientes de correlación por encima del 0.7 y tres de ellos por encima del 0.9. El valor más elevado es el de la correlación entre el PIB y el IPC teniendo entre ellos un coeficiente de 0.98. La variable que tienen coeficientes de correlación más pequeños es la de la población, cuyo coeficiente más elevado es con el euribor con una correlación de 0.48. La variables de la actividad de la competencia también tiene coeficientes bajos, siendo el más elevado 0.36 con el euribor.

Los valores tan elevados de correlación entre el PIB, IPC, el número de trabajadores afiliados a la seguridad social y el euribor nos podría traer problemas de colinealidad a la hora de ajustar un MMM. Por lo tanto sólo se añadirán algunas de estas a los modelos.

Figuras eliminadas por confidencialidad

Figura 3.6

3.4. Relación de la variable objetivo con las variables explicativas

Una vez estudiadas las correlaciones entre las variables de contexto y las de gasto por separado, es necesario realizar un estudio de la correlación de estas con la variable de altas semanales. Además, se aprovecha esta sección para estudiar la correlación entre las variables de gasto y las de contexto. Analizando así si el modelo podría presentar problemas de colinealidad.

El primer paso a tomar es observar la matriz de correlación de todas las variables en conjunto. Esta se puede observar en la Figura 3.6

Debido al número tan elevado de coeficientes de correlación se optó por una figura más ilustrativa de los coeficientes de correlación. En tonos azulados están representados aquellos que tienen valores positivos. Por otro lado, los valores con tonos rojizos son los que tienen correlación negativa. Además, cuanto más elevado sea el valor absoluto del coeficiente, más grande está escrito el número. El coeficiente de correlación más elevado que con la variable de altas semanales lo tiene la variable de contexto del PIB con un valor de 0.66. Cabe destacar que el gasto en televisión tiene un coeficiente de correlación de 0.64. Por otro lado, hay variables que no parecen tener una correlación significativa con la serie de altas semanales. Los anuncios display tienen un coeficiente de correlación de 0.16 con la serie de altas semanales y la población entre 25 y 64 tiene un coeficiente de correlación de -0.08.

Parece que va a haber correlación entre algunas variables de contexto y de marketing. El gasto en anuncios en buscadores tiene un coeficiente de correlación bastante elevado con todas las variables de contexto menos con la de población y de competencia. Destacar que la variable de inversión en anuncios de programación tiene un coeficiente de correlación negativo con todas las variables menos con la de inversión en anuncios tipo display y en redes sociales.

Capítulo 4

Resultados estadísticos

En este capítulo se presentan los principales resultados de los modelos formulados en la Sección 2.5 para los datos descritos en el Capítulo 3. Por ello, es importante recordar que estos modelos se utilizan para explicar el número de altas del banco según los gastos de inversión, tanto en publicidad televisiva, como online. Además, se han incorporado variables de contexto que, históricamente, han sido cruciales para entender el transcurso de las decisiones económicas y de marketing. Finalmente, se realizan tareas de diagnóstico, se calculan medidas de bondad de ajuste y se visualizan los resultados.

Con esta idea en mente, se dispone de datos desde el 1 de enero de 2021 hasta el 31 de diciembre de 2022. En concreto, se ha trabajado con datos agregados semanalmente, lo que resulta en 104 instantes temporales. A parte de las variables de contexto del Capítulo 3, se ha prescindido de variables económicas como el IPC, el número de trabajadores afiliados a la seguridad social y el euribor. De hecho, la alta correlación presente entre las series de tiempo asociadas a la inversión y las variables de marketing justifica la exclusión de las primeras.

4.1. Modelo lineal

El modelo lineal nos aportará una idea de las variables que tendrán un efecto más relevante a la hora de estimar el número de altas del banco.

A continuación, se ajusta el modelo lineal y se presenta un resumen del mismo.

```
> summary(ajuste_lineal)
```

Call:

```
lm(formula = alt ~ BUSC + COMP + DISP + PROG +  
SN + VOL + TV + PIB + POB + COMP.TOTAL, data = dt_input)
```

Residuals:

Min	1Q	Median	3Q	Max
-691.44	-110.05	9.13	116.38	567.84

Coefficients:

	Estimate	Std. Error	t value	Pr(> t)
(Intercept)	6.796e+03	2.476e+04	0.275	0.78430
BUSC	-1.421e-02	6.719e-03	-2.115	0.03714 *
COMP	9.668e-04	1.488e-02	0.065	0.94833
DISP	5.384e-04	2.001e-03	0.269	0.78844

```

PROG      -1.571e-02  6.205e-03  -2.533  0.01300 *
SN         5.056e-03  1.268e-02   0.399  0.69087
VOL        -4.108e-03  1.332e-02  -0.308  0.75849
TV          1.445e-03  3.162e-04   4.570  1.5e-05 ***
PIB         2.974e-03  1.042e-03   2.853  0.00533 **
POB        -3.213e-01  9.120e-01  -0.352  0.72543
COMP.TOTAL  1.015e-01  6.212e-02   1.634  0.10574

```

```
---
```

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Residual standard error: 230.5 on 93 degrees of freedom
```

```
Multiple R-squared:  0.6384, Adjusted R-squared:  0.5995
```

```
F-statistic: 16.42 on 10 and 93 DF,  p-value: < 2.2e-16
```

Se puede observar que la mayoría de los coeficientes asociados a las variables explicativas tienen valores muy pequeños, algunos incluso son no significativos. Como solución a este problema se utiliza la función *step* para seleccionar las variables del modelo. Obteniendo el siguiente ajuste:

```
> summary(step(ajuste_lineal,direction = c("backward"),trace = F))
```

```
Call:
```

```
lm(formula = alt ~ BUSC + PROG + TV + PIB + COMP.TOTAL,
    data = dt_input)
```

```
Residuals:
```

```

Min      1Q  Median      3Q      Max
-695.44 -114.43   17.39  111.95  571.25

```

```
Coefficients:
```

```

              Estimate Std. Error t value Pr(>|t|)
(Intercept) -1.901e+03  6.676e+02  -2.848  0.00536 **
BASC         -1.572e-02  5.164e-03  -3.044  0.00300 **
PROG         -1.327e-02  4.188e-03  -3.168  0.00205 **
TV            1.427e-03  2.337e-04   6.109 2.03e-08 ***
PIB           3.136e-03  5.800e-04   5.408 4.51e-07 ***
COMP.TOTAL    1.058e-01  5.992e-02   1.766  0.08051 .

```

```
---
```

```
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1
```

```
Residual standard error: 225 on 98 degrees of freedom
```

```
Multiple R-squared:  0.6369, Adjusted R-squared:  0.6184
```

```
F-statistic: 34.39 on 5 and 98 DF,  p-value: < 2.2e-16
```

El modelo final solo utiliza las variables de marketing relativas al gasto en buscadores, programación y en televisión. Destacar que el valor de β_0 es negativo, este parámetro se podía interpretar como las altas base del banco, no tiene sentido que su estimación sea negativa.

Dentro de las variables de contexto, se ha estimado que el PIB y la variables de la competencia tienen un efecto positivo en el número de altas semanales. Esto último no tiene demasiado sentido, lo más normal y lógico es que la actividad de la competencia provoque una bajada en el número de clientes que llegan al banco.

Todos los coeficientes de regresión son significativos en un nivel del 10% de significación. Y todos menos el de la competencia lo son a un nivel del 5%. Por último comentar que el R^2 ajustado de este

modelo tiene un valor de 0.618.

Para poder obtener conclusiones de este modelo, primero se deben comprobar las hipótesis mencionadas en la Sección 2.5.1, además comprobar si la serie de residuos es . En la figura 4.1 se pueden observar distintos gráficos de los residuos de este modelo. Los de la parte superior se corresponden con las ACF y PACF de la serie de residuos y se observan valores no nulos, lo que nos indica que no se ha eliminado la dependencia temporal. Se puede observar que no parece haber heterocedasticidad en los residuos en el gráfico inferior izquierdo, para contrastarlo se lleva a cabo el test de Breusch-Pagan que obtuvo en p -valor de 0.74, por lo que no hay evidencias para rechazar la homocedasticidad de estos. Por otro lado, en el gráfico inferior derecho se muestra la estimación de la densidad tipo núcleo junto con la de la normal máximo verosímil. Estos no son demasiado similares, se puede observar que la densidad tipo kernel presenta colas más pesadas. Se utilizó el contraste de Shapiro-Wilks para contrastar si la distribución de los residuos era gaussiana, siendo este test rechazado con un p -valor de 0.003. Por último, se comprobó que los residuos tenían media igual a 0 utilizando el test de la t -Student obteniendo un p -valor muy próximo a 1.

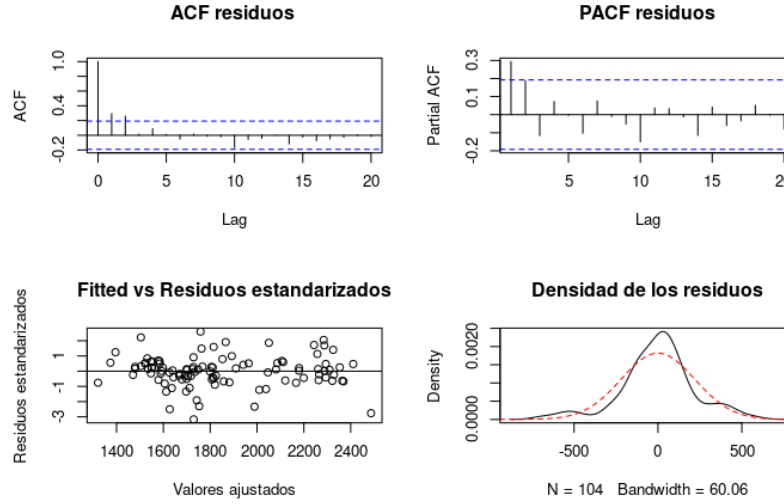


Figura 4.1: Análisis de los residuos del modelo de regresión lineal múltiple.

Además de esto, se debe cumplir que los residuos sean ruido blanco. Para comprobar esto, se utilizó el test de Ljung-Box para contrastar si las observaciones de estos residuos presentan correlación entre ellas. Es decir, dadas dos observaciones X_r y X_{r+k} se contrasta si ρ_k es significativo. Para ello se utiliza el siguiente estadístico:

$$Q_r = T(T+2) \sum_{k=1}^r \frac{\hat{\rho}_k^2}{T-k}$$

donde T es el número de observaciones y r es el número de retardos a contrastar. Asumiendo que T es grande se tiene que $Q_r \approx \chi_r^2$. Por tanto se rechaza el contraste si el estadístico Q_r es mayor que el percentil 0.95 de la distribución χ_r^2 . Se realiza el contraste de Ljung-Box para $r = 1, \dots, 10$, siendo rechazado en todos los casos. Esto implica que los residuos no son independientes entre sí y por tanto no son ruido blanco.

4.2. Modelo de regresión penalizada

Como se menciona en la Sección 2.5.2, los modelos de regresión penalizada son una buena elección cuando se trabaja con variables con correlación elevada. El problema planteado presenta este problema, por lo que a priori estos modelos son una buena elección. Los modelos de regresión penalizada más comunes se presentaron en la Sección 2.5.2, de todos ellos se optó por trabajar con un modelo de regresión ridge.

Una de las grandes cuestiones de estos modelos es la selección del parámetro de penalización λ . En particular, a medida que λ aumenta, menores serán las estimaciones de los coeficientes de regresión. La función `cv.glmnet` permite optimizar λ en términos de error cuadrático medio. En nuestro caso, $\hat{\lambda} = 325.18$.

A continuación, se visualizan los resultados de la estimación de los coeficientes de regresión.

