



FACULTAD DE CIENCIAS ECONÓMICAS Y EMPRESARIALES

# INDICADORES SINTÉTICOS

Medición de la recuperación del consumo en España  
tras el Covid-19

Autor: Elena Campón Acuña

Director: Dr. David Roch Dupré

Madrid. Abril 2023.



## RESUMEN

Un indicador simple o individual es un conjunto de datos que caracterizan una realidad concreta, de los que se puede inferir su futura evolución. Los indicadores sintéticos o compuestos, como su nombre indica, son una agregación de aquéllos. De esta manera, se constituyen como una herramienta útil que permite monitorear la eficacia de las políticas implementadas y el desempeño de un país en un ámbito concreto.

El presente trabajo tiene como objetivo principal la construcción de un indicador sintético lo más objetivo y preciso posible, que va a mostrar el consumo en España, con el fin de determinar la evolución y recuperación del Covid-19.

## PALABRAS CLAVE

Indicador sintético; consumo; datos; SmartVision 360; Covid-19; España.

## ÍNDICE

<b>1. INTRODUCCIÓN: MARCO TEÓRICO</b> .....	4
<b>1.1. El consumo y la iniciativa 360° Smart Vision</b> .....	5
<b>1.2. Indicadores sintéticos: estado del arte</b> .....	8
<b>1.3. Motivación</b> .....	11
<b>1.4. Objetivos, recursos y metodología</b> .....	13
<b>2. CONSTRUCCIÓN DEL INDICADOR SINTÉTICO</b> .....	15
<b>2.1. Selección de los datos</b> .....	15
<b>2.2. Imputación de la información ausente</b> .....	24
<b>2.3. Análisis Multivariante</b> .....	35
<b>2.4. Normalización de los datos</b> .....	43
<b>2.5. Pesos y agregaciones</b> .....	46
<b>2.6. Análisis de la robustez y sensibilidad</b> .....	56
<b>3. CONCLUSIONES</b> .....	59
<b>4. BIBLIOGRAFÍA Y REFERENCIAS.</b> .....	60

### 1. INTRODUCCIÓN: MARCO TEÓRICO

## 1.1. El consumo y la iniciativa 360° Smart Vision

La Enciclopedia Económica define el consumo como la «acción de adquirir bienes y servicios para satisfacer necesidades» (2021). En el ámbito de la macroeconomía, numerosos autores, como Simon Kuznets, James Duesenberry, Franco Modigliani o Milton Friedman, han desarrollado teorías sobre este fenómeno. Sin embargo, fue John Maynard Keynes el primero en afirmar que el consumo depende fundamental y principalmente de la renta de las personas, y si bien otros factores pueden influir en el mismo, no tienen la importancia suficiente como para llegar a ser determinantes en el nivel de consumo (Morettini, 2022). Así, este autor presenta el consumo como resultado de la siguiente ecuación (1):

$$\begin{aligned} C &= f(Yd) \\ C &= a + b * (Yd) \end{aligned} \tag{1}$$

Donde:

- $C$  es el consumo privado o final, que viene determinado por los siguientes factores.
- $a$  es el consumo autónomo, que no depende de la renta.
- $b$  es la propensión marginal a consumir (PMC), que indica cuánto varía el consumo cuando la renta aumenta en una unidad.
- $Yd$  es el ingreso disponible.

Así, el consumo privado de las personas depende, por un lado, del consumo autónomo, y por otro, de la renta. En el consumo autónomo se engloban ciertos factores que influyen en menor medida en el nivel de consumo de las personas, como son el nivel de riqueza de estas o del país en el que viven, el nivel de incertidumbre de sus ingresos futuros, la edad, las expectativas de evolución del país, o el tipo de interés del momento (Cabrera et al., 2019). Keynes afirmó que el consumo aumenta con la renta, si bien no en la misma proporción. Así, la PMC es una constante que determina en cuánto aumenta el consumo por cada punto más de renta (Morettini, 2022).

Como, de acuerdo con Keynes, el consumo y la renta no crecen en la misma proporción, el porcentaje de consumo de una persona respecto de su renta tiene una tendencia decreciente. En otras palabras, una persona con menor renta, consume más -en proporción a su renta- que una persona con una renta más elevada. La lógica que subyace

de este razonamiento tiene un gran sentido a nivel práctico, ya que una persona que obtenga de renta 2.000 € al mes, deberá destinar un porcentaje mucho mayor a la satisfacción de las necesidades básicas de la vida que aquella que tenga una renta de 6.000 €. Así, se podría decir que las personas con mayor renta ahorran más.

Con todo lo anterior, se puede ver que la renta, el consumo y el ahorro son tres términos altamente relacionados, en tanto que el consumo es aquella parte de la renta que se destina a la inmediata satisfacción de las necesidades básicas, o de deseos o caprichos que tienen las personas. Todo aquello que no se destine al consumo, constituirá el ahorro de las personas.

Esto nos muestra cómo el consumo es una herramienta útil para determinar en qué estadio se encuentra la economía de un país. Cualquier lector podría haber llegado a esta conclusión, ya que, dejando de un lado el plano más teórico y adentrándonos en uno más práctico, se podría decir que el consumo está presente en la vida diaria de las personas, y lo ha estado durante milenios. Las distintas ramas de la economía, y cada empresa en particular, destinan sus recursos para el estudio de los hábitos de consumo del mercado, con la finalidad de conseguir una economía más eficiente y una posición más competitiva para el caso de las empresas. No obstante, estos hábitos se han visto transformados por la reciente pandemia Covid-19, ya que, el consumo, al igual que la mayoría de los factores económicos de los que depende la economía de un país, es un fenómeno que se ha visto altamente afectado por mencionada pandemia. Ello ha resultado en consumidores con una menor capacidad económica y por tanto más analíticos en cuanto a la relación calidad-precio de los productos y servicios que adquieren, más preocupados por la salud y la seguridad en general, y mucho más digitales (KPMG, 2020; PwC, 2021).

En primer lugar, el nuevo mercado se encuentra con consumidores más preocupados con el precio de los productos que adquiere. De acuerdo con un estudio llevado a cabo por KPMG en noviembre de 2020, la paralización de la actividad empresarial durante la pandemia ha conllevado un aumento del desempleo, y una consiguiente bajada en los ingresos de las familias, tanto a nivel mundial como nacional. En concreto, un 57% de los españoles entrevistados por la compañía indican haber notado un empeoramiento en su economía, frente a un 47% globalmente (p.7). Por ello, el consumidor actual tiene más presente su nivel de gasto, y por tanto prioriza la calidad-precio de los productos que adquiere. En concreto, el consumidor actual procura tener el menor nivel de gasto posible,

y se preocupa más por el ahorro. Un sondeo llevado a cabo por PwC en junio del 2021 demuestra cómo un 52% de los españoles están preocupados por el ahorro, frente a un 31% antes de la pandemia (p.15). Asimismo, el precio de los productos y servicios orientaba al 29% de los consumidores españoles en sus compras antes del Covid-19, pero hoy día un 59% de compradores se ven orientados por el precio (p.15).

En segundo lugar, la crisis sanitaria que ha sido producto de la pandemia ha llevado a la población a preocuparse más por la salud y seguridad. Un cuarto de los consumidores globales está preocupado por estos aspectos tras la pandemia (PwC, 2021, p.7). Esta preocupación se manifiesta en una doble vertiente, de acuerdo con los estudios llevados a cabo por KPMG (2020) y PwC (2021). Por un lado, los ciudadanos tienden a evitar los espacios públicos. Así, los compradores son menos propensos a comprar en tiendas físicas y empiezan a preferir la compra mediante métodos digitales. Los trabajadores, a su vez, toman el teletrabajo como una buena alternativa para garantizar la salud y seguridad. Por otro lado, los consumidores son más locales, en el sentido de que son más propensos a consumir productos locales y marcas nacionales, y buscan una relación de confianza con las marcas, que deben garantizar cierta cercanía a sus clientes.

En tercer lugar, los ciudadanos se han digitalizado. Esto también se expresa en una doble vertiente. Por una parte, se ha extendido la compra por internet de muchos productos que antes apenas se adquirían por dicho medio. Por ejemplo, los alimentos. Por otra parte, la facilidad de acceder a lo digital y la proliferación de los servicios tecnológicos ha generalizado el teletrabajo, y hoy día muchas empresas lo ofrecen como una alternativa al trabajo presencial.

Teniendo en cuenta todos estos aspectos, se puede ver cómo la pandemia ha influido considerablemente en el comportamiento del consumidor medio, y tal y como indica el análisis llevado a cabo por PwC, se trata de una «transformación histórica en el comportamiento del consumidor» (2021, p.2), en tanto que los cambios se están manteniendo en el tiempo, y no se vuelve a la realidad previa a la pandemia.

Junto con esta transformación en el comportamiento del consumidor, las economías mundiales se enfrentan ahora a la recuperación económica de la crisis provocada por la pandemia. Por ello, la Universidad Pontificia Comillas y Deloitte han creado la plataforma 360° Smart Vision, que tiene como principal objetivo ayudar de una manera objetiva y transparente a que los agentes económicos, sociales y políticos determinen la

eficacia de las políticas implementadas para la recuperación económica de los mercados. Con este análisis, se podría buscar maneras de mejorar el rendimiento de los países, y encontrar soluciones más efectivas. La plataforma divide sus datos en ocho categorías: actividad empresarial, mercado laboral, mercados financieros, sostenibilidad, ciencia ciudadana, sanidad, movilidad, y consumo. El presente trabajo abordará la cuestión desde el punto de vista del consumo.