```
> coef(fit.ridge)
11 x 1 sparse Matrix of class "dgCMatrix"
s1
(Intercept)  1.959379e+04
BUSC         1.740533e-03
COMP         1.099110e-02
DISP         1.033569e-04
PROG        -8.650647e-03
SN           4.047088e-03
VOL          7.365300e-03
TV           6.400635e-04
PIB           8.186507e-04
POB          -7.190771e-01
COMP.TOTAL   7.364800e-02
```

Cabe destacar que todos los coeficientes asociados a las variables de marketing son positivos, excepto aquel relacionado con los anuncios en programática. Por otro lado, los coeficientes asociados al PIB y a la actividad de la competencia tienen valores positivos. Sin embargo, el coeficiente asociado a la variable de la población es menor que 0. Destacar que no tiene demasiado sentido que invertir en anuncios en programática implique que se pierden altas.

Una vez ajustados estos modelos, se deben de comprobar las hipótesis de los residuos. Para ellos en la Figura 4.2 se muestran distintos gráficos de esta serie. Destacar que con el ajuste del modelo no se ha conseguido eliminar la dependencia temporal. Se observan valores de las ACF y PACF no nulos. Por otro lado, no parece haber evidencias de homocedasticidad y los residuos no parecen ser normales. El contraste de media nula obtuvo un p -valor muy próximo a 1. Por otro lado, el contraste de Shapiro-Wilks se rechazó con un p -valor de 0.009. Por último, se han realizado contrastes de Ljung-Box sobre la serie de residuos para $r = 1, \dots, 10$ y el test fue rechazado en todos los casos.

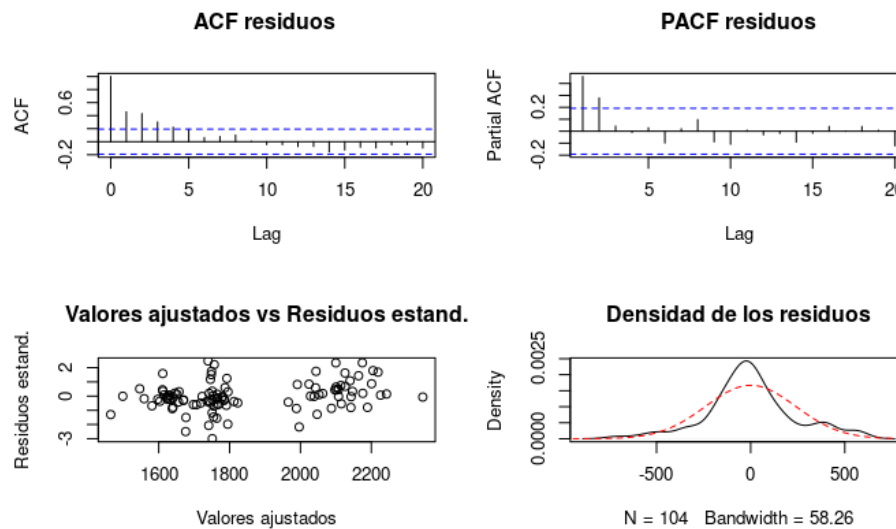


Figura 4.2: Análisis de los residuos del modelo de regresión Ridge.

4.3. Modelo semiparamétrico

Los modelos paramétricos no han presentado un buen ajuste sobre la serie de altas semanales. Por ello se prueban los modelos semiparamétricos. Estos permiten ajustar algunas variables con efectos no lineales mientras que otras se ajustan con efectos lineales. En la fórmula (2.6) se puede observar que se ajustan las variables de marketing con funciones que captan los efectos no lineales. Por otro lado, en esa misma ecuación se fijan efectos lineales para las variables de contexto.

Se han ajustado los efectos suaves utilizando thin plate splines. El primer modelo ajustado contenía todas las variables que se utilizaron para ajustar el modelo de regresión ridge. Sin embargo al ajustar el modelo semiparamétrico muchos de los efectos, tanto suaves como lineales eran no significativos. Se han ido eliminando una a una las variables no significativas siguiendo un procedimiento similar al de la función `step`. El criterio elegido para comparar estos modelos ha sido el AIC. El modelo con mejor valor ha sido el que contenía a las variables de gasto en televisión y en anuncios en buscadores por parte de las variables de marketing. Por otro lado, las variables de contexto presentes en el modelo son las de PIB y la actividad de la competencia.

```
> mod_gam1 = gam(alt ~ s(BUSC_spend,bs="tp")+s(TV_IMP,bs="tp")+PIB+COMP.TOTAL,
  data = dt_input)
> summary(mod_gam1)
```

```
Family: gaussian
Link function: identity
```

```
Formula:
alt ~ s(BUSC_spend, bs = "tp") + s(TV_IMP, bs = "tp") + PIB +
COMP.TOTAL
```

```
Parametric coefficients:
              Estimate Std. Error t value Pr(>|t|)
(Intercept) -2.125e+03  7.729e+02  -2.750  0.00715 **
```

```

PIB          3.097e-03  6.232e-04  4.969 2.98e-06 ***
COMP.TOTAL   9.962e-02  5.681e-02  1.754 0.08273 .
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Approximate significance of smooth terms:
              edf Ref.df      F p-value
s(BUSC)       2.753  3.472  5.051 0.00124 **
s(TV)         3.605  4.413 10.948 < 2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

R-sq.(adj) = 0.676   Deviance explained = 70.2%
GCV = 47217   Scale est. = 42969      n = 104

```

Contenido eliminado por temas de confidencialidad

Los residuos del modelo parece que presentan mayor varianza cuando la estimación es pequeña, sin embargo esto no es demasiado claro. Se realizó el contraste de Breusch-Pagan para comprobar si estos son heterocedásticos, obteniendo un p -valor de 0.33, por tanto no hay evidencias estadísticas para rechazar la homocedasticidad de los residuos. Por otro lado, se puede observar la estimación tipo kernel de la densidad junto con la normal máximo verosímil de los residuos. La estimación kernel de la densidad presenta una ligera asimetría y una kurtosis bastante elevada. Se realizó un contraste de media nula en el que no se rechazó la hipótesis nula. Sin embargo, en el test de Shapiro-Wilks fue rechazado por lo que se descarta la normalidad de los residuos. Por último, se realizaron los contrastes de Ljung-Box para $r = 1, \dots, 10$ siendo rechazados en todas las ocasiones.

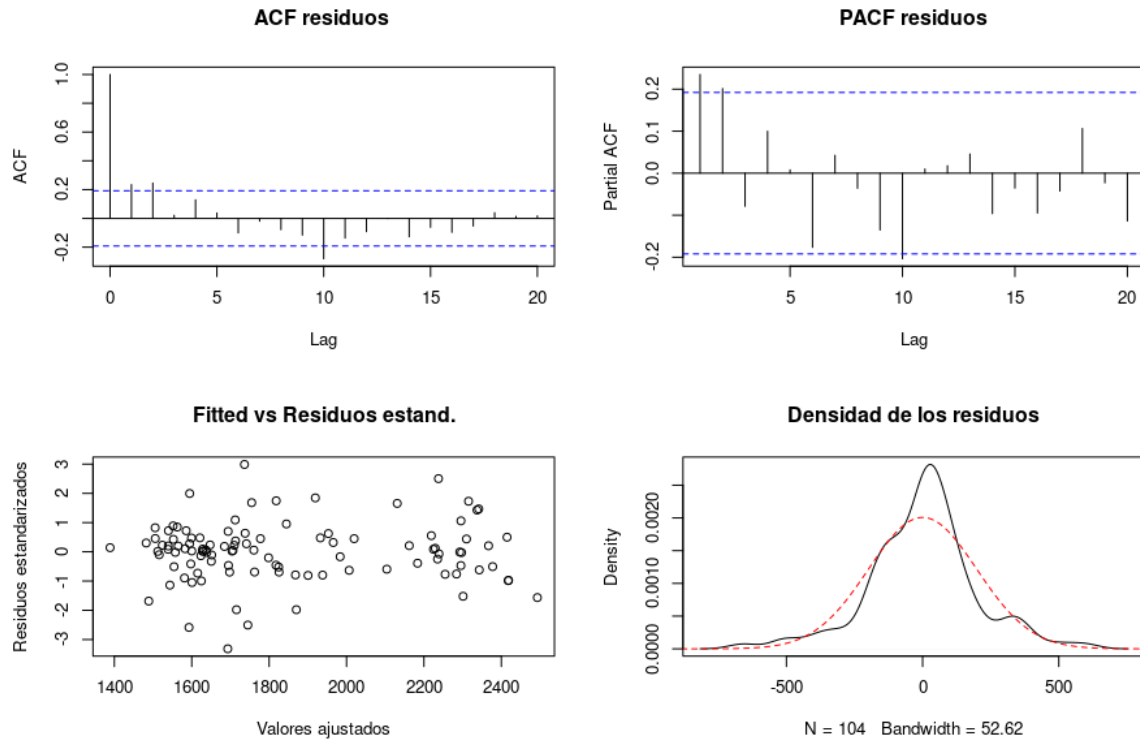


Figura 4.3: Análisis de los residuos del modelo semiparamétrico.

Al rechazarse los residuos de Ljung-Box se puede determinar que los residuos no son ruido blanco, por lo que no se cumplen las hipótesis de los MMMs utilizando este modelo.

4.4. Modelo de regresión lineal con transformaciones

Jin et al. (2017) propone utilizar las transformaciones de *carryover* y *shape effect* para ajustar los MMMs. Estas dependen de hiperparámetros como se mencionó en la Sección 2.3. ABANCA no dispone de información sobre los posibles valores de estos. Sin embargo, en la Sección 2.5.4 se presentaron cotas para estos según las propuestas por Zhou et al. (2023).

En la Sección 2.5.4, se explicó el procedimiento seguido para ajustar este modelo de regresión. La malla generada contiene 5000 posibles combinaciones de los hiperparámetros. Para ajustar el modelo de regresión lineal con restricciones se ha utilizado la función `glmnet` de la librería `glmnet`, donde se ha indicado que el parámetro de penalización λ tiene valor nulo. Además, con las opciones `lower.limits` y `upper.limits` se pueden indicar cotas para los coeficientes de regresión del modelo. En este caso se indicó que el límite inferior es 0 en las variables de marketing. Destacar que al añadir las transformaciones de las variables de marketing, se pueden obtener estimaciones de la curva respuesta (2.7) para estas.

Una vez ajustado el modelo para cada una de las combinaciones de la malla, el mejor de los ajustes obtuvo un valor del $RMSE$ de 223.37 y el del R^2 fue de 0.6202. Se ha mejorado ligeramente el ajuste del modelo lineal sin transformaciones, aunque la mejora no es demasiado significativa. En la Tabla 4.1 se muestra la combinación de hiperparámetros que obtuvo el mejor ajuste. Se puede observar que algunas variables de marketing tienen estimaciones muy similares, como los anuncios tipo display y los de vídeos-online.

Figuras eliminadas por confidencialidad

Figura 4.4

	$m=\text{BUSE}$	$m=\text{COMP}$	$m=\text{DISP}$	$m=\text{PROG}$	$m=\text{SN}$	$m=\text{VOL}$	$m=\text{TV}$
α_m	2.7	0.5	2.7	1.6	2.6	0.6	0.9
γ_m	0.7	0.9	0.6	1	0.6	0.8	0.3
θ_m	0.15	0.2	0.05	0.1	0.1	0.2	0.3

Tabla 4.1: Combinación de hiperparámetros que proporcionó un mejor ajuste para el modelo de regresión lineal con transformaciones.

Los coeficientes de regresión del modelo que presentó un mejor ajuste son los siguientes:

```
> coef[ind,]
Intercepto  BUSE      COMP      DISP      PROG      SN
2.493537e+04 0.000000e+00 0.000000e+00 0.000000e+00 0.000000e+00 0.000000e+00
VOL         TV       PIB       POB       COMP.TOTAL
2.315639e+02 7.458129e+02 1.017764e-03 -9.049588e-01 6.339265e-02
```

Se puede observar que hay muchos coeficientes de regresión nulos en este caso. Las únicas variables de marketing que tienen coeficientes no nulos son las de la televisión y anuncios en VOL. Esto significa que son las únicas cuyas curvas respuesta no van a ser nulas. Se pueden observar las curvas estimadas en la Figura 4.4. Con la inversión aumenta de forma monótona el número de altas estimadas que obtiene cada canal. Además se percibe una saturación en este crecimiento, llegando a crecer de forma muy lenta a medida que las altas se acercan al valor β_m . Se puede interpretar este coeficiente como el valor máximo de altas que se obtienen en cada canal.

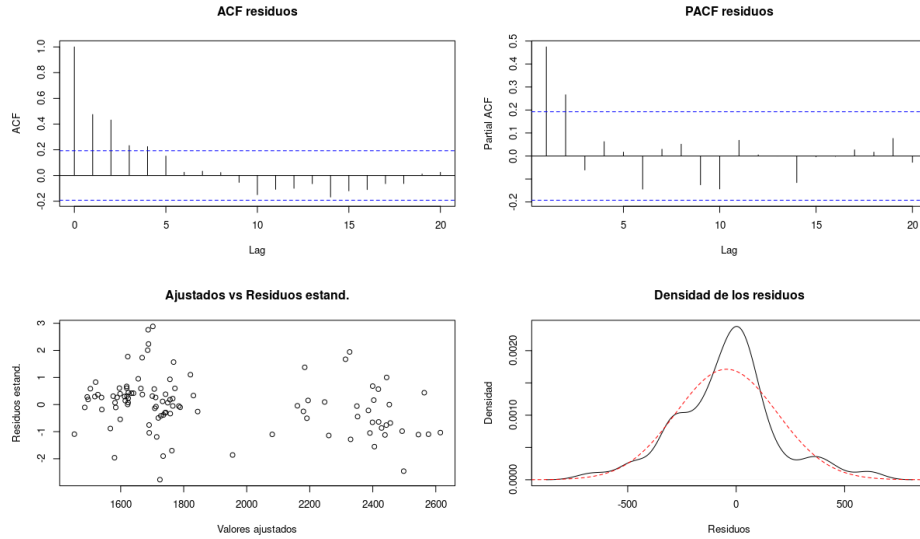


Figura 4.5: Gráficos del análisis de los residuos del modelo de regresión lineal con transformaciones.

De las variables de contexto, la asociada la población tiene un coeficiente negativo, pero los otros dos son positivos. Esto quiere decir que el aumento de la población disminuye el número de altas que obtiene el banco. Por otro lado, el coeficiente de regresión asociado a la competencia tiene un efecto positivo sobre el número de altas. Sería interesante estudiar el porqué y añadir restricciones sobre el coeficiente de regresión de esta variable.

Por último, se realiza un análisis de los residuos de este modelo siguiendo el mismo procedimiento que para los anteriores. En la Figura 4.5 se observa que hay valores de las ACF y las PACF que son significativos. Por otro lado, los residuos no parecen homocedásticos y la estimación de la densidad tipo kernel de estos no se asemeja demasiado a la normal máximo verosímil, pues se observa una ligera asimetría y colas más pesadas.

Se realizaron tests para contrastar las hipótesis de los residuos. El test de media nulo obtuvo un p -valor de 0.055 por lo que no hay evidencias estadísticas para rechazar la hipótesis de media nula a un nivel de significación del 5 %, pero sí para el 10 %. Posteriormente se realizaron los tests de Shapiro-Wilks y los de Ljung-Box para $r = 1, \dots, 10$ y todos ellos rechazaron la hipótesis nula. Por tanto, no hay evidencias de que los residuos del modelo sean ruido blanco.

4.5. Modelos de regresión penalizada con transformaciones

Para ajustar este modelo, la idea es similar a la del modelo lineal con transformaciones. La diferencia está a la hora de estimar los parámetros, donde se ha añadido una penalización a la función objetivo del problema de minimización.

Para la selección de los hiperparámetros del modelo se ha seguido el mismo procedimiento que para el anterior modelo. La diferencia se encuentra en la estimación de los parámetros. El estimador se puede observar en la Sección 2.5.5. Para ajustar este modelo se ha utilizado la función `cv.glmnet` de la librería `glmnet`, esta selecciona de forma automática el parámetro de penalización λ que cumple la regla del error estándar. Se ajustará un modelo de regresión penalizada con restricciones y transformaciones para cada una de las combinaciones obtenidas en la malla.

	$m=\text{BUSE}$	$m=\text{COMP}$	$m=\text{DISP}$	$m=\text{PROG}$	$m=\text{SN}$	$m=\text{VOL}$	$m=\text{TV}$
α_m	0.5	1.3	2.2	2	1.2	2.8	0.7
γ_m	0.3	0.7	0.5	0.6	0.5	1	1
θ_m	0.05	0.15	0.25	0.2	0.25	0.7	0.35

Tabla 4.2: Combinación de hiperparámetros que proporcionó un mejor ajuste para el modelo de regresión ridge con transformaciones.

Una vez ajustados, el mejor modelo obtuvo un $RMSE$ de 226.518 y un R^2 de 0.609. En la Tabla 4.2 se muestra la combinación de hiperparámetros utilizada para este modelo. Destacar que en este caso, el hiperparámetro θ asociado a la televisión es considerablemente mayor que los de la publicidad online. Esto provoca que la curva asociada a la televisión se sature mucho más tarde.

El ajuste del modelo proporciona las siguientes estimaciones para los coeficientes de regresión:

```
> coef[ind,]
Intercepto  BUSC          COMP          DISP          PROG          SN
2.274947e+04 0.000000e+00 0.000000e+00 0.000000e+00 0.000000e+00 0.000000e+00
VOL         TV          PIB          POB          COMP.TOTAL
3.926704e+02 6.499469e+02 1.115436e-03 -8.281744e-01 7.054650e-02
```

Se observa que las únicas variables de marketing con coeficiente no nulo vuelven a ser las de anuncios en VOL y en televisión. Por otro lado, en las variables de contexto los coeficientes asociados tienen valores muy próximos a los obtenidos para el modelo lineal con transformaciones.

Contenido eliminado por temas de confidencialidad

Para finalizar con esta sección, se analizarán los residuos de este modelo siguiendo el mismo esquema que el empleado en el resto de secciones del capítulo. En la Figura 4.6 se observan gráficos de la serie de residuos del modelo. Hay valores fuera de las bandas de significación de las ACF y las PACF. Por otro lado, no parece haber un problema de heterocedasticidad en los residuos y no se asemeja la estimación tipo núcleo de la densidad a la normal máximo verosímil de estos. Se realizaron distintos contrastes sobre estos para obtener evidencias estadísticas de estos supuestos. El contraste de media nula no fue rechazado, sin embargo se rechazaron tanto el test de Shapiro-Wilks, como los contrastes de Ljung-Box para $r = 1, \dots, 10$. Por tanto, existen evidencias estadísticas de que los residuos de este modelo no son ruido blanco.

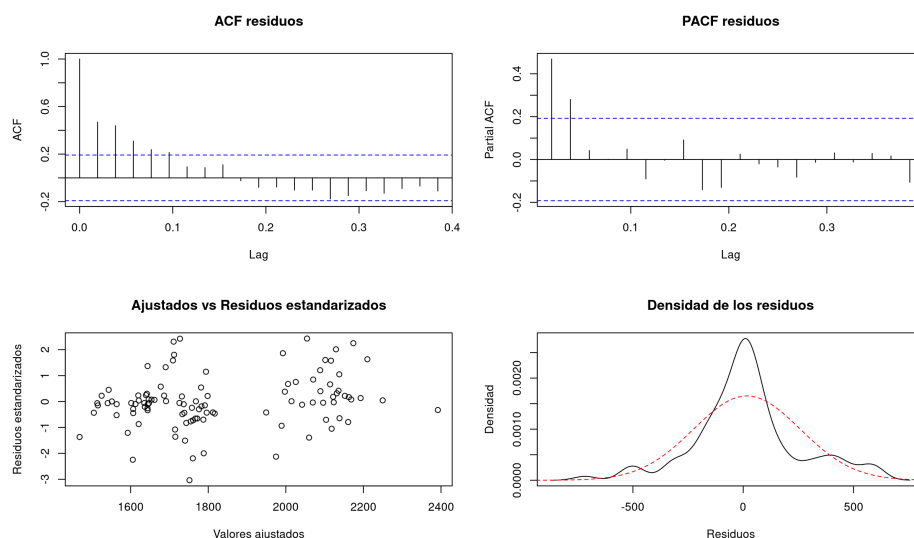


Figura 4.6: Gráficos de los residuos del modelo ridge con transformaciones en las variables de marketing.

4.6. Modelo Robyn

Estos modelos son complicados de ajustar debido a la gran cantidad de variables que se utilizan este problema y la correlación que presentan entre ellas. Además, los modelos ajustados hasta ahora no han cumplido la hipótesis de ruido blanco de los residuos. Robyn es el MMM más utilizado por las empresas para hacer el reparto de su presupuesto en marketing. En ABANCA se ha optado por estudiar

estos modelos y aplicarlo. Para esta sección se han utilizado las funciones de la librería **Robyn**. Este es un modelo de regresión penalizada con restricciones pero se encuentran algunas diferencias con 2.5.5.

Contenido eliminado por temas de confidencialidad

La gran diferencia de este modelo con el de la sección anterior está en la selección de los hiperparámetros del modelo. Para esta selección se lleva a cabo un problema de optimización multiobjetivo. Los criterios que se tienen en cuenta para esta se presentaron en la Sección 2.5.6.

Recordar que la inicialización del algoritmo es aleatoria, por lo que se han tomado distintos inicios y para cada inicio se realizan el mismo número de iteraciones. En Zhou et al (2023) se recomienda utilizar cinco trials y 2000 iteraciones. Sin embargo, tras un estudio, se determina que, para este problema los criterios de error que tiene en cuenta el algoritmo de optimización no mejoran de forma considerable a partir de las 500 iteraciones. Por tanto se optó por tomar 10 trials y 500 iteraciones en cada uno de ellos. De esta forma se ha mejorado considerablemente la velocidad de ajuste de estos modelos.

Multi-objective Evolutionary Performance

2D Pareto fronts with TwoPointsDE, for 10 trials with 500 iterations each

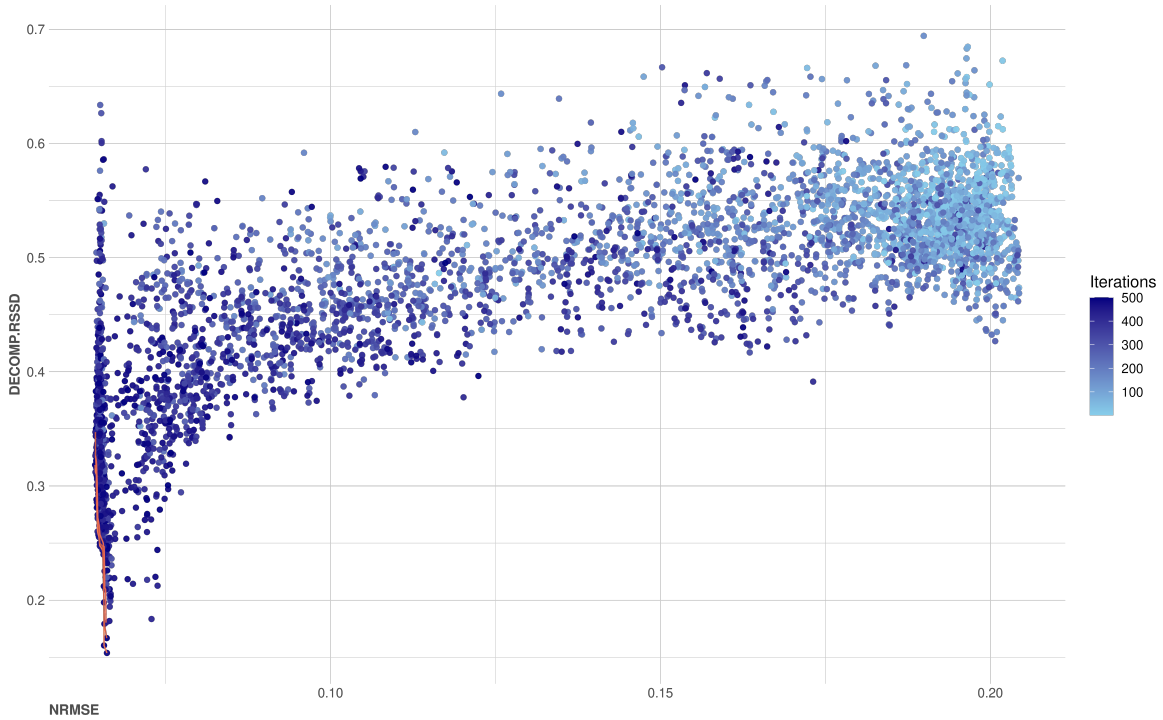


Figura 4.7: Evolución de los criterios de error a lo largo de las iteraciones del modelo.

En la Figura 4.7 se observan los valores de los criterios de error para los modelos ajustados. Los puntos de un color azul claro representan los modelos de las primeras iteraciones de los trials. Los de un color más oscuro son los modelos finales de cada uno de los trials. En la esquina inferior izquierda se pueden observar tres curvas de color rojo que unen varios puntos. Una de estas es la frontera de Pareto, definida en el Apéndice B. Esta contiene a los modelos con mejores valores para los criterios de error. La segunda curva contiene a los modelos que formarían la frontera de Pareto si no se tienen

en cuenta los elementos de la primera y se denomina segunda frontera de Pareto. La tercera curva contiene los modelos que formarían la frontera de Pareto si no se tuvieran en cuenta la primera y la segunda. Se podrían seguir definiendo fronteras de Pareto excluyendo las anteriores.

Las tres fronteras de Pareto contienen en total 59 modelos. Se ha disminuido el número de modelos entre los que elegir de 5000 a 59. De todas formas este sigue siendo un número elevado. Para disminuirlo han obtenido clústers según los valores del ROI de las variables de media. La técnica utilizada es la de k-means. El problema es que no se tiene conocimiento de cuantos grupos se deben obtener. Para ello, se ha hecho la agrupación con $k = 1$ hasta $k = 30$ y se ha tomado el valor de k para el que WSS del clúster haya disminuido en un 5 %. En la Figura 4.8 se puede observar el valor del WSS según el número de grupos generados.

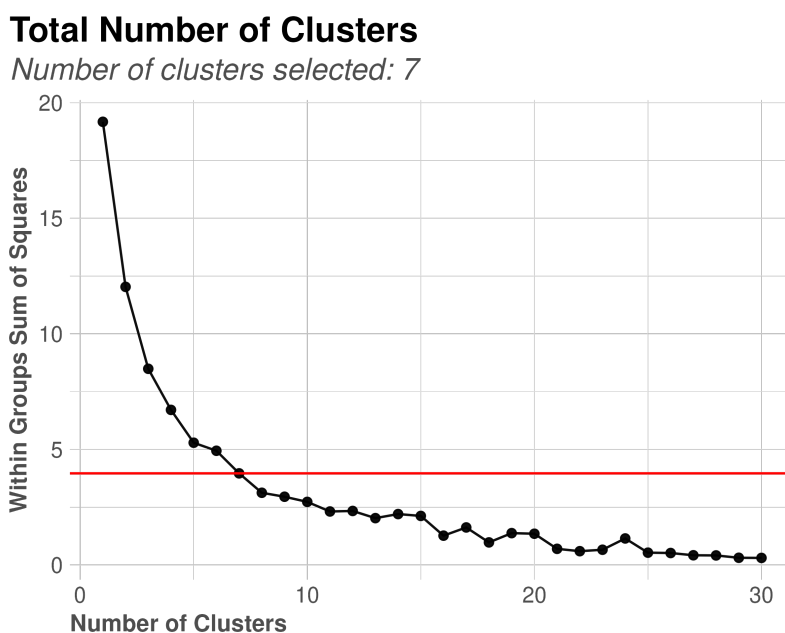


Figura 4.8: Evolución del WSS según el número de clusters generados.

Una vez obtenidos los clústers, los modelos candidatos son los modelos más próximos al centroide de cada grupo (utilizando la distancia euclídea entre los vectores de ROIs), de esta forma se reduce el número de modelos candidatos de 59 a 7. En la Figura 4.9, en el gráfico de la izquierda, se observa un diagrama de barras con los valores del ROI estimado para cada modelo candidato. En el de la derecha se puede observar el diagrama de dispersión de los modelos utilizando los criterios de error. Se observa que la diferencia entre los modelos candidatos es mínima en cuanto a NRMSE, siendo la más elevada de 0.0015. Es más notable si se tiene en cuenta el $DECOMP_RSSD$, donde la diferencia es de 0.2. Recordar que el $DECOMP_RSSD$ permite descartar las soluciones más irreales. Se observa que el que tiene n menor valor de este criterio de error es el modelo del cuarto clúster y se debe a que es el que tiene menos valores del vector de ROI nulos.

Una vez obtenidos, se debe elegir cuál es el modelo que mejor se ajusta a los datos reales. Para ello puede haber tres enfoques distintos:

- Conocimiento del negocio. Si la empresa ha realizado previamente estudios donde se hayan obtenido estimaciones del ROI, curvas de respuesta para cada variable, etc... En este caso, se

Figuras eliminadas por confidencialidad

Figura 4.9

seleccionaría el modelo que mejor se ajuste a estos estudios previos.

- Mediante parámetros estadísticos como el R^2 o el modelo con el menor $NRMSE$.
- Experimentando a futuro. En caso de disponer más datos a parte de los utilizados para ajustar el modelo, se tomaría el modelo que mejor ajuste esos datos. Un enfoque de entrenamiento/test similar al utilizado en el aprendizaje estadístico.

Tras estudiar los 7 candidatos, el elegido es el del cuarto clúster debido a que es el que más se aproxima a las estimaciones de ROI de ABANCA de otros modelos de atribución. Este tiene un valor del $NRMSE$ de 0.0661 y el $DECOMP.RSSD$ de 0.15.

Los coeficientes de regresión de este modelo son:

Contenido eliminado por temas de confidencialidad

Para finalizar esta sección se realiza un estudio de los residuos del modelo. Para ello se observa la Figura 4.10, donde se observan distintas gráficas de estos. En las ACF y las PACF se han obtenido valores nulos. En la gráfica de la esquina inferior izquierda no parece haber problemas con la homocedasticidad. Por último, se comparó la estimación tipo kernel de la densidad con la normal máximo verosímil, siendo estas bastante similares.

Se han realizado los mismos contrastes que se realizaron sobre los residuos del resto de modelos. Ninguno de ellos ha sido rechazado, además tras observar la serie de tiempo de estos residuos, estos parecen tener varianza constante, por lo que se puede determinar que los residuos fueron ruido blanco.

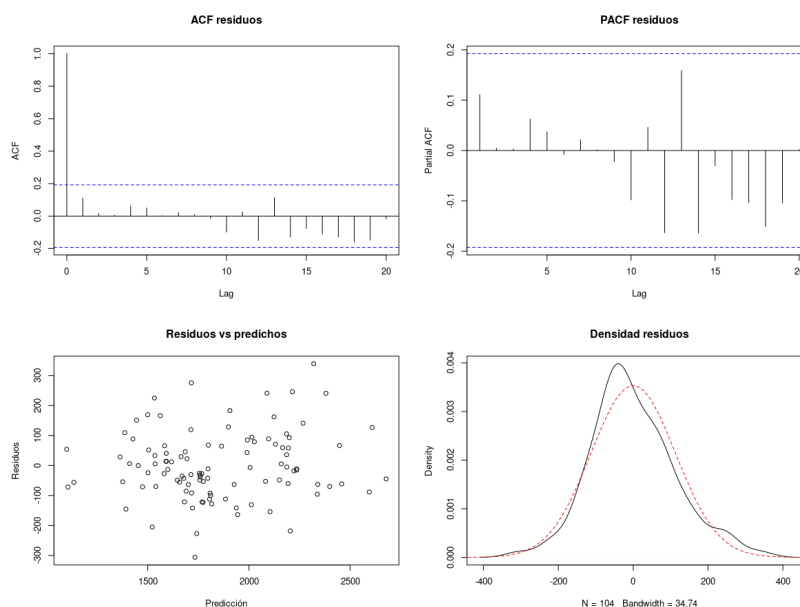


Figura 4.10: Análisis de los reiduos del modelo Robyn.

4.7. Comparación de los modelos

En esta sección se comparan brevemente los modelos ajustados a lo largo de este capítulo.

En la Tabla 4.3 se observan los valores del $RMSE$ de estos modelos y del R^2 , se puede concluir que el modelo que presenta el mejor ajuste es Robyn. Además, este fue el único modelo para el que se cumplieron las hipótesis de los residuos.

A continuación, en la Figura 4.11 se pueden observar la serie de valores reales de las altas junto con los ajustados por los modelos más competitivos. Se observa que en 2021 es más complicado ajustar los valores de la serie. Esto se debe a que la inversión en muchos tipos de anuncios introducidos en los modelos es nula hasta comienzos de 2022. Sin embargo, el modelo que mejor se ajusta a la serie de altas original en 2021 es la serie de Robyn, capaz de captar mejor la subida de septiembre de ambos

Modelo	RMSE	R^2
Lineal	218.42	0.61
Ridge	242.66	0.59
Semiparamétrico	197.74	0.7
Lineal con transformaciones	223.33	0.62
Ridge con transformaciones	226.1	0.61
Robyn	112.83	0.88

Tabla 4.3: Valores del RMSE y del R^2 para los modelos ajustados.

años. Esto ocurre porque, al añadir las series de tendencia, estacionalidad y efectos festivos el modelo pudo ajustar mejor estas subidas y bajadas según la época del año.

Figuras eliminadas por confidencialidad

Figura 4.11

Debido a que Robyn es el modelo que presenta mejor ajuste de todos los estudiados es el que se eligió para llevar a cabo la optimización del presupuesto. Para ello serán claves las curvas de respuesta

estimadas.

Capítulo 5

Optimización del presupuesto

En el año 2022 se ha En la Sección 1.3 se comenta brevemente que los MMMs tienen dos objetivos. El primero de ellos es ajustar un modelo de regresión para explicar el valor de un KPI utilizando las variables de marketing de la empresa, esto se realizó en el Capítulo 4.3, estudiando cuales eran los canales más efectivos a la hora de atraer atraer nuevos clientes. La segunda es optimizar el reparto del presupuesto en esas variables que será el problema al que se dedica este capítulo. Para esto serán claves las curvas respuesta definidas en (2.7) y estimadas por los modelos.