Una herramienta útil que permite monitorear la eficacia de dichas políticas y proponer soluciones más eficientes son los indicadores sintéticos, que, como primera aproximación, se podrían definir como la agregación de otros indicadores simples para representar una realidad multidimensional. Este trabajo tiene como principal objetivo la construcción de un indicador sintético que monitoree el consumo en España y sea de utilidad práctica para los distintos agentes que forman parte de la economía del país.

## **1.2. Indicadores sintéticos: estado del arte**

Un indicador simple o individual es un conjunto de datos que caracterizan una realidad concreta, de los que se puede inferir su futura evolución. Así, los indicadores resultan una herramienta útil en el análisis de aquellos hechos que se desarrollan de manera prolongada en el tiempo y cuya evolución es de interés. Sin embargo, hay muchas realidades que no dependen de un solo factor, sino que se componen de varias dimensiones. Ejemplos de ello pueden ser la competitividad, el comercio, o el impacto de las políticas implementadas por las administraciones públicas en un país. Especialmente en un momento como el actual, se está presenciando un desarrollo e incremento en el número de las dimensiones de las que una realidad depende, debido al auge de los datos a los que se tiene acceso. En otras palabras, la revolución tecnológica que ha traído consigo una mayor generación, acceso y almacenamiento de datos, implica una expansión en el conocimiento de nuevas -o ya existentes- dimensiones de las que dependen distintas realidades., pero cuya extracción de datos para su análisis resultaba prácticamente imposible antes- Estas realidades, debido a su complejidad e influencia por diversos componentes, son imposibles de representar mediante un único indicador, que se ve limitado por analizar una sola dimensión (Greco *et al.*, 2018). Por ello, se ha desarrollado lo que se conocen como indicadores sintéticos o compuestos, que se podrían definir como



un resumen, una combinación de índices simples, que tienen por objetivo representar de manera más fehaciente una realidad que se compone de múltiples dimensiones.

Verdaderamente no existe una única definición oficial para los indicadores sintéticos, pero son muchos los autores que han tratado de responder a esta pregunta. Así, el centro de investigación de la Comisión Europea ha desarrollado un informe sobre de los indicadores sintéticos (Saisana y Tarantola, 2002, p.5), definiéndolos de la siguiente manera:

*«Composite indicators are based on sub-indicators that have no common meaningful unit of measurement and there is no obvious way of weighting these sub-indicators»*

Una traducción dada a la anterior definición de indicadores sintéticos es:

«Los indicadores sintéticos son aquella combinación (o agregación) matemática de los indicadores que representan los distintos componentes del concepto que se pretende evaluar a partir de un sistema de partida, proporcionando una evaluación multidimensional del mismo» (Blancas et al., 2011, p.3)

Freudenberg (2003) nos da una definición más breve, refiriéndose a los indicadores compuestos como «índices sintéticos de otros indicadores individuales o simples» (p.5). En la misma línea, Saltelli (2007) indica que los indicadores sintéticos son una «manipulación de los distintos indicadores simples, así como sus posibles pesos, para generar una unidad de medida agregada ordinal o cardinal que pueda determinar rendimiento de un país» (p.65). Este último apunte que hace Saltelli sobre la necesidad de crear indicadores sintéticos para estudiar la evolución de un país es de especial interés para el presente trabajo, pues es objeto del mismo crear un indicador sintético que sirva como herramienta para determinar la eficacia de las políticas implementadas en España para la recuperación del Covid-19, y ver la evolución del país, concretamente en el ámbito del consumo.

Como se puede observar, todas las definiciones dadas convergen en la idea de que los indicadores compuestos sintetizan la información de otros indicadores más simples, que se agregan con el fin de mostrar una realidad multidimensional. No obstante, si bien el uso de los indicadores sintéticos para el análisis de diferentes realidades se ha expandido notablemente en los últimos años, y tal y como anuncian Blancas et al., (2011) «se han configurado como uno de los instrumentos de medición analítica más utilizados en la práctica en multitud de campos de la realidad social» (p.3), no dejan de existir ciertas

objecciones a su uso, pues algunos críticos se cuestionan su utilidad objetiva, así como cuál debe ser su alcance.

Resulta incuestionable que los indicadores sintéticos son una herramienta de gran utilidad práctica, pero no por ello se debe dejar atrás los inconvenientes que traen consigo. Uno de los aspectos más cuestionados es que resulta difícil conseguir indicadores sintéticos objetivos. Como ya hemos dicho, un indicador compuesto es una agregación de otros indicadores simples. A cada uno de estos indicadores simples se les va a dar una ponderación concreta, que determinará su peso en el indicador compuesto. La persona que determina dichos pesos es el creador del indicador compuesto. Por ello, se critica a los indicadores compuestos de ser subjetivos y poco representativos de la realidad. Así, una de las cuestiones más relevantes y de la que se hablará en puntos posteriores del presente trabajo es cómo conseguir determinar de la manera más objetiva y real posible los pesos que se debe dar a cada índice.

Si bien la posible subjetividad de los indicadores sintéticos son el punto más desaprobado por los críticos, hay otros que no se deben pasar por alto. Siguiendo a la OECD (2008, pp. 13 y 14), las ventajas y desventajas de los indicadores sintéticos se pueden resumir en los puntos presentados en la tabla 1:

VENTAJAS	INCONVENIENTES
<ul style="list-style-type: none"> <li>- Resumen realidades complejas y multidimensionales de tal manera que ayudan a la toma de decisiones eficiente.</li> <li>- Es más fácil interpretar un único indicador sintético que varios simples.</li> <li>- Evalúan la evolución de los países a lo largo del tiempo.</li> <li>- Campo visual más reducido, pero sin pérdida de información; recoge todos los datos importantes.</li> <li>- Se incluye más información en un campo visual más reducido.</li> <li>- El rendimiento y progreso de los países está en el punto central de estudio.</li> <li>- Facilitan la comunicación con el público en general, y son buena fuente de justificación de ciertas políticas.</li> <li>- Ayudan a la comprensión de la cuestión a personas legas en la materia.</li> </ul>	<ul style="list-style-type: none"> <li>- La falta de rigor en su construcción o interpretación pueden llevar a conclusiones erróneas.</li> <li>- Las conclusiones que se obtienen de ellos pueden ser simplistas.</li> <li>- Si no se construyen de manera objetiva y transparente, las agregaciones y datos se pueden manipular, de tal manera que se lleguen a conclusiones cuya implementación satisfaga intereses preestablecidos.</li> <li>- La selección de los indicadores y sus respectivos pesos se podría hacer de manera arbitraria y siguiendo una política.</li> <li>- Con la falta de transparencia en su construcción se pueden ocultar defectos en algunas dimensiones, y aumentar la dificultad para encontrar remedios a los mismos.</li> <li>- Puede llevar a políticas inapropiadas e incorrectas si no se incluyen las dimensiones difíciles de medir.</li> </ul>

<p>- Permiten a los usuarios comparar dimensiones complejas de una manera efectiva.</p>	
---	--

**Tabla 1:** Tabla de ventajas y desventajas de los indicadores sintéticos.  
*Fuente: OECD (2008).*

Con el fin de evitar la creación de indicadores compuestos arbitrarios e inútiles, la OECD ha desarrollado un manual orientativo que establece las bases para su creación, de tal manera que se consigan indicadores sintéticos lo más objetivos y transparentes posible. Se trata de un proceso que consta de diez pasos, conocido como *checklist* (Greco *et al*, 2018), y son los siguientes:

- 1) Construcción de un marco teórico que centre los objetivos del indicador sintético.
- 2) Selección de los datos y variables a utilizar.
- 3) Imputación de valores perdidos.
- 4) Análisis multivariable para evitar la selección arbitraria de los datos, y sin correlación.
- 5) Normalización de los datos.
- 6) Determinar los pesos y agregaciones.
- 7) Análisis de la robustez y sensibilidad.
- 8) Resumen.
- 9) Conexión con otros indicadores.
- 10) Presentación visual del resultado final.

### **1.3. Motivación**

El nivel de consumo de un país es una herramienta útil para determinar en qué estadio se encuentra la economía del mismo. Tanto es así, que el consumo es uno de los componentes del Producto Interior Bruto o PIB (guías jurídicas, s. f.), que es definido por la Enciclopedia Económica como «un indicador económico que muestra el valor total de bienes y servicios que se han generado en el ámbito económico durante un período de tiempo» (2018), y es el indicador más generalizado y mundialmente utilizado para medir la economía y prosperidad de un país.

Existen tres métodos para el cálculo del PIB: el método del valor añadido, el método de las rentas, y el método del gasto. Este último es el que interesa a efectos del presente trabajo, y consiste en la agregación de los gastos finales de los distintos agentes

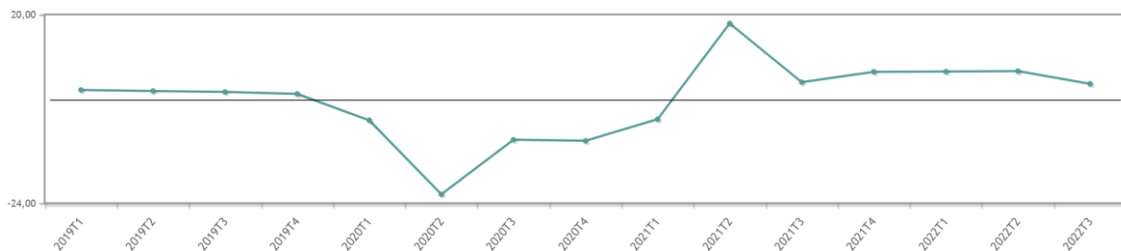
económicos en una economía. Se representa de la forma expresada en la ecuación (2) (Gastón Lorente, 2020):

$$PIB = C + I + G + X - M \quad (2)$$

Donde:

- C es el consumo
- I es la inversión de las empresas y familias
- G es el gasto en consumo final del sector público
- X – M es el valor de las exportaciones netas: X (exportaciones) – M (importaciones)

La figura 1 muestra cómo el PIB español cayó a valores negativos durante la pandemia, lo cual podría ser, en caso de tener una correlación positiva con el índice de consumo, prueba de una caída del mismo, pues «cuando en una economía varía la producción, la renta de los integrantes de esa economía también varía y, al variar la renta, también lo hace el consumo» (Durán Durán, 2022).



**Figura 1:** Variación anual del PIB español en los últimos 15 trimestres.

*Fuente:* Instituto Nacional de Estadística (2022).

Las políticas modernas centran sus esfuerzos en la recuperación de las crisis producidas por el Covid-19. Muchas de ellas se relacionan directamente con el consumo, que se vio altamente afectado por la pandemia, y hoy se está recuperando de la misma. Así, surge la necesidad de crear un indicador sintético que monitoree el consumo, para poder determinar, la eficacia de las políticas implementadas, y asegurar una recuperación del Covid-19 lo más eficaz y rápida posible. Cuando las economías actuales se recuperen de la pandemia, el indicador sintético será igualmente útil en tanto que aparecerán nuevos retos, y el indicador siempre ilustrará de una manera transparente y objetiva el desempeño de las economías en cualquier momento.

#### 1.4. Objetivos, recursos y metodología

El presente trabajo tiene como objetivo principal la construcción de un indicador sintético lo más objetivo y preciso posible, que va a mostrar el consumo en España, con el fin de determinar la eficacia de las políticas implementadas en los últimos años para la recuperación del Covid-19. Para ello, se va a seguir los pasos descritos en el manual oficial publicado por la Comisión Europea (OECD, 2008). Así, se hará tanto la aproximación teórica a cada paso, como la construcción del indicador (que será la aplicación práctica de dicha teoría). En concreto, las tareas a seguir para conseguir la construcción de dicho indicador son las que se exponen a continuación.

La primera tarea es la selección de los datos. Se va a acudir a los proporcionados por la plataforma 360° Smart Vision para la categoría de consumo. El primer paso consta en estudiar el conjunto de datos del que se dispone y determinar los años con los que se va a trabajar, así como su frecuencia, y las categorías en las que se van a agrupar los distintos KPIs.

Debido a que los datasets suelen presentar valores faltantes, o datos en distintas frecuencias, la segunda tarea es la imputación de los valores ausentes. Existen diversas técnicas de imputación, por lo que se precisa de un estudio teórico de las mismas para determinar cuál procede llevar a cabo en este caso. Tantas técnicas como se concluya que pueden ser adecuadas para la imputación de los datos faltantes en este caso se llevarán a la práctica, y nos quedaremos con los valores de aquella que proporcione el menor margen de error. Para ello se hará uso de dos herramientas distintas: IBM SPSS *Statistics* y Matlab, siendo aquella un programa estadístico informático, y ésta una plataforma de programación y cálculo que ofrecen distintos métodos de imputación de valores ausentes.

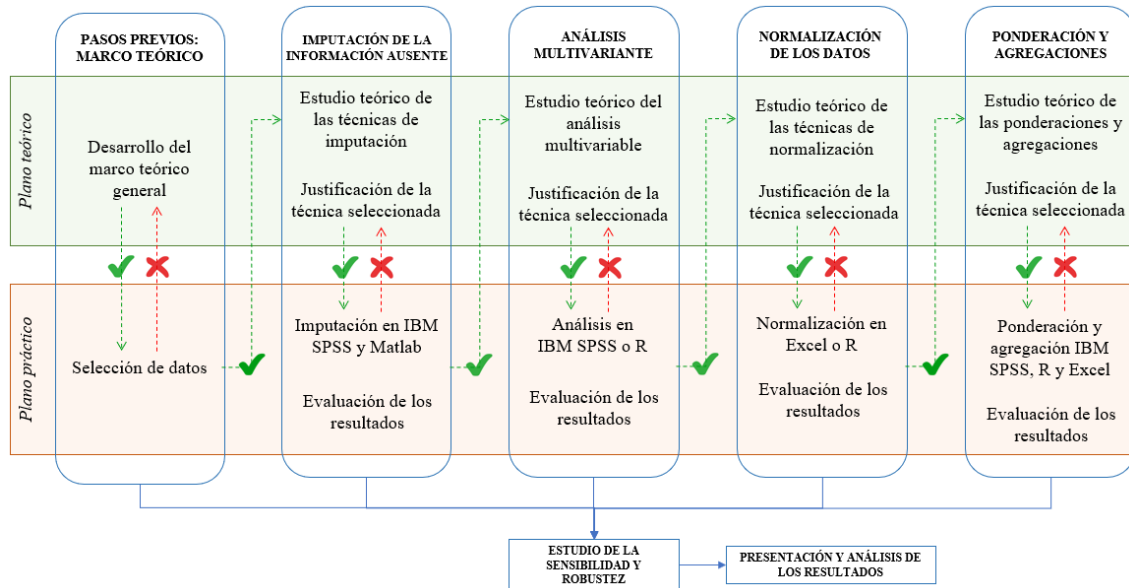
La siguiente tarea es el análisis multivariable, que consiste en determinar las características que presenta el dataset. En otras palabras, permite analizar su estructura antes de empezar a trabajar con el mismo. Es útil para determinar las correlaciones que puedan existir entre los distintos KPIs, los grupos en que se puede agrupar el dataset, y averiguar otros patrones que no se puedan apreciar a simple vista. Hay diversas técnicas que permiten llevar a cabo un análisis multivariable, por lo que se hará un estudio de cada una de ellas y se procederá con aquella que resulte más adecuada para los datos en cuestión. Para este análisis se hará uso de las herramientas RStudio y SPSS.

Una vez se tenga analizada la estructura del dataset completo e imputado, procede normalizar los datos. Así, todos los datos pasarían a estar en la misma escala y unidad, evitando que unos tengan mayor peso que otros. De nuevo, se hará una aproximación teórica a las técnicas de normalización, y se aplicará la más adecuada para los datos en cuestión. Los recursos con los que se llevarán a cabo este paso van a ser RStudio o Microsoft Excel. R es una plataforma de programación y análisis estadístico que permite normalizar un conjunto de datos en una sola línea de código, mientras que Excel es una hoja de cálculo que permite llevar a cabo operaciones de una manera más manual. Una vez normalizados los datos, se evaluarán antes de pasar a la fase siguiente.

La quinta tarea a llevar a cabo para el logro de los objetivos del presente trabajo es determinar los pesos y agregaciones que cada variable va a tener en el indicador sintético. Se trata de la parte más controvertida de la construcción de los indicadores, pues se puede caer en arbitrariedades y tergiversar el indicador sintético por darle a ciertos KPIs un peso mayor o menor al que verdaderamente debería corresponderle. Por ello, se acudirá, en la medida de lo posible, a métodos estadísticos que justifiquen la asignación de pesos, y se reflejará de manera transparente todos los pasos llevados a cabo, para evitar caer en errores que puedan dejar inútil el indicador. Existen, asimismo, diferentes métodos para dar pesos y agregaciones a los indicadores, por lo que, al igual que en la imputación de los valores ausentes, se hará una primera aproximación teórica. De todas las técnicas, se escogerá aquella o aquellas que se consideren apropiadas para el caso concreto, y se llevarán a la práctica. Este paso se llevará a cabo con las herramientas Excel, SPSS y R.

Por último, se llevará a cabo un análisis de la sensibilidad y robustez del indicador, para determinar si el indicador es efectivamente acorde a la realidad o no. Si lo es, se procederá a la presentación de los resultados.

De manera más visual y sintética, todos estos pasos se resumen en el gráfico de diagrama de flujo representado en la figura 3:



**Figura 2:** Gráfico de diagrama de flujo.  
Fuente: Elaboración propia.

## 2. CONSTRUCCIÓN DEL INDICADOR SINTÉTICO

### 2.1. Selección de los datos

El punto de partida de un indicador sintético es la extracción y selección de datos sobre los que el mismo se va a construir. Ya se adelantó en la Introducción que los datos con los que se va a trabajar en el presente trabajo son los facilitados por la plataforma SmartVision 360°. En concreto, los relativos al consumo, y se va a tratar de medir la recuperación del consumo en España tras la pandemia del Covid-19.

La estructura de la base de datos es la siguiente: hay un total de 20.616 observaciones para 7 ámbitos territoriales distintos, y que se agrupan en un total de 167 KPIs o indicadores sintéticos. Estos, a su vez, se agrupan en tres grandes grupos o categorías (consumo general, consumo sectorial y *consumer tracker*), de las que se hablará más adelante. La Tabla 2 hace un resumen visual de la estructura de la base de datos en cuestión:

RESUMEN BBDD CONSUMO					
TERRITORIO	Nº OBS.	Nº KPIs.	Nº CAT.	FREC.	FECHAS
Eurozona	32	1	1	Mensual	oct.19 - may.22
Portugal	34	5	2	Mensual	ene.20 - ago.20
Italia	361	21	3	Mensual	ene.20 - ago.21
Alemania	362	21	3	Mensual	ene.20 - ago.21
Francia	362	21	3	Mensual	ene.20 - ago.21
Reino Unido	362	21	3	Mensual	ene.20 - ago.21
España	19.103	77	3	Mensual	jun.19 - may.22

**Tabla 2:** Resumen general de la base de datos relativa al consumo<sup>1</sup>.

*Fuente:* Elaboración propia.