En Jin et al (2017) se propone el siguiente problema de optimización para obtener el reparto óptimo del presupuesto C .

$$\begin{aligned} & \max_{x_1, \dots, x_M} \sum_{m=1}^M \hat{Y}(x_1, \dots, x_M, 0, \dots, 0) \\ & \text{sujeto a } \sum_{m=1}^M x_m = C \\ & \quad x_m \geq 0 \quad \forall m \in \{1, \dots, M\} \end{aligned} \tag{5.1}$$

donde $\hat{Y}(x_1, \dots, x_M, 0, \dots, 0)$ es la predicción del modelo, cuando las variables de publicidad toman valores x_1, \dots, x_M y las de contexto toman valores nulos. Tomando la ecuación (2.6) para $\hat{Y}(x_1, \dots, x_M, 0, \dots, 0)$ este problema es equivalente a

$$\begin{aligned} & \max_{x_1, \dots, x_M} \beta_0 + \sum_{m=1}^M \beta_m \text{Hill}(x_m^{ad}, \alpha_m, \gamma_m) \\ & \text{sujeto a } \sum_{m=1}^M x_m = C \\ & \quad x_m \geq 0 \quad \forall m \in \{1, \dots, M\}. \end{aligned} \tag{5.2}$$

Debido a que β_0 es constante, optimizar este problema es igual a optimizar la suma de las curvas respuesta (2.7). Por otro lado en Zhou et al. proponen un problema similar a (5.1) pero añadiendo cotas a las variables de marketing. De esta manera, se maximiza el número de altas obtenidas, dentro de un reparto lógico del presupuesto.

A continuación, se denota por \bar{X}_m^k a la media de los valores no nulos de la variable X_m en las últimas k semanas. De la misma manera, para cada canal de marketing se definen las constantes $l_m \in [0, 1]$ y

$u_m \in [1, +\infty)$ que ayudan a fijar los límites del cambio en la variable m . Así para cada una de ellas el valor óptimo se debe encontrar dentro del intervalo $[l_m \bar{X}_m^k, u_m \bar{X}_m^k]$. Esto provoca que los resultados obtenidos no den lugar a cambios drásticos en las inversiones de publicidad y el resultado tenga sentido para los agentes de marketing de la empresa.

Por ejemplo, si en las últimas 8 semanas se ha invertido una cantidad de $\bar{X}_{m_k} = 15000$ euros en publicidad en buscadores y el departamento de marketing está dispuesto a subir esa cantidad en un 33 % o bajarla en un 50 % entonces $u_m = 1.33$ y $l_m = 0.5$. De esta forma, el valor óptimo se encuentra entre 7500 y 19950 euros. De esta forma el problema se reformula de la siguiente manera:

$$\begin{aligned} & \underset{x_1, \dots, x_M}{\text{máx}} \sum_{m=1}^M \beta_m \text{Hill}(x_m^{ad}, \alpha_m, \gamma_m) \\ & \text{sujeto a } \sum_{m=1}^M x_m = C \\ & \quad l_m^* \leq x_m \leq u_m^* \quad \forall m \in \{1, \dots, M\}, \end{aligned} \tag{5.3}$$

donde $l_m^* = l_m \bar{X}_{m_k}$ y $u_m^* = u_m \bar{X}_{m_k}$ y C es el presupuesto para la semana $T + 1$.

La suma de curvas de respuesta no será lineal, sin embargo las restricciones de este problema si lo son. Existen muchos algoritmos para solucionar problemas de optimización similares a (5.3). Sin embargo, en Zhou et al. (2023) se proponen dos para solucionar este problema:

- Algoritmo de lagrangiano aumentado propuesto por Conn et al (1991).
- Algoritmo de Sequential Least Squares Quadratic Programming (SLSQP) propuesto por Kraft (1988).

De estos dos algoritmos se ha optado por utilizar el del lagrangiano aumentado.

5.1. Algoritmo del lagrangiano aumentado

La idea del algoritmo del lagrangiano aumentado es combinar las virtudes de las dos versiones de los algoritmos de penalización exterior (absoluta y cuadrática).

El primer paso que se debe tomar antes de presentar el algoritmo es definir el concepto de lagrangiano aumentado de un problema de optimización.

Definición 5.1.1. El lagrangiano aumentado de un problema de optimización P con J restricciones de desigualdad y Q de igualdad

$$\begin{aligned} & \text{PROBLEMA } P \\ & \underset{x}{\text{minimizar}} \quad f(x) \\ & \text{sujeto a} \quad g_j(x) \geq 0 \quad \forall j \in \{1, \dots, J\} \\ & \quad h_q(x) = 0 \quad \forall q \in \{1, \dots, Q\} \end{aligned} \tag{5.4}$$

se define su lagrangiano aumentado $L_P^A(x, u, v)$ como

$$L_P^A(x, u, v) = \frac{\rho}{2} \left(\sum_{j=1}^J \max\{0, g_j(x) + \frac{u_j}{\rho}\}^2 + \sum_{q=1}^Q (h_q(x) + \frac{v_q}{\rho})^2 \right) + c \tag{5.5}$$

donde u y v son los multiplicadores de Lagrange de las restricciones g y h respectivamente.

El algoritmo del lagrangiano aumentado es iterativo, en cada iteración t se dispone de un parámetro de penalización ρ^t y de unas aproximaciones de los multiplicadores de lagrange u^t y v^t . Para esos valores se buscará el mínimo de $L_\rho^A(x, u^t, v^t)$. Suponiendo que x^{t+1} es el mínimo de este problema, entonces se actualizan los valores de u^t a $u_j^{t+1} = \max\{0, u_j^t + \rho^t g_j(x^{t+1})\}$ y v_q^t a $v_q^{t+1} v_j^t + \rho^t h_j(x^{t+1})$. El valor de ρ no se actualizará en todas las iteraciones, provocando que no se obtengan penalizaciones tan elevadas como en otros métodos de optimización. Por último, el algoritmo se para cuando se cumple un criterio de parada, para ello se define un valor ϵ para el que se debe cumplir que para los distintos criterios de error calculados $C < \epsilon$.

En Conn et al (1991) se estudia la convergencia tanto local como global de este algoritmo. Además, se demuestra la convergencia global y se establece que un parámetro de penalización demasiado pequeño puede ser potencialmente problemático.

5.2. Resultados

Para la primera versión de este problema, se plantea que el presupuesto tome un valor $C = \sum_{m=1}^M \bar{X}_{m_k}$. A la hora de añadir las cotas, no tiene sentido fijar las mismas para todas las variables de publicidad. Se ha comprobado a lo largo del trabajo que la variable de televisión afecta de una forma muy positiva al número de altas. Además, la inversión en esta es mucho más elevada que en el resto de variables de marketing. Ser agresivo con las cotas para esta variable implica permitir que cambie mucho el reparto del presupuesto, por lo que se han fijado valores de l_m y u_m más próximos a 1 que en el resto de canales de marketing.

- Debido a que los anuncios de televisión son tan caros, no es posible aumentar estos de forma considerable, por lo tanto se ha optado por fijar $u_m = 1.5$, de forma que sólo se permite que aumente en un 50 % el presupuesto. Por otro lado, con los resultados tan positivos que se han obtenido, se ha fijado $l_m = 0.7$ de forma que sólo se permite que el presupuesto baje en un 30 %.
- Los anuncios de publicidad online son más baratos que los de televisión. Esto implica que aumentar la inversión supone un esfuerzo menor que en el caso de los televisivos. Por lo tanto se ha propuesto $u_m = 2$, de forma que los anuncios online se pueda aumentar la inversión en un 100 %. Por otro lado, no da tanto miedo bajar de forma considerable los anuncios web por lo que se ha fijado $l_m = 0.5$ de forma que se permite reducir la inversión en un 50 %.

Por otro lado, para este problema inicial se ha tomado $k = 104$, es decir los datos de los dos años anteriores. Además, se ha supuesto que $C = \sum_{m=1}^M \bar{X}_{m_k}$.

Por último, comentar que aquellas variables con β asociado igual a 0 no se han tenido en cuenta para obtener la optimización del presupuesto. De esta forma en vez de ser $M = 7$, se tiene $M = 5$ debido a que los coeficientes asociados a las variables inversión en Vídeos-On-Line y en programática tienen valores nulos.

Recordar además que se habían definido los valores de l_m y u_m de forma que se los límites inferiores y superiores para las variables de inversión tienen las siguiente cotas:

Contenido eliminado por temas de confidencialidad

Una vez definidas las constantes presentes en el problema (5.3), se puede resolver el problema de optimización. Para ello se utiliza la función `robyn_allocator` de la librería `Robyn`.

Para solucionar el problema, utiliza la función `nloptr` de la librería `nlopt`. Esta contiene distintos algoritmos para resolver problemas de optimización no lineal, entre estos se encuentran el algoritmo del lagrangiano aumentado propuesto por Conn et al (1991) y el algoritmo de SLSQP propuesto por Kraft (1988). Si se quiere conocer más sobre las funciones de esta librería y los algoritmos que contiene consultar Johnson (2023) .

Contenido eliminado por temas de confidencialidad

Con el gasto promedio, se estima que se atraen a 63 clientes nuevos cada semana, siendo el medio publicitario que más aporta a esto la televisión. Por otro lado, con el reparto óptimo del presupuesto se han obtenido 79 altas por semana. Se aumenta el número de clientes en 16, es decir, el las altas producidos por las actividades de marketing de ABANCA cada semana crece en un 25.4%.

Para tener una imagen más clara del reparto óptimo, se muestra en la Figura ?? las curvas de respuesta junto con los puntos que representan el reparto medio del presupuesto y los del óptimo. Destacar que si se hubiera dado más libertad a las variables de inversión, se hubiera obtenido un estimación mucho mayor de la inversión óptima en los anuncios en redes sociales y en tipo display.

Contenido eliminado por temas de confidencialidad

Por último, comentar que en el caso de que $C \neq \sum_{m=1}^M \bar{X}_{m_k}$ se debería tomar otro punto de inicio debido a que \bar{X}_k ya no sería un punto factible del problema de optimización. Para tomar un punto inicial, se propone tomar el vector del reparto del presupuesto y multiplicarlo por C . Es decir, se puede observar que en la Figura 5.1 el porcentaje de gasto inicial en cada uno de los canales. Multiplicando estos valores por C se puede obtener un punto factible del problema que permite comparar el hipotético gasto realizado con el óptimo.

La solución de este problema permite a los agentes de marketing comparar el reparto del presupuesto realizado con el óptimo. De esta forma se puede valorar la inversión realizada y estudiar como mejorarla.

Figuras eliminadas por confidencialidad

Figura 5.1

Capítulo 6

Conclusiones y líneas futuras

En este capítulo final, se resumen brevemente las tareas realizadas a lo largo del trabajo. Además, se estudian las distintas líneas futuras que pueden llevarse a cabo en este proyecto:

6.1. Conclusiones

La gran inversión realizada a lo largo de 2022 en publicidad, generó la necesidad de modelos estadísticos que aportasen ideas de como mejorar el reparto del presupuesto. *ABANCA* se interesó por los MMMs y el objetivo de este proyecto es conocer estos modelos y estudiar su potencial dentro de la entidad.

Estos modelos permiten a las empresas valorar las campañas publicitarias realizadas en el periodo de estudio, viendo cuales son los canales publicitarios que obtienen mejores resultados. Se han utilizado los datos del número de altas obtenidas en 2021 y 2022 agrupados de forma semanal, además de los datos de la inversión en distintos medios publicitarios, variables macroeconómicas y una variable de actividad de la competencia en ese mismo periodo de estudio.

En el desarrollo de los MMMs se han utilizado distintos modelos de regresión clásico. Se comienza utilizando un modelo de regresión múltiple en el que se realiza una selección de variables debido a la no significación de muchos coeficientes de regresión. Esto se debe a la elevada correlación que presentan las variables introducidas en el modelo. Por esta misma razón, se utilizan modelos de regresión penalizada. Estos ajustes no presentaron buenos resultados pero dieron ideas de cuáles tendrían más importancia a la hora de estimar el número de altas.

Una vez obtenido el ajuste de estos modelos paramétricos, se opta por utilizar los modelos parcialmente lineales para afrontar este problema. De nuevo, por la gran correlación de las variables estos no ofrecieron un buen ajuste, obteniendo muchos de los coeficientes de regresión nulos. Tras una selección de variables se obtuvo que los anuncios en buscadores y en televisión presentaban un efecto positivo sobre el número de altas. Las hipótesis de los MMMs no se cumplen para estos modelos, por lo que no pueden ser utilizados para obtener conclusiones.

Debido a que los modelos conocidos por el autor de este trabajo de fin de máster no presentaron buenos resultados en el ajuste de los datos, se optó por utilizar las transformaciones más comunes en la literatura de los MMMs sobre las variables de publicidad de *ABANCA*. Estas transformaciones dependían de hiperparámetros. Se exploraron distintas combinaciones de hiperparámetros para trabajar con estos datos. Primero se creó una malla de posibles valores de estos y se realizaron ajustes de regresión restringida utilizando las técnicas de mínimos cuadrados y mínimos cuadrados penalizados.

Esta búsqueda aleatoria no proporcionó un buen ajuste, obteniendo valores del R^2 muy pequeños y residuos que presentaban dependencia temporal.

Posteriormente, se optó por estudiar el ajuste que proporciona Robyn, el MMM creado por facebook que utilizaba estas mismas transformaciones. Este incorpora un algoritmo de optimización que hace una búsqueda de los posibles valores de los hiperparámetros. Este proceso consiste en un algoritmo iterativo que ayuda a buscar los valores de estos. Para cada combinación de hiperparámetros obtenida por el algoritmo de optimización se ajusta un modelo de regresión ridge. Se ajustan muchos modelos, de los cuales, una gran cantidad de ellos presentan un buen ajuste, pero puede ocurrir que los hiperparámetros escogidos sean muy distintos. Por lo tanto, se debe de escoger el modelo que presente un buen ajuste y a la par se aproxime a las estimaciones del ROI obtenidas en otros estudios de ABANCA.

Una vez seleccionado uno de los modelos de Robyn, se estudió como hacer el reparto del presupuesto. Se propone un problema de optimización no lineal con restricciones para esta tarea. Entre los algoritmos utilizado para resolverlo, se escogió el del lagrangiano aumentado.

Contenido eliminado por temas de confidencialidad

Por último, destacar que se estudiaron más técnicas de las mostradas. Debido al numero tan elevado de variables para los datos disponibles se intentó utilizar análisis de componentes principales dinámicas previo al ajuste del modelo. Sin embargo, al utilizar únicamente las variables de marketing para la optimización, se descartó esta opción. Además, se estudiaron las relaciones entre las variables de inversión con los clicks e impresiones. Para ello se utilizaron modelos de regresión lineal simple y no lineales como el de Michaelis-Menten.

6.2. Líneas de trabajo futuras

Estos modelos son innovadores para ABANCA y abren un gran abanico de posibilidades para la entidad:

▪ Línea eliminada por temas de confidencialidad

- **Ajustar un modelo utilizando como variables de contexto solo las de tendencia, estacionalidad y efecto vacacional.** De esta forma el número de variables del problema original disminuye y seguimos teniendo el contexto explicado por las variables obtenidas por Prophet. Los problemas de correlación entre las variables no permitieron desagregar más las de marketing. En el caso de los anuncios en televisión, es clave conocer si provienen más clientes de canales como Antena 3 o Telecinco. Esto se debe a que se cree que los clientes que ven estos canales televisivos tienen perfiles distintos.
- **MMM Lightweight (Google).** Este es otro MMM popular entre los que utilizan las empresas. La formula es muy similar a la que utiliza Robyn. La gran diferencia se encuentra en la estimación de los parámetros e hiperparámetros del modelo.
- **Creación de una herramienta de simulación para los agentes de marketing.** Sería de gran utilidad para el equipo de Marketing de ABANCA crear una herramienta de simulación en la que, según el presupuesto indicado, utilizando las curvas de respuesta, se obtuviera una estimación del número de altas obtenidas. Muchas veces no se puede invertir una cantidad deseada debido a

que las compañías que llevan las campañas de marketing cobran cantidades establecidas. Por lo que estos modelos permitirán tener una idea de cuantas altas pueden proporcionar las cantidades establecidas por marketing.

- **Ajustar MMMs para los distintos niveles de captación.** En la Sección 3.1 se introdujeron los niveles por los que pasa un alta antes de convertirse en nuevo cliente. Sería interesante ver las estimaciones de las curvas de respuestas para distintos niveles de captación. De esta manera se podría comprobar si hay anuncios que atraen más leads que altas o viceversa. Se podría investigar de esta forma si hay medios publicitarios que no aporten tanto a las altas y sin embargo, generan un gran número de leads.

Apéndice A

Prophet

Prophet es una metodología de predicción de series de tiempo. Para ello, se ajusta un modelo de regresión donde se introducen efectos de tendencia, estacionales y de festivos o eventos. Es decir, se intenta modelar la serie de tiempo sobre la que se quiere hacer predicciones de la siguiente manera. Este apéndice está basado en Taylor y Letham (2021).

$$y_t = g_t + s_t + h_t + \varepsilon_t$$

donde g_t es una serie de tiempo que modela de tendencia, s_t representa una serie de tiempo en la que se reflejan los cambios periódicos de la serie de tiempo y h_t representa la los efectos de las vacaciones que ocurren puntualmente.

A continuación, se va a presentar brevemente como se obtienen estas series de tiempo. La primera serie que se presenta es la de tendencia, esta se puede estimar utilizando un enfoque no lineal con crecimiento saturado o un enfoque lineal por partes. La fórmula del modelo no lineal es similar a la del crecimiento de la población en ecosistemas. Este tipo de crecimiento se suele modelar con la función:

$$g(t) = \frac{C}{1 + \exp(-k(t - m))}$$

donde C es el punto de saturación, m un parámetro que permite ralentizarla y k el ratio de crecimiento. Puede ocurrir que el parámetro C o k no sean constantes en el tiempo por distintos eventos ocurridos en el periodo de estudio. Prophet tiene implementado cambios de tendencia dentro del modelo. Estos se producen en lo que se denominan puntos de cambio, son los momentos en los que se modifican estos parámetros.

Dados S puntos de cambio en los momentos s_j con $j = 1, \dots, S$. Se define el vector $\delta \in \mathbb{R}^S$ donde δ_j es el cambio que se produce en el ratio en el momento s_j . Por lo tanto el ratio de cambio en el momento s_j es:

$$k + \sum_{j: t > s_j} \delta_j$$

También se define un vector $a(t) \in \{0, 1\}$ de forma que sea 1 cuando $t \geq s_j$ y nulo en otro caso. De esta forma, en el momento t el ratio de cambio será $k + a(t)^T \delta$. El parámetro m también debe de ser cambiado para conectar correctamente los segmentos. Así que se define γ_j como

$$\gamma_j = \left(s_j - m - \sum_{l < j} \gamma_l \right) \left(1 - \frac{k + \sum_{l < j} \delta_l}{k + \sum_{l \leq j} \delta_l} \right)$$

Se puede reformular $g(t)$ como sigue

$$g(t) = \frac{C(t)}{1 + \exp(-(k + a(t)^T \delta)(t - (m + a(t)^T \gamma)))}$$

Otra forma de estimar la serie de tendencia es utilizar un modelo lineal por partes que sigue la siguiente fórmula:

$$g(t) = (k + a(t)^T \delta)t + (m + a(t)^T \gamma)$$

donde k sigue siendo el parámetro de crecimiento mientras que m sirve para reajustar el modelo y que siga siendo lineal, por otro lado en este caso el parámetro $\gamma_j = -s_j \delta_j$ para que la función sea continua.

Los puntos de cambios s_j se podría elegir por conocimiento debido a alguna crisis o algún cambio en nuestra política de inversiones. De todas formas Prophet tiene un método de selección automática dado un conjunto de candidatos.

Por otro lado para estimar el efecto estacional, se utilizan las series de Fourier. Se denota a P como el periodo que tiene la serie de tiempo de estudio. En López Pouso (2019) se demuestra como aproximar una función arbitraria con efectos suaves utizando series de Fourier:

$$s(t) = \frac{a_0}{2} + \sum_{n=1}^N a_n \cos\left(\frac{2\pi nt}{P}\right) + b_n \sin\left(\frac{2\pi nt}{P}\right). \quad (\text{A.1})$$