Parece clara cuál debe ser la primera decisión a tomar para la construcción del indicador en cuestión, y es qué país o países se pretende medir en el mismo.

- La Eurozona dispone de un total de 32 observaciones para 32 meses (de octubre del año 2019 a mayo 2022, ambos meses incluidos), relativos a una sola variable (o KPI) y por tanto, también una única categoría. Si bien la variable dispone de observaciones suficientes para extraer conclusiones respecto de la misma, no parece que una realidad como lo es el consumo pueda ser medida desde una única variable, en este caso, el índice de confianza del consumidor (a partir de ahora, ICC), que representa la intención de gasto del consumidor, y es una única dimensión de todas las que componen el consumo. Por ello, se descarta la Eurozona como posible objeto de análisis, por carecer ésta de suficientes datos sobre los que asentar el indicador sintético.

- Portugal cuenta con un total de 34 observaciones, pertenecientes a 5 KPIs. Dos de ellos tienen datos de enero a agosto del año 2020, y los tres restantes cuentan con datos de enero a junio del mismo año. Todos ellos en frecuencia mensual. Teniendo en cuenta que, por un lado, ninguna de estas cinco variables presenta datos para ningún año completo -por lo que se tendría que imputar muchos datos- y, por otro lado, tal y como ocurre con la Eurozona, el número de KPIs disponible no es del todo representante de un área tan amplio como lo es el consumo, Portugal queda descartado como país objeto de análisis.

- Italia, Alemania, Francia y Reino Unido presentan una estructura de datos idéntica (nótese que Italia cuenta con una observación menos). Esto es porque carece del dato correspondiente al mes de abril 2020 del Índice de Confianza del Consumidor (en

<sup>1</sup> “BBDD” son las siglas para “bases de datos”. “Nº obs.” hace referencia al número de observaciones del territorio en cuestión; “Nº KPIs” al número de variables en que se agrupan tales observaciones; “Nº cat.” al número de categorías en que se agrupan tales variables; “frec.” a la frecuencia en que se encuentran los datos en cuestión.



adelante, ICC)), que se puede ver en la Tabla 3. En ella, las casillas en verde representan los meses en que hay valores disponibles, en frecuencia mensual. Las casillas verdes con “x2” indican una frecuencia bisemanal. Las casillas en rojo representan ausencia de datos para tales meses. Los meses de septiembre 2020 a agosto 2021 presentan la misma estructura, por lo que se ha sintetizado con puntos suspensivos.

RESUMEN BBDD CONSUMO PARA IT, AL, FR, RU											
KPI	CATEGORÍA	ene-20	feb-20	mar-20	abr-20	may-20	jun-20	jul-20	ago-20	sep-20	... ago-21
ICC	Consumo general										
ICPM	Consumo general										
INC alimentación	Consumer tracker				x2	x2	x2	x2	x2		
INC electrónica	Consumer tracker				x2	x2	x2	x2	x2		
INC entrenimiento	Consumer tracker				x2	x2	x2	x2	x2		
INC gasolina	Consumer tracker										
INC Internet/móvil	Consumer tracker				x2	x2	x2	x2	x2		
INC libros	Consumer tracker										
INC medicamentos	Consumer tracker				x2	x2	x2	x2	x2		
INC muebles	Consumer tracker				x2	x2	x2	x2	x2		
INC restauración	Consumer tracker				x2	x2	x2	x2	x2		
INC textil	Consumer tracker				x2	x2	x2	x2	x2		
INC TV	Consumer tracker				x2	x2	x2	x2	x2		
INC viajes	Consumer tracker				x2	x2	x2	x2	x2		
M.V.	Consumo sectorial										
M.V. empresas	Consumo sectorial										
M.V. particulares	Consumo sectorial										
Preoc. Económica	Consumer tracker				x2	x2	x2	x2	x2		
Preoc. Salud	Consumer tracker				x2	x2	x2	x2	x2		
Preoc. Pagos	Consumer tracker				x2	x2	x2	x2	x2		
Retraso compras	Consumer tracker				x2	x2	x2	x2	x2		

**Tabla 3:** Resumen estructura de datos disponibles para Italia, Alemania, Francia y Reino Unido.

*Fuente:* Elaboración propia.

Véase cómo estos países ya presentan una base de datos más sólida, con un mayor número de KPIs y de categorías a las que los mismos pertenecen. Ya se ha adelantado que la base de datos cuenta con un total de tres categorías en las que divide el consumo, y son, consumo general, consumo sectorial, y *consumer tracker*. Para el caso concreto de Italia, Alemania, Francia y Reino Unido véase cómo el consumo general queda constituido únicamente por dos variables, y consumo sectorial por tres (de las que se dispone únicamente de un 50% de los datos para el año 2020 (de enero a junio de tal año)). Se podría concluir que la única categoría que verdaderamente cuenta con un número significativo de variables y observaciones es *consumer tracker*.

- España dispone de poco más de 19 mil observaciones, que a su vez se dividen por el total de todo el país y por Comunidad Autónoma. Como los datos de total España se entiende que ya contemplan los pertenecientes a cada Comunidad Autónoma, se escogerá los datos de total país para la construcción del indicador sintético, que suman un total de

10.165 observaciones. Tal y como está estructurada la base de datos en cuestión<sup>2</sup>, el consumo se divide en tres grandes categorías:

-- Consumo general, que recoge datos del consumo, en diferentes ámbitos generales. Se compone de las variables que figuran en la Tabla 4.

VARIABLES QUE COMPONEN CONSUMO GENERAL	
KPIs	SIGNIFICADO
BBVA_ATM	Gasto en cajeros ATM
BBVA_Entertainment	Gasto en entretenimiento
BBVA_Food	Gasto en alimentación
BBVA_Health	Gasto en salud
BBVA_Hotels	Gasto en hoteles
BBVA_Online	Compras <i>online</i>
BBVA_Physical	Gasto en compras físicas
BBVA_Restaurants	Gasto en restaurantes
BBVA_Total.Consumption	Gasto en consumo total
BBVA_Transport	Gasto en transporte
ICC	Intención de gasto del consumidor
ICM	Evolución de las ventas y el empleo en el comercio minorista
Consumo eléctrico	Consumo de la electricidad

**Tabla 4:** Explicación de las variables que componen la categoría consumo general<sup>3</sup>.

*Fuente: Elaboración propia.*

-- Consumo sectorial, que representa el consumo, estudiado por sectores, y lo construyen las variables mostradas en la Tabla 5.

<sup>2</sup> Datos total España.

<sup>3</sup> “ICC” se refiere a “Índice de Confianza del Consumidor”; “ICM” se refiere a “Índice de Comercio al por Menor”.

VARIABLES QUE COMPONEN CONSUMO SECTORIAL			
KPIs	SIGNIFICADO	KPIs	SIGNIFICADO
AlimentaciónHipermercados	Gasto en hipermercados	OtrosOtros	Otros gastos
AlimentaciónResto	Gasto en alimentación	OtrosOtrosComercios	Gasto en otros comercios
BellezaCalzado	Gasto en calzado	OtrosParking	Gasto en parking
BellezaComplementos	Gasto en complementos	OtrosTributos	Pago de impuestos
BellezaRopa	Gasto en ropa	RestauraciónBar	Gasto en bares
HogarBricolajeymenaje	Gasto en bricolaje y menaje	RestauraciónCafeterías	Gasto en cafeterías
HogarHogar	Gasto para el hogar	RestauraciónRestaurantes	Gasto en restaurantes
HogarMascotas	Gasto para mascotas	SaludSalud	Gasto en salud
HogarMuebles	Gasto en muebles	SaludBelleza	Gasto en belleza
HogarReparaciones	Gasto en reparaciones	SaludDeporte	Gasto en deporte
HogarSeguros	Gasto en seguros	TransporteAvion	Gasto en avión
HogarTelefonía	Gasto en telefonía	TransporteCoche	Gasto en coche
OcioDeportes	Gasto en deportes	TransporteMetro	Gasto en metro
OcioElectrónica	Gasto en electrónica	TransporteOtros	Gastos en otros medios de transporte
OcioEstudios	Gasto en estudios	TransporteTaxi	Gasto en taxi
OcioFotografía	Gasto en fotografía	TransporteTren	Gasto en tren
OcioJuguetes	Gasto en juguetes	TurismoAgenciasdeViajes	Gasto en agencias de viajes
OcioLoterías	Gasto en loterías	TurismoHoteles	Gasto en hoteles
OcioOcio	Gasto en ocio	TurismoViajes	Gasto en viajes
OtrosAutomovil	Gasto en automóvil	M. V.	Matriculación de vehículos: compra de vehículos, en general
OtrosBazares	Gasto en bazares	M. V. para empresas	Matriculación de vehículos: compra de vehículos, para empresas
OtrosDonaciones	Donaciones	M. V. para particulares	Matriculación de vehículos: compra de vehículos, para particulares
OtrosGasolinerías	Gasto en gasolinerías	Número de pernотaciones hoteleras	
Gasto de turistas internacionales en alojamiento en España			

**Tabla 5:** Explicación de las variables que componen la categoría consumo sectorial.

*Fuente: Elaboración propia.*

-- *Consumer tracker*, que, tal y como su nombre indica, hace un seguimiento del consumidor, en tanto que engloba datos relativos a la intención de consumo del mismo en distintas áreas (como puede ser la alimentación, la restauración o el entretenimiento) y a las distintas preocupaciones que afectan a su nivel de consumo (como la preocupación económica o por la propia salud<sup>4</sup>). En concreto, esta categoría está compuesta por las variables que se puede ver en la Tabla 6.

<sup>4</sup> Especialmente relevante en un contexto pandémico como el vivido en el año 2020.

VARIABLES QUE COMPONEN CONSUMER TRACKER	
KPIs	SIGNIFICADO
INC alimentación	Intención neta de consumo de alimentación
INC electrónica	Intención neta de consumo de electrónica
INC entretenimiento	Intención neta de consumo de entretenimiento
INC Internet/Móvil	Intención neta de consumo de internet en el móvil
INC medicamentos	Intención neta de consumo de medicamentos
INC muebles	Intención neta de consumo de muebles
INC restauración	Intención neta de consumo de restauración
INC textil	Intención neta de consumo de textil
INC TV	Intención neta de consumo de televisión
INC viajes	Intención neta de consumo o gasto en viajes
INC gasolina	Intención neta de consumo de gasolina
INC libros	Intención neta de consumo de libros
Preoc. económica	Media del % de personas preocupadas por no realizar pagos y % de personas que retrasan compras de gran volumen
Preoc. salud	Media del % de personas preocupadas por su propia salud y % de personas preocupadas por la salud ajena
Preoc. pagos futuros	Preocupación por no poder afrontar próximos pagos
Retraso compras	Retraso en compras de gran volumen

**Tabla 6:** Explicación de las variables que componen la categoría *consumer tracker*.  
*Fuente:* Elaboración propia.

Estas últimas tres tablas presentan las variables o KPIs que componen cada una de las categorías, sumando un total de 77 para la base de datos total. En cuanto a su extensión temporal, estas variables se distribuyen de la siguiente manera: las pertenecientes a consumo general contienen datos diarios (salvo dos KPIs, que están en frecuencia mensual) entre enero del año 2020 y mayo del 2022. Consumo sectorial dispone de datos desde junio del 2019 hasta el mismo mes del año siguiente, en frecuencia mensual. Por último, las variables de *consumer tracker* cuentan con datos desde abril del 2020 hasta agosto del 2021, en frecuencia generalmente mensual, excepto entre los meses de abril y agosto (ambos incluidos) del año 2020, que se encuentran en frecuencia bisemanal. Véase la disponibilidad temporal de las variables en la Tabla 7, de una manera más visual.

RESUMEN DE LOS DATOS DISPONIBLES PARA CADA CATEGORÍA						
CATEGORÍA	Nº OBS.	FREC.	2019	2020	2021	2022
Consumo general	9.201	Diaria		Ene-Dic.	Ene.-Dic.	Ene.-May.
Consumo sectorial	636	Mensual	Jun-Dic.	Ene.-Jun.		
<i>Consumer tracker</i>	328	Mensual		Abr.-Dic.	Ene.-Ago.	

**Tabla 7:** Tabla-resumen de la disponibilidad temporal de los datos.  
*Fuente:* Elaboración propia.

Como se ha venido viendo hasta ahora, cada categoría cuenta con un importante número de KPIs. También se ha visto que la disponibilidad temporal de los datos no es

homogénea, por lo que se debe tratar de seleccionar aquellas variables y aquel rango temporal que presenten datos suficientes de los que poder extraer análisis y conclusiones objetivos y veraces.

Es evidente que la ausencia de datos dota de imprecisión a un conjunto de datos. No es lo mismo querer hallar los datos de un año en frecuencia mensual teniendo disponibles únicamente los datos correspondientes a los meses de enero y febrero, en cuyo caso, habría que imputar los valores de los diez meses restantes, que, con el mismo objetivo, tener disponibilidad de los datos correspondientes a los meses de enero a octubre, y tener que imputar, mediante el método que se considere oportuno, los valores correspondientes a los meses de noviembre y diciembre. Este último conjunto de datos es, cuanto menos, mucho más fidedigno que el primero. Por ello, se va a tratar de escoger aquel rango temporal que presente la mayor cantidad de valores posibles, de manera que se procure que la base de datos no presente un porcentaje de valores faltantes mayor al 30%. Siguiendo este objetivo, se descartan los años 2019 y 2022. Para el año 2021 únicamente se dispone de los datos correspondientes a una categoría “y media”: consumo general cuenta con datos para el año entero, consumo sectorial no dispone de datos para tal año, y *consumer tracker* tiene únicamente datos de enero a agosto del año en cuestión. El indicador sintético se construirá sobre las observaciones correspondientes al año 2020 del total España, en frecuencia mensual, ya que es el año del que más datos se dispone: consumo general cuenta con datos de enero a diciembre, consumo sectorial de enero a junio, y *consumer tracker* de abril a diciembre.

Una vez establecido el año objeto de análisis, procede un estudio pormenorizado del comportamiento de cada una de las variables, para determinar si la variable en cuestión dispone de todos los datos correspondientes, y en caso de que no sea así, indicar si procede su imputación, o eliminación (en este último caso, si no se cuenta con un mínimo aproximado del 70% de los datos, tal y como se comentó anteriormente):

-- Las variables pertenecientes a la categoría consumo general cuentan con todos los datos correspondientes al año 2020, salvo el KPI consumo eléctrico, que tiene observaciones hasta agosto. Por esta razón, se reemplazan los datos de esta variable por los datos de consumo eléctrico proporcionados por el Ministerio de Asuntos Económicos

y Transformación Digital<sup>5</sup>, que sí que cuentan con los datos correspondientes a todos los meses del año 2020.

-- La mayoría de los KPIs de consumo sectorial presentan datos de junio 2019 a junio 2020. Tomando únicamente el año 2020 como punto de referencia, faltan datos correspondientes a los últimos seis meses, que representan el 50% sobre el total de doce meses. Es decir, se dispone del 50% restante de los datos para la categoría en cuestión. Para poder imputar los KPIs de consumo sectorial y no tener que eliminarlos, se ha buscado una solución para tratar de disponer de un mayor porcentaje de datos: en lugar de hallar los porcentajes sobre el total de 12 meses que componen el año 2020, se hará sobre el total de 19 meses que van desde junio del 2019 hasta diciembre del 2020. De esta manera, en lugar de tener un 42% de datos en la mayoría de los KPIs que componen esta categoría, tenemos un total de 13 datos sobre el total de 19, esto es 68% de datos disponibles, y un 32% de datos faltantes. Si bien estos porcentajes se encuentran en el umbral mínimo de datos disponibles para poder, posteriormente, proceder a la imputación de los faltantes, se consigue un porcentaje de datos mucho mayor al que se tiene si se tiene como base los doce meses del año 2020. No obstante, hay que mencionar separadamente el caso de cinco variables. Dos de ellas, a saber, gasto de turistas internacionales en alojamiento en España, y número de pernoctaciones hoteleras, contienen datos únicamente entre octubre 2019 y mayo 2020. Si se coge el total de 15 meses (de octubre 2019 a diciembre 2020), se cuenta con alrededor del 53% de los datos. Si se coge el total de 12 meses (de enero a diciembre 2020), se tiene todavía menos datos disponibles, alrededor de un 42%. Se ve que, en ningún caso, se llega a los porcentajes mínimos de disponibilidad de datos para poder proceder a su imputación, por lo que se concluye que estas dos variables deben ser eliminadas. Distinto es el caso de las otras tres variables a las que hay que hacer mención: matriculación de vehículos, matriculación de vehículos para empresas, y matriculación de vehículos para particulares, variables que cuentan con los datos correspondientes a todos los meses, desde enero hasta diciembre, del año 2020, por lo que no corresponde realizar ningún tipo de imputación, ya que no hay información ausente.

-- En general, todos los KPIs de *consumer tracker* presentan información ausente entre los meses de enero a marzo, y tienen los datos correspondientes a los meses de abril a

---

<sup>5</sup> Véase, en el siguiente enlace: <https://serviciosede.mineco.gob.es/Indeco/sie/capituloSIEInf.aspx> Producción y Demanda > 17.- Consumo de energía eléctrica

diciembre 2020. Así, se tiene, en la mayoría de las variables un total del 75% de datos disponibles y un 25% de datos faltantes, por lo que se podría imputar los datos de acuerdo con la norma matemática ya mencionada. Sin embargo, hay dos KPIs que no presentan el mismo comportamiento, y son tanto la intención neta de consumo de gasolina como la intención neta de consumo de libros. Para estas dos variables, se tienen datos desde abril hasta agosto del 2020, esto es, se tienen datos de 5 meses de los 12 que componen al año, lo cual significa que se cuenta, como en el caso anterior, con un 42% de los datos, y el 58% restante son datos faltantes. Falta más de la mitad de los datos, por lo que estas dos variables se deben eliminar, ya que si se procede a imputar los datos faltantes, los resultados podrían quedar poco veraces.