Estas se utilizan para estimar la serie de tendencia. Sin embargo, en Taylor y Letham (2017) [21] se omite el sumando $\frac{a_0}{2}$ de la ecuación (A.1) debido a que ya se ha estimado una serie para la tendencia. Ajustar un modelo para la estacionalidad supone tener que estimar $2N$ parámetros $\beta = [a_1, b_1, \dots, a_N, b_N]$. Para ello se ha construido el siguiente vector

$$X(t) = \left[\cos\left(\frac{2\pi(1)t}{P}\right), \sin\left(\frac{2\pi(1)t}{P}\right), \dots, \cos\left(\frac{2\pi(N)t}{P}\right), \sin\left(\frac{2\pi(N)t}{P}\right) \right]$$

de esta forma se puede escribir la componente estacional como:

$$s(t) = X(t)\beta$$

Por último, en este modelo se añade el efecto que pueden tener las vacaciones o distintos eventos dentro de la empresa como el lanzamiento de un producto o alguna promoción dentro de la empresa. Estos eventos no entran dentro de la componente estacional debido a que no siguen un patrón periódico. Por ejemplo, el periodo de Semana Santa no es algo que ocurra periódicamente debido a que cada año este evento es en una semana distinta, por ejemplo en 2021 el Jueves Santo fue el día 2 de abril mientras que en el año 2022 fue el día 15 de abril.

Para implementar este efecto se debe que indicar cuando ocurren estos eventos. La librería de **prophet** dispone de un conjunto de datos donde están almacenados los festivos de 59 países desde el 1 de enero de 1995 hasta finales del año 2044. Dentro de las vacaciones asociadas España se pueden encontrar alguna como el día de La Inmaculada Concepción (vacación periódica) o el Viernes Santo (no periódico).

Para cada festivo i se le asigna un indicador D_i de forma que sea 1 si en la semana t ocurre el evento i y 0 en otro caso. Además a cada vacación un parámetro k_i que se corresponde con el efecto de este evento. El procedimiento seguido es similar al tomado para la estacionalidad. Suponinedo que hay L festivos por año de forma que:

$$Z(t) = [\mathbb{I}(t \in D_1), \dots, \mathbb{I}(t \in D_L)]$$

de esta forma se puede definir $h(t)$ cómo:

$$h(t) = Z(t)k$$

La estimación de los parámetros asociados a las series de tendencia, estacionalidad y festividades se hacen mediante inferencia bayesiana. Para más información sobre las distribuciones a priori de estos y el algoritmo utilizado para su estimación consultar Taylor y Letham (2021).

Contenido eliminado por temas de confidencialidad

Para acabar, se puede observar la serie de los eventos. Destacar que la mayoría de ellos son negativos. Los únicos festivos con efectos positivos son la semana del día del trabajador y la semana de Navidad. Añadir que se aprecian festivos no periódicos en esta serie, en 2021 la semana santa fue en la última semana de Marzo y en 2022 fue la segunda de abril, por ello en 2021 hay más valores entre ese evento y el del día del trabajador.

Apéndice B

Algoritmos de evolución diferencial

Una de las diferencias entre el modelo Robyn y el de regresión penalizada con transformaciones en las variables de media está en la selección de los hiperparámetros de estas transformaciones y en la elección del parámetro λ de la penalización.

El algoritmo utilizado para esta tarea pertenece a la familia de algoritmos de evolución diferencial. Esta familia de algoritmos se inspiran en la teoría de la evolución de Darwin. Se ha utilizado para resolver distintos problemas en los campos de la optimización y la ingeniería. Estos son introducidos en Storn y Price (1997).

Estos pertenecen a los algoritmos de búsqueda metaheurística (MSA), que obtienen soluciones candidatas a resolver el problema en base a una forma de búsqueda inspirada en distintos fenómenos de la naturaleza. Dentro de los MSA se distinguen cuatro categorías (ver Ahmad et al. (2022)):

- Algoritmos evolutivos, basados en la teoría de la evolución. Un ejemplo de este tipo de algoritmos son la familia de la evolución diferencial. También se conocen como algoritmos genéticos.
- Algoritmos de inteligencia de colmena, como el algoritmo de optimización de enjambre de partículas (PSO).
- Algoritmos basados en comportamientos humanos, como el algoritmo de comportamiento cognitivo.
- Algoritmos basados en leyes físicas, como el algoritmo de búsqueda gravitacional.

Los algoritmos evolutivos se basan en su mayoría en la teoría de la evolución de Darwin y el concepto de supervivencia del más apto en el que se basa la selección natural. Muchos de los conceptos de esta teoría como la mutación y la selección son utilizados en estos algoritmos.

Los algoritmos de evolución diferencial son los más populares entre los algoritmos evolutivos para tratar con problemas de optimización complejos. Este genera nuevas soluciones en base a recombinar las obtenidas previamente y estas serán reemplazadas si y sólo si la nueva solución mejora el valor en la función objetivo.

Algunas de las ventajas que tiene el algoritmo de evolución diferencial sobre otros tipos de algoritmos de búsqueda metaheurística es que:

- La implementación de este algoritmo es más simple y directa que la de otros MSA
- A pesar de su simpleza, el algoritmo de evolución diferencial ha demostrado ser uno de los MSA con mejor resultados en distintos problemas de optimización.

Hay muchas versiones distintas de estos algoritmos, las más comunes de ellas vienen recogidas y brevemente explicadas en Ahmad et al. (2022) que es la referencia principal de este apéndice. Estos algoritmos tienen cuatro fases denominadas: inicialización, mutación, recombinación y selección. La inicialización es un proceso que se realiza una única vez, mientras que los otros procesos se repiten de forma recurrente. Para introducir estos algoritmos se va a suponer se quiere minimizar la función $f(x_1, \dots, x_D)$ donde tenemos D variables decisoras x_1, \dots, x_D . Donde los límites de estas variables vienen dados por $X_j \in [a_j, b_j]$ con $j = 1, \dots, D$.

B.1. Fase de inicialización

Es el primer paso de los algoritmos de evolución diferencial. Antes de comenzar a hablar de cómo se inicializa el algoritmo, antes de comenzar se debe aclarar que al igual que la teoría de Darwin, este algoritmo trabaja con poblaciones o generaciones.

Cada generación está compuesta por NP soluciones, es decir tendremos NP vectores de dimensión D dentro de cada población. Durante la t –ésima población, cada i –ésima solución se puede representar cómo:

$$X_i^t = (X_{i,1}, X_{i,2}, \dots, X_{i,D}) \quad (\text{B.1})$$

donde $i = 1, 2, \dots, NP$. Para obtener la primera población lo primero que se debe hacer es definir los límites de las variables en los siguientes vectores:

$$X_{min} = (a_1, a_2, \dots, a_D) \quad (\text{B.2})$$

$$X_{max} = (b_1, b_2, \dots, b_D) \quad (\text{B.3})$$

La generación inicial se puede obtiene mediante un proceso aleatorio. Dado un $i \in \{1, 2, \dots, NP\}$ para la i –ésima solución de algoritmo de diferenciación evolutiva se define la j –ésima componente cómo un valor generado de forma aleatoria dentro del intervalo $[a_j, b_j]$. Es decir, dado un $u_{i,j} \in U[0, 1]$ se define:

$$X_{i,j}^{(0)} = X_{min,j} + u_j(X_{max,j} - X_{min,j}) = a_j + u_{i,j}(b_j - a_j) \quad (\text{B.4})$$

donde $X_{i,j}^{(0)}$ es la componente j –ésima de la solución i –ésima de la generación inicial. Una vez obtenidas las NP soluciones primera generación se finaliza el proceso de inicialización y se avanza a la fase de mutación.

B.2. Mutación

En términos biológicos se define como un cambio instantáneo de característica observado en un cromosoma. En el contexto de la evolución diferencial es una perturbación aleatoria sobre las variables de decisión elegidas. A continuación, se define el vector Y_i^t construido a partir de los elementos de la generación t utilizando las estrategias de mutación más comunes dentro de los algoritmos de evolución diferencial, entre ellas destacan las siguientes:

- DE/rand/1:

$$Y_i^t = X_{r_1}^t + F(X_{r_2}^t - X_{r_3}^t) \quad (\text{B.5})$$

- DE/rand/2:

$$Y_i^t = X_{r_1}^t + F(X_{r_2}^t - X_{r_3}^t) + F(X_{r_4}^t - X_{r_5}^t) \quad (\text{B.6})$$

- DE/best/1:

$$Y_i^t = X_{best}^t + F(X_{r_1}^t - X_{r_2}^t) \quad (B.7)$$

- DE/best/2:

$$Y_i^t = X_{best}^t + F(X_{r_1}^t - X_{r_2}^t) + F(X_{r_3}^t - X_{r_4}^t) \quad (B.8)$$

- DE/current-to-best/1:

$$Y_i^t = X_i^t + F(X_{best}^t - X_i^t) + F(X_{r_1}^t - X_{r_2}^t) \quad (B.9)$$

En las anteriores ecuaciones (B.5),(B.6),(B.7),(B.8),(B.9) los elementos X_{r_i} con $i \in \{1, 2, 3, 4, 5\}$ son los r_i -ésimos elementos de la generación t tomados de forma aleatoria donde $r_i \in [1, \dots, NP]$ para $i \in \{1, 2, 3, 4, 5\}$ son distintos entre sí y a su vez distintos de la solución X_i , por otro lado X_{best}^t es la mejor solución de la generación t . F es un vector que sirve para controlar el proceso de mutación que está entre $[0, 1]$. Elegir un valor apropiado de F es crucial debido a que nos dá un equilibrio entre la exploración y explotación de nuevas soluciones para evitar inconvenientes cómo puede ser convergencia prematura o lenta.

B.3. Recombinación

El objetivo de esta fase está en combinar los elementos de la generación t con los vectores mutantes definidos en la anterior sección. Esto le permite a la siguiente generación a heredar atributos del vector mutante. Hay dos maneras distintas de llevar a cabo esta recombinación:

- Uniforme, se define el vector Z_i^t cómo el vector recombinado. El elemento j -ésimo de este vector viene dado de la siguiente forma:

$$Z_{i,j}^t = \begin{cases} X_{i,j}^t & \text{si } v_{i,j} > CR \quad y \quad j \neq k_i \\ Y_{i,j}^t & \text{si } v_{i,j} \leq CR \quad o \quad j = k_i \end{cases} \quad (B.10)$$

donde $v_{i,j}$ pertenece a una distribución uniforme en el intervalo $[0,1]$ y k_i es un elemento de una distribución uniforme discreta en el intervalo $[0, D]$ garantizando así que haya como mínimo un elemento distinto entre Z_i^t y X_i^t . Por otro lado CR (crossover ratio) debe de tener un valor en el intervalo $[0,1]$ y mide cuánto se van a recombinar los vectores X_i^t y Y_i^t , debido a que valores próximos al 0.5 provocarán que haya bastante combinación, mientras que valores próximos al 0 o al 1 harán que apenas se combinen estos vectores.