Tras haber eliminado las variables que no cuentan con información suficiente para ser imputadas y haber delimitado la franja temporal objeto de estudio, la base de datos sobre la que se va a construir el indicador sintético va de enero a diciembre 2020, en frecuencia mensual, sobre los datos relacionados con el consumo del total España. Tras esta primera limpieza de los datos, cada categoría queda de la manera que se ve en la siguiente tabla:

CONSUMO GENERAL		CONSUMER TRACKER	
KPIs		KPIs	
BBVA_ATM	BBVA_Restaurants	INC alimentación	INC textil
BBVA_Entertainment	BBVA_Total.Consumption	INC electrónica	INC TV
BBVA_Food	BBVA_Transport	INC entretenimiento	INC viajes
BBVA_Health	ICC	INC Internetmóvil	Preoc. No Pagos
BBVA_Hotels	ICM	INC medicamentos	Preoc. Salud
BBVA_Online	Consumo eléctrico	INC muebles	Preoc. Próx. Pagos
BBVA_Physical		INC restauración	Retraso Compras

CONSUMO SECTORIAL			
KPIs			
AlimentaciónHipermercados	OcioDeportes	OtrosOtrosComercios	TransporteOtros
AlimentaciónResto	OcioElectrónica	OtrosParking	TransporteTaxi
BellezaCalzado	OcioEstudios	OtrosTributos	TransporteTren
BellezaComplementos	OcioFotografía	RestauraciónBar	TurismoAgenciasdeViajes
BellezaRopa	OcioJuguetes	RestauraciónCafeterías	TurismoHoteles
HogarBricolajeymenaje	OcioLoterías	RestauraciónRestaurantes	TurismoViajes
HogarHogar	OcioOcio	SaludSalud	M. V.
HogarMascotas	OtrosAutomovil	SaludBelleza	M. V. para empresas
HogarMuebles	OtrosBazares	SaludDeporte	M. V. para particulares
HogarReparaciones	OtrosDonaciones	TransporteAvion	
HogarSeguros	OtrosGasolineras	TransporteCoche	
HogarTelefonia	OtrosOtros	TransporteMetro	

**Tabla 8:** Variables finales que componen cada categoría.

*Fuente: Elaboración propia.*



## 2.2. Imputación de la información ausente

Una vez se tiene el conjunto de datos con el que se quiere trabajar, es determinante precisar cómo se va a tratar los valores perdidos. Según la OECD (2008), existen tres tipos de valores perdidos: *missing completely at random* (MCAR), *missing at random* (MAR) y *not missing at random* (NMAR). El primero es un supuesto en el que los datos faltantes lo hacen de una manera completamente aleatoria, y por tanto, es difícilmente identificable la razón de su ausencia: por ejemplo, se tiene una base de datos en frecuencia diaria para un año concreto, y faltan los datos correspondientes al 8 de enero y al 27 de noviembre. El segundo supuesto, MAR, se da cuando «la distribución de los valores observados no depende del patrón de comportamiento de los registros sin información [...]» (Medina y Galván, 2007, p.19). Por último, los NMAR, como su propio nombre indica, son casos en los que existe una razón subyacente que motiva la ausencia del dato en cuestión. Un claro ejemplo de este último tipo de valores perdidos es el conjunto de datos objeto de análisis en el presente trabajo, que, en dos de sus categorías, faltan datos en los primeros o últimos meses del año en cuestión.

Para la categoría *consumer tracker*, faltan los valores correspondientes a los meses de enero, febrero y marzo del 2020, para todas las variables. Es posible que su razón sea que se trate de variables que se empezaron a medirse en mayor medida tras el Covid-19, ya que recoge, en general, la intención neta de consumo en distintas áreas (comportamiento que, como ya se vio *supra*, se ha visto muy transformado y ha estado en un principal foco de atención tras la pandemia), y la preocupación de los consumidores tanto a nivel económico como salud. En cuanto a la categoría de consumo sectorial, faltan datos de julio a diciembre del año objeto de estudio, probablemente porque el gran impacto de pandemia en el consumo sectorial desvirtuó en gran manera tales datos. En cuanto a la tercera categoría, consumo general, ya se ha indicado que se cuenta con los valores correspondientes a todos los meses de 2020, por lo que no existe ningún valor perdido, ni por azar, ni siguiendo un patrón.

Ninguna de las categorías presenta datos perdidos al azar (ni completamente al azar), pues entre las fechas inicial y final de disponibilidad de datos, no se da ningún caso aislado de ausencia de dato.

Parece lógico que una de las características que se pretende conseguir siempre en un análisis de datos es que la base de datos esté lo más completa posible, y cuanto mayor sea



el número de observaciones de que se dispone, mejor. Ello dotará a las conclusiones extraídas de mayor objetividad, precisión y veracidad. Sin embargo, la realidad es que normalmente es muy complicado conseguir un conjunto de datos que cumpla con estas características, ya que no siempre el número de observaciones o variables es suficiente para ser representativo de la realidad -como ocurría con el caso de Portugal, por ejemplo-, y muchas veces nos encontramos, como hemos venido comentando hasta el momento, con valores faltantes. Así, si una base de datos contiene valores ausentes, éstos deben ser tratados de una de las siguientes maneras (Medina y Galván, 2007; OECD, 2008; Muñoz y Álvarez, 2009):

- Eliminar la observación. Tal y como indica su propio título, se trata de eliminar por completo aquellas observaciones de las que no se dispone. Este método tiene dos problemas fundamentales. El primero, que ignora precisamente el porqué de esa ausencia en el valor, aspecto relevante para los MAR y NMAR. El segundo, que si se eliminan todos los casos con valores perdidos, la base de datos se reducirá considerablemente, y consecuentemente, el error estándar del producto final aumentará. Como ya se ha comentado, conclusiones basadas en un mayor número de observaciones son más precisas, objetivas, y representan mejor la realidad. Por ello, si más del 5% de los valores de una variable son valores perdidos, normalmente no se aplica este método. En el presente trabajo se ha concluido la eliminación de ciertas variables, ya justificado en el apartado de selección de los datos.

- Ignorar el valor perdido. Como su nombre indica, esta técnica consta en dejar los valores perdidos en la base de datos, y simplemente ignorarlos, de tal manera que únicamente se tenga en cuenta los datos de cuyo valor se dispone. Esta solución no se va a aplicar en el presente trabajo, ya que la falta de datos en tantas variables tanto en los primeros como en los últimos meses del año no puede ser ignorada, en tanto que el indicador final se podría ver altamente desvirtuado, representado de manera fehaciente únicamente los datos entre los meses de abril y junio 2020.

- Imputación simple. Mediante la imputación simple se pretende dar un único valor estimado al ausente. Ello se puede hacer realizar con distintos métodos, que la OECD los clasifica en implícitos y explícitos (2008). Aquéllos presumen ciertas características de la base de datos de manera implícita (OECD, 2008), y dos de los más conocidos son los siguientes:

-- *Hot deck imputation*. Lo que se hace con este método es “duplicar” uno de los valores ya existentes en la base de datos, para dárselo al faltante cuyas características en el resto de las variables sean muy similares (Otero, 2011). Por ejemplo, a una persona que en una encuesta no haya indicado su renta anual, se le daría la renta de aquella persona que hubiera respondido igual en el resto de variables (como puede ser lugar de residencia, nivel de estudios, años que lleva en la compañía, etc.).

-- *Cold deck imputation*. De manera similar a como ocurre con el método anterior, se le da al valor faltante un valor de otra observación similar, pero obtenido de una base de datos distinta a la que es objeto de imputación (Otero, 2011).

Estos métodos asumen dos cosas: por un lado, que la ausencia de datos se da de manera completamente aleatoria (MCAR). Por otro lado, que los valores faltantes siguen la misma distribución que los no faltantes (Primorac *et. al.*, 2020). Tal y como indica Otero (2011), un riesgo del método *hot deck* es que un mismo valor se duplique muchas veces, lo cual ocurre cuando «hay muchos valores faltantes y pocos valores registrados» (p.21). Teniendo en cuenta todo ello, en principio se descarta cualquiera de estos métodos de imputación para nuestra base de datos. Primero, porque tal conjunto de datos no presenta valores faltantes aleatoriamente. Segundo, porque teniendo el año 2020 como objeto de análisis (momento en que se vivió una pandemia que creó una crisis económica y sanitaria a nivel mundial), si se busca la asignación de los datos faltantes con el valor de las variables en cuestión para otros años (como ocurre con la imputación *cold deck*), los datos finales podrían quedar muy sesgados y desvinculados de la realidad del momento. Tercero (para *hot deck*), porque con la cantidad de datos faltantes que presenta la base de datos, se considera que un método más sofisticado, que tenga en cuenta la evolución de una misma variable, o incluso su relación con otras, dotaría a la misma de mayor precisión.

Al segundo grupo de métodos de imputación simple, la OECD (2008) los clasifica como “explícitos”, pues se trata de modelos estadísticos que las asunciones se muestran de una forma explícita, y algunos de los más conocidos son:

-- *Imputación por media, mediana o moda*, y se puede dar de manera incondicionada o condicionada. En el primer caso, el valor que se le dará a los valores perdidos es la media (o mediana o moda) de la variable a la que pertenezcan. En el segundo, se hace una primera agrupación de variables en función de sus correlaciones, y se imputa a los valores perdidos la media, mediana o moda del grupo al que pertenece. En ambos casos se

presume que los datos perdidos son MCAR, y es un método que no deja de presentar inconvenientes, en tanto que afecta a la distribución de la variable imputada, e incluso a su relación con otras variables. Así, en caso de gran ausencia de datos, los valores imputados podrían estar altamente sesgados (Medina y Galván, 2007; Otero, 2011).

-- *Regression imputation*, o regresión, en el que la variable dependiente es el indicador simple con valores perdidos, y los regresores son el resto de indicadores simples (OECD, 2008). En palabras de Otero (2011), «[e]ste procedimiento consiste en eliminar las observaciones con datos incompletos y ajustar la ecuación de la regresión para predecir los valores faltantes». Uno de los requisitos es que la base de datos debe estar muy correlada.

-- *Imputación por interpolación*, en el que los valores perdidos se hallan mediante una interpolación del último valor válido antes del valor perdido y el primero después del mismo (Sandoval, 2021).

-- *Imputación por máxima verosimilitud*. Probablemente el algoritmo más generalizado sea el *Expectation Maximization (EM) imputation*. Se trata de un algoritmo iterativo que parte de una estimación inicial de los valores faltantes. Cada iteración consta de dos pasos: el E (*expectation*), que estima el valor de los datos faltantes en función de las estimaciones iniciales (para el caso de la primera iteración. Para el resto, la referencia son las estimaciones obtenidas en la iteración inmediatamente anterior) y de los valores observados. En el segundo paso, M (*maximization*), se maximiza la función obtenida en E<sup>6</sup> (OECD, 2008; Otero, 2011). De esta manera, se van actualizando y optimizando los valores perdidos tantas veces como iteraciones se lleven a cabo en el proceso.

Todo lo presentado hasta el momento son métodos de imputación simple, que dan un único valor a los faltantes. No obstante, hay técnicas de imputación múltiple que, a partir de la combinación de imputaciones y estimaciones independientes, dan más de una estimación de valor a los ausentes, estimaciones con buenas propiedades estadísticas y cuya variabilidad es fácilmente hallable (Chen & Shao, 1999; Otero, 2011). La ventaja que tiene este tipo de imputación de valores ausentes es que disminuye considerablemente el error estándar, puesto que se tienen en cuenta distintos posibles valores para la variable,

---

<sup>6</sup> Para una descripción más detallada del proceso que se sigue en esta técnica, véase Otero (2011), pp. 25-27.

y no únicamente uno. Los métodos más utilizados actualmente son la aproximación bayesiana (*bootstrap*), y el método Monte Carlo con cadenas de Markov (OECD, 2008).

Si bien los métodos de imputación múltiple son los más robustos y precisos, no necesariamente se debe acudir a ellos, en tanto que, en muchas ocasiones, los resultados obtenidos mediante imputación simple son igualmente robustos, y mucho más fáciles de llevar a cabo. Numerosos autores se han pronunciado al respecto. Entre ellos, Fay (1991), Rao y Shao (1992), y Rao (1996) (citados por Muñoz y Álvarez (2009)) han puesto de manifiesto ciertas limitaciones de la imputación múltiple, que no siempre la hacen idónea para el caso en cuestión.

Teniendo en cuenta lo comentado en el párrafo anterior, proceden las siguientes determinaciones:

- Se descarta la imputación múltiple por las razones comentadas *supra*, siendo éstas, fundamentalmente, la gran complejidad operacional que ella conlleva, y la posibilidad de imputar valores muy poco sesgados por otros métodos mucho más accesibles.

- Se descarta la imputación simple implícita por las razones ya señaladas, destacando la existencia de métodos mucho más completos con capacidad de dotar a la imputación de mayor precisión. Por estas mismas razones, se descarta la imputación explícita por media, mediana o moda.

- Cuando proceda la imputación de datos, ésta se llevará a cabo mediante uno de los siguientes métodos: regresión, interpolación, o máxima verosimilitud (con algoritmo EM).

- Algunos indicadores simples han procedido a ser eliminados por el bajo número de número de datos de que disponían (véase el razonamiento final del apartado anterior).

- En ningún momento se dejará e ignorará valores perdidos en el presente conjunto de datos. Todos ellos han sido tratados mediante la eliminación o lo serán mediante la imputación.

La imputación de los datos se puede hacer por indicador simple, por categorías, o por el total de los datos disponibles. Como la primera opción puede ignorar las relaciones que guardan ciertos indicadores entre sí -lo cual puede ser interesante para una imputación más completa-, y la última parece excesiva en tanto que siempre, en una primera instancia,

se va a tomar cada categoría por separado, se decide que, en el presente trabajo, la imputación se llevará a cabo por cada una de las categorías que componen el consumo (consumo sectorial y *consumer tracker*, ya que, recuérdese, que no se necesita imputar consumo general, ya que se dispone de todos los datos).

Ya se ha adelantado que se va a imputar los datos en cuestión mediante la regresión, interpolación, o máxima verosimilitud. La forma de determinar cuál es la más idónea, es llevar a cabo una partición, que consiste en crear una nueva base de datos que comprende únicamente los datos de los que se dispone, y eliminar una parte de ellos. Esos datos que se han eliminado, pero cuyo valor verdaderamente se conoce, se imputan mediante los distintos métodos posibles, y se compara el resultado de la imputación con el valor real. De todos los métodos utilizados, se escoge aquél que más se acerque al valor real.

Para considerar distintos escenarios dentro del desempeño de cada método, se ha realizado tres particiones distintas: 80-20, 75-25 y 70-30. Esto significa que para el primer caso se ha dejado un 80% de los datos, y se ha eliminado el 20% restante. Con las otras dos particiones se ha hecho lo mismo, pero con porcentajes distintos: se deja, respectivamente, un 75% y un 70%, y se elimina el 25% y 30% restante.

Como en ambas categorías a imputar -recuérdese que consumo general no precisa de imputación- hay datos faltantes del año 2020, y el total de meses que correspondería a este año no llega a 12, el número de observaciones total por variable es muy reducido, por lo que las imputaciones podrían resultar muy al azar. Por este motivo, se ha decidido incluir, en las bases de datos, los datos de que se dispone tanto del año anterior como posterior a 2020. Para el caso de *consumer tracker*, el conjunto de datos de que se dispone comprende entre abril 2020 y agosto 2021 (ambos incluidos, es decir, total de 17 meses). Para consumo sectorial se incluyen datos de junio 2019 a junio 2020 (ambos incluidos, lo que se corresponde con un total de 13 meses). Dentro de cada una de estas bases de datos -en las que sí contamos con todos los datos-, eliminamos un 20%, 25% y 30% de las observaciones en cada caso, y procedemos a imputarlas. En concreto, se han empleado los métodos de imputación simple que se exponen a continuación.

- La regresión se ha llevado a cabo con la herramienta IBM SPSS Statistics. En el diálogo de análisis de valores perdidos, esta herramienta permite llevar a cabo una regresión lineal múltiple, que tiene en cuenta no únicamente los datos observados de la

variable en cuestión, sino también del resto de variables disponibles en la base de datos (IBM, 2019).

- Se ha empleado tres técnicas diferentes de interpolación, todas ellas con la herramienta Matlab, y haciendo uso de la función *fillmissing* (MathWorks, 2022):

-- Interpolación lineal (función denominada *linear*), que, de forma lineal, imputa el valor de los datos faltantes con aquellos que guardan características más parecidas, teniendo en cuenta únicamente los valores observados de la variable en cuestión, sin tener en cuenta al resto de variables.

-- Interpolación cúbica con la función *spline* a trozos, similar al caso anterior, pero definido por un polinomio cúbico.

-- Interpolación cúbica con la función *pchip* a trozos, una variante del caso anterior que conserva la forma.

- El algoritmo *Expectation Maximization* que, siguiendo el método ya delimitado anteriormente, calcula las estimaciones de máxima verosimilitud (IBM, 2019).

Las imputaciones se comparan con el valor real que corresponde a cada observación, lo cual se hace hallando distintos índices de medición, que indican cómo de buena ha sido tal imputación. Las métricas utilizadas para determinar la conveniencia de cada método de imputación simple son las siguientes:

-  $R^2$  o coeficiente de correlación, hace una medición de cuánto se ajusta la imputación a los datos reales, por lo que, cuanto mayor sea esta métrica, mejor será la imputación. Siendo  $N$  el número de imputaciones,  $P_i$  el valor imputado,  $O_i$  el valor observado,  $\bar{P}$  la media de los valores imputados,  $\bar{O}$  la media de los valores observados, y  $\sigma_P$  la desviación típica de los valores imputados y  $\sigma_O$  de los valores observados,  $R^2$  se expresa de la siguiente manera (3):

$$R^2 = \left[ \frac{1}{N} \frac{\sum_{i=1}^N [(P_i - \bar{P})(O_i - \bar{O})]}{\sigma_P \sigma_O} \right]^2 \quad (3)$$

- d o índice de concordancia de Willmott, que determina, en un rango del 0 al 1, cómo de correcta ha sido la estimación. Si el índice presenta un valor de 1, significa que la estimación ha sido perfecta, por lo que se busca un índice de Willmott lo más cerca de 1 posible. Con los mismos términos que en la ecuación (3), y con  $k$  pudiendo tomar los valores 1 o 3, este índice se halla de la forma expresada en la siguiente ecuación (4):

$$d = 1 - \left[ \frac{\sum_{i=1}^N (P_i - O_i)^k}{\sum_{i=1}^N [ (|P_i - \bar{O}| + |O_i - \bar{O}|)^k ]} \right]^2 \quad (4)$$

- La raíz del error medio cuadrado o *root mean squared error (RMSE)* representa el error medio de la imputación, por lo que, como lo que se mide es el error y no lo bien hecha que está la imputación, se busca la medida más pequeña. Véase la ecuación (5).