- Exponencial, se toma un elemento $n \in 1, 2, \dots, D$ de forma aleatoria y otro elemento $L \in 1, 2, \dots, D$ que denota el número de componentes que van a ser recombinadas. Una vez tomados estos valores podemos definir Z_i^t cómo sigue:

$$Z_{i,j}^t = \begin{cases} Y_{i,j}^t & \text{si } j \equiv n + k - 1 \pmod{D} \text{ para } k = 1, \dots, L \\ X_{i,j}^t & \text{en otro caso} \end{cases} \quad (B.11)$$

De esta forma podemos garantizar que vamos a tener L elementos del vector mutante Y_i^t en el vector Z_i^t , todos de forma consecutiva.

Dependiendo del problema al que se enfrente el algoritmo se debe tomar el enfoque uniforme o el exponencial. Según Ahmad et al. (2022) el método exponencial funciona mejor en aquellos problemas con presencia de correlación entre variables de decisión vecinas.

B.4. Selección

El proceso de selección permite al algoritmo de evolución diferencial a encontrar una solución de la siguiente generación. Debemos de elegir entre un vector recombinado o un elemento de la generación anterior. Una vez se obtenga la nueva población se repite el proceso de mutación, recombinación y selección se recorren de forma iterativa hasta la convergencia.

El proceso de selección se lleva a cabo cómo sigue:

$$X_i^{t+1} = \begin{cases} Z_i^t & \text{si } f(Z_i^t) \leq f(X_i^t) \\ X_i^t & \text{en otro caso} \end{cases}$$

donde $f(\cdot)$ es la función objetivo o en nuestro caso un criterio de error sobre una posible solución. En el caso de que Z_i^t mejore la solución que nos aporta X_i^t el vector recombinado reemplazará al de la generación t en la iteración $t + 1$.

B.5. Evolución diferencial en Robyn

Muchos problemas cómo el que nos plantea Robyn provocan que se deban optimizar varias funciones de forma simultanea a la vez que de forma competitiva. Para estos problemas muchas veces no existe una única solución óptima, pero si un conjunto de soluciones que son más competitivas que el resto. Este conjunto se conoce como las soluciones de Pareto.

Antes de nada, recordar que se está optimizando la selección de los hiperparámetros del modelo, no el reparto del presupuesto. Se definen en la Sección 2.5.4 las cotas de los valores de estos.

El conjunto de soluciones de un problema de optimización multiobjetivo está compuesto por todos los vectores de decisión para los que una componente del vector f no puede ser mejorada sin empeorar otra, este conjunto es conocido cómo frontera de Pareto. Se define la dominancia de Pareto como:

Definición B.5.1 (Dominancia y frontera de pareto). Dados dos vectores $a, b \in \mathbb{R}^D$ decimos que a domina a b si se cumple:

$$\forall i \in \{1, 2, \dots, n\} : f_i(a) \leq f_i(b) \wedge \exists j \in \{1, 2, \dots, n\} : f_j(a) < f_j(b)$$

Los vectores de decisión no dominados son las soluciones de Pareto y el conjunto de todas estas soluciones forma la frontera de Pareto.

Cabe destacar que esta nos aporta un conjunto de modelos que entre los que elegir según las funciones a optimizar de nuestro problema multiobjetivo

Este concepto permite quedarse únicamente con los modelos que presenten valores competitivos de los criterios de error. Los modelos que son dominados por las soluciones de pareto son descartados debido a que presentar peores valores para los criterios de error definidos por Robyn.

Hay que definir un criterio con el que seleccionar el mejor modelo en el caso de estar utilizando $DE/curr - to - best/2$ en la fase de mutación y para poder tomar una decisión en la fase de selección definida en B.4. Si la solución candidata domina a la solución original en la fase de selección del modelo se toma esa solución como la i -ésima solución de la siguiente generación.

La librería de R **Robyn** conecta con la librería de Python **Nevergrad** donde se encuentran múltiples algoritmos, entre ellos el algoritmo de evolución diferencial utilizado para la optimización de los hiperparámetros del modelo.

Hay muchas variaciones dentro de los algoritmos evolutivos. Robyn escogió utilizar la transformación [B.9](#) en el proceso de mutación y la [B.11](#) para la fase de recombinación. Por otro lado entre los parámetros del algoritmo que deben elegir se tomaron $NP = 30$ individuos por generación. En el proceso de mutación se toma $F = 0.8$, posteriormente en la fase de recombinación toma L de una forma aleatoria. En cuanto a la adaptación multivariante, afecta tanto al apartado de mutación (al utilizar [B.9](#)) como al de selección del algoritmo. Está implementado. Además, como los marketing mix models suelen tener variables de marketing similares entre sí para la fase de recombinación del algoritmo se ha optado por utilizar la recombinación exponencial.

Al estar trabajando con NP individuos por población, las iteraciones se pueden obtener de forma paralela. Esto provoca que el algoritmo sea algo más rápido que otros algoritmos donde tienes que ir ajustando iteraciones una a una. Para ello en la función *robyn_run* se utiliza la librería **doParallel** de R que contiene la función **foreach** que permite ajustar los modelos de Robyn de forma paralela.

Bibliografía

- [1] Ahmad, M. F., Mat Isa, N. A., Lim, W. H. y Ang, K. M. (2022), Differential evolution: A recent review based on state-of-art works. *Alexandria Engineering Journal*, **61**(5),. <https://doi.org/10.1016/j.aej.2021.09.013>
- [2] Borden N. H.,(1964) The Concept of the Marketing Mix. *Journal of Advertising Research*, **4**(2), 2-7.
- [3] Britannica, The Editors of Encyclopaedia. "John Wanamaker". Encyclopedia Britannica (2022), <https://www.britannica.com/biography/John-Wanamaker>. Acceso 13 Abril 2023.
- [4] Conn, A. R., Gould N.I.,M, Toint P. L. (1991) A globally augmented Lagrangian algorithm for optimization with general constraints and simple bounds. *SIAM Journals on Numerical Analysis*, **28**(2), p. 545-572
- [5] Taylor, S., Letham, B. (2021) prophet: Automatic Forecasting Procedure. <https://cran.r-project.org/web/packages/prophet/prophet.pdf>. Versión 1.0
- [6] Sax, C., Steiner, P. (2013) tempdisagg: Methods for Temporal Disaggregation and Interpolation of Time Series. <http://CRAN.R-project.org/package=tempdisagg>. Versión 0.22. [p80]
- [7] Crujeiras Casais, R.M., Conde Amboage, M. (2019) El modelo de regresión lineal simple. Inferencia Estadística-Grado en Matemáticas.
- [8] Gesztelyi, R., Zsuga, J., Kemeny-Beke, A., Varga, B., Juhasz, B., Tosaki, A. (2012), The Hill equation and the origin of quantitative pharmacology. *Archive for History of Exact Sciences*, **66** (4), 427-438. [doi:10.1007/s00407-012-0098-5](https://doi.org/10.1007/s00407-012-0098-5)
- [9] Gobierno de España, Ministerio de Asuntos Económicos y Transformación digital. Indicadores e informes macroeconómicos. https://serviciosedes.mineco.gob.es/Indeco/BDSICE/Busquedas/Busquedas_new.aspx
- [10] Hoerl, A. E., Kennard, R. W.(1970) Ridge Regression: Applications to Nonorthogonal Problems, *Technometrics*, **12**:1, 69-82. [10.1080/00401706.1970.10488635](https://doi.org/10.1080/00401706.1970.10488635)
- [11] Johnson, S. G. (2023) The NLOpt nonlinear-optimization package, <http://github.com/stevengj/nlopt>. Accedido en 15 de abril de 2023.
- [12] Judge, G. G., Takayama, T. (1966) Inequality restrictions in regression analysis. *Journal of the American Statistical Association*, **61**(313), 166-181. <https://doi.org/10.2307/2283052>
- [13] Jin, Y., Shobowale, S., Koehler, J., Caso, H. (2012) The incremental reach and cost efficiency of online video ads over tv ads. research.google.com. Accedido 14-05-2023
- [14] Jin, Y., Wang, Y., Sun, Y., Chan, D., Koehler, J. (2017) Bayesian Methods for Media Mix Modeling with Carryover and Shape Effects. Google Inc. research.google.com. Accedido 05-02-2023

- [15] Kraft D. (1988) A software package for sequential quadratic programming, *Technical Report DFVLR-FB-88-28*, Institut für Dynamik der Flugsysteme, Oberpfaffenhofen.
- [16] Lawless, J. F. (2002) Basic Concepts and Models. In *Statistical Models and Methods for lifetime data*. John Wiley & Sons.
- [17] López Pouso, R. (2019) Series de Fourier y ecuaciones en derivadas parciales. USC Editora.
- [18] Nield, T. (2016) Getting Started with SQL: A Hands-On Approach for Beginners. O'Reilly Media, Inc.
- [19] Shumway, R. H., Stoffer, D. S. (2017) Time Series Analysis and Its Applications. With R Examples. Springer (4^a Edición)
- [20] Storn, R., Price, K. (1997) Differential Evolution- A Simple and Efficient Heuristic for global Optimizaton over continuous Spaces. *Journal of Global Optimization*, **11**, 341-359
- [21] Taylor, S., Letham, B. (2017) Forecasting at Scale. *The American Statistician*, **72**, 37-45
- [22] Tellis, G. J. (2006) Modeling Marketing Mix. In R. Grover & M. Vriens (Eds.), *Handbook of marketing research: uses, misuses, and future advances*. Thousand Oaks, CA (p. 506-522)
- [23] Tibshirani, R. (1996) Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, **58**(1), 267-288
- [24] Toker, S., Siray, G.Ü., Kaçiranlar, S. (2013) Inequality constrained ridge regression. *Statistics & Probability Letters*, **83**(10) 2391-2398. <https://doi.org/10.1016/j.spl.2013.06.023>
- [25] Wang, Y., Jin, Y., Sun, Y., Chan, D., Koehler, J. (2017), A Hierarchial Bayesian Approach to Improve Media Mix Models Using Category Data. Google Inc., research.google.com
- [26] Wood, S. (2017) Generalized Additive Models: an introduction with R. Springer 2^a Edición.
- [27] Zhou, G., Sentana, L., Skokan, I., Prada, A. (2021), Marketing Mix Modelling Using Multi-objective Hyperparameter Optimization, Technical Disclosure Commons, https://www.tdcommons.org/dpubs_series/4627.
- [28] Zhou, G., Sentana, L., Skokan, I., Lares, B. (2023) Robyn: Semi-Automated Marketing Mix Modeling (MMM) from Meta Marketing Science. Meta Platforms, Inc <https://cran.rstudio.com/web/packages/Robyn/Robyn.pdf>, <https://facebookexperimental.github.io/Robyn/>. Accedido 13 de enero de 2023.
- [29] Zou, H., Hastie, T. (2005) Regularization and variable selection via the elastic net, *Journal of the Royal Statistical Society*, **67**(2), 301-320.