$$RMSE = \left( \frac{1}{N} \sum_{i=1}^N (P_i - O_i)^2 \right)^{1/2} \quad (5)$$

- Error absoluto medio o *mean absolute error (MAE)*, que representa, como bien indica su nombre, el error medio de la imputación, en valor absoluto. Al igual que en el caso anterior, es una métrica que mide el error de imputación, por lo que se busca el menor error posible. Véase la ecuación (6).

$$MAE = \frac{1}{N} \sum_{i=1}^N |P_i - O_i| \quad (6)$$

Con la herramienta Excel, se ha hallado el resultado de las ecuaciones (3) a (6) para cada valor imputado. Dentro de cada categoría, se ha procedido a sumar los resultados, obteniendo los presentados en las siguientes tablas:

PARTICIÓN 70-30 EN CONSUMER TRACKER					
MÉT. IMP.	R <sup>2</sup>	d CON k=1	d CON k=2	RMSE	MAE
Lineal	0,76	0,96	0,99	0,18	0,13
Spline	0,24	0,30	0,01	7,33	4,45
Pchip	0,15	0,50	0,06	2,48	1,38
E.M.	0,907	0,98	0,998	0,16	0,10
Regresión	<b>0,909</b>	<b>0,99</b>	<b>0,999</b>	<b>0,10</b>	<b>0,06</b>

**Tabla 9.1:** Resultados de los distintos métodos de imputación para *consumer tracker* en partición 70-30.  
**Fuente:** Elaboración propia.

PARTICIÓN 75-25 EN CONSUMER TRACKER					
MÉT. IMP.	R <sup>2</sup>	d CON k=1	d CON k=2	RMSE	MAE
Lineal	0,87	<b>0,999</b>	0,9993	<b>0,099</b>	0,072
Spline	0,35	0,45	0,27	4,04	2,59
Pchip	0,38	0,88	0,76	0,84	0,43
E.M.	0,910	0,985	0,998	0,15	0,09
Regresión	<b>0,914</b>	0,990	<b>0,9994</b>	0,104	<b>0,071</b>

**Tabla 9.2:** Resultados de los distintos métodos de imputación para *consumer tracker* en partición 75-25.

*Fuente:* Elaboración propia.

PARTICIÓN 80-20 EN CONSUMER TRACKER					
MÉT. IMP.	R <sup>2</sup>	d CON k=1	d CON k=2	RMSE	MAE
Lineal	0,88	<b>0,994</b>	<b>0,99924</b>	<b>0,10</b>	0,08
Spline	0,18	0,54	0,27	3,58	2,02
Pchip	0,36	0,89	0,80	0,75	0,35
E.M.	<b>0,90</b>	0,989	0,9989	0,13	0,08
Regresión	0,88	0,992	0,99917	0,12	<b>0,07</b>

**Tabla 9.3:** Resultados de los distintos métodos de imputación para *consumer tracker* en partición 80-20.

*Fuente:* Elaboración propia.

Se resaltan los mejores resultados para cada una de las técnicas de imputación, en cada una de las particiones, y para cada métrica de evaluación de acierto de la imputación. Claramente, se puede ver que quedan descartados los métodos de imputación *spline*, *pchip* y *expectation maximization*, y que las técnicas que mejor resultado proporcionan son, bien la interpolación lineal, bien la regresión lineal múltiple, ya que, si bien en la partición 70-30 parece claro que este último método proporciona unos valores más precisos, no todas las particiones presentan el mismo patrón. Así, se ha hecho un cuadro porcentual de las diferencias, en valor absoluto, de los resultados obtenidos en cada caso, con el fin de determinar, con un fundamento más razonado y probado, qué tipo de imputación se debe seguir en el caso de esta categoría. El cuadro de diferencia de valores en puntos porcentuales se representa a continuación, en la Tabla 10. Para que se vea de manera más visual, se han eliminado los porcentajes correspondientes a los otros tres métodos de imputación.

DIFERENCIA PORCENTUAL DE RESULTADOS DE LAS IMPUTACIONES EN CONSUMER TRACKER								
MÉT. IMP.	PARTICIÓN	R <sup>2</sup>	d CON k=1	d CON k=2	RMSE	MAE	SUMA	TOTAL
Lineal	70-30	16%	4%	1%	80%	101%	<b>202%</b>	<b>221%</b>
	75-25	5%	0%	0%	0%	2%	<b>7%</b>	
	80-20	2%	0%	0%	0%	10%	<b>12%</b>	
Regresión	70-30	0%	0%	0%	0%	0%	<b>0%</b>	<b>23%</b>
	75-25	0%	1%	0%	5%	0%	<b>6%</b>	
	80-20	2%	0%	0%	14%	0%	<b>17%</b>	

**Tabla 10:** Diferencias porcentuales del método de imputación frente al mejor en *consumer tracker*.

*Fuente:* Elaboración propia.



El cálculo seguido para llegar a estos porcentajes el siguiente: se ha tenido como referencia la mejor técnica de imputación para uno de los índices de error, aquí presentados como columnas. A continuación, se ha establecido la tasa de variación de cada uno de los resultados sobre esa referencia. Para cada técnica de imputación  $t$  y métrica de error  $m$ , el dato  $F_{tm}$  obtenido en las figuras 9.1 a 9.3 se compara con la imputación de referencia  $Z_t$  de manera que nos da la diferencia porcentual  $D_{mt}$  del valor estudiado frente al de referencia. Véase la ecuación (7).

$$\left| \frac{F_{tm} - Z_t}{Z_t} * 100 \right| = D_{mt} \quad (7)$$

Si bien para la métrica  $R^2$  en partición 80-20 el mejor valor lo presenta la *expectation maximization*, el resto de casos tienen sus mejores imputaciones en los métodos lineal y de regresión, razón por la que en la Tabla 10 se han incluido únicamente estas dos filas.

Para dar una correcta interpretación a los resultados de la Tabla 10, se debe tener en cuenta que un menor valor significa mayor semejanza con el valor de la mejor imputación en cuestión. Por ello, se busca la menor suma posible. La imputación lineal suma 221%, mientras que la regresión tiene un total de 23%, por lo que parece claro que el mejor método para *consumer tracker* es la regresión con SPSS.

Los mismos análisis y lógica se han llevado a cabo para la categoría consumo sectorial, cuyas tablas se muestran en las Tablas 11 y 12.

PARTICIÓN 70-30 EN CONSUMO SECTORIAL					
MÉT. IMP.	$R^2$	d CON k=1	d CON k=2	RMSE	MAE
Lineal	0,80	0,9970	0,9936	15,16	6,06
Spline	0,13	0,9994	0,47	151,48	63,65
Pchip	0,31	0,99	0,91	36,27	11,25
E.M.	<b>0,84</b>	<b>1,0000</b>	<b>0,9939</b>	<b>14,59</b>	<b>5,58</b>
Regresión	0,83	0,9997	0,991	15,47	5,76

**Tabla 11.1:** Resultados de los distintos métodos de imputación para consumo sectorial en partición 70-30.

*Fuente:* Elaboración propia.

PARTICIÓN 75-25 EN CONSUMO SECTORIAL					
MÉT. IMP.	R <sup>2</sup>	d CON k=1	d CON k=2	RMSE	MAE
Lineal	0,75	<b>1,0000000</b>	0,99	27,32	8,86
Spline	0,41	0,99999997	0,69	128,86	44,38
Pchip	0,36	0,99	0,96	30,80	9,22
E.M.	<b>0,87</b>	<b>1,0000000</b>	<b>1,00</b>	<b>11,97</b>	<b>4,39</b>
Regresión	0,81	0,99986	0,99	15,21	5,58

**Tabla 11.2:** Resultados de los distintos métodos de imputación para consumo sectorial en partición 75-25.

*Fuente: Elaboración propia.*

PARTICIÓN 80-20 EN CONSUMO SECTORIAL					
MÉT. IMP.	R <sup>2</sup>	d CON k=1	d CON k=2	RMSE	MAE
Lineal	0,73	<b>1,0000</b>	0,98	31,04	10,97
Spline	0,53	0,96	0,74	120,15	30,00
Pchip	0,34	0,98	0,95	35,54	12,84
E.M.	0,847	0,9997	0,997	13,19	4,43
Regresión	<b>0,850</b>	0,9999	<b>0,998</b>	<b>12,87</b>	<b>4,03</b>

**Tabla 11.3:** Resultados de los distintos métodos de imputación para consumo sectorial en partición 80-20.

*Fuente: Elaboración propia.*

DIFERENCIA PORCENTUAL DE RESULTADOS DE LAS IMPUTACIONES EN CONSUMO SECTORIAL								
MÉT. IMP.	PARTICIÓN	R <sup>2</sup>	d CON k=1	d CON k=2	RMSE	MAE	SUMA	TOTAL
E.M.	70-30	0%	0%	0%	0%	0%	<b>0%</b>	<b>13%</b>
	75-25	0%	0%	0%	0%	0%	<b>0%</b>	
	80-20	0%	0%	0%	2%	10%	<b>13%</b>	
Regresión	70-30	1%	0%	0%	6%	3%	<b>10%</b>	<b>72%</b>
	75-25	7%	0%	0%	27%	27%	<b>61%</b>	
	80-20	0%	0%	0%	0%	0%	<b>0%</b>	

**Tabla 12:** Diferencias porcentuales del método de imputación frente al mejor en consumo sectorial.

*Fuente: Elaboración propia.*

Visualizando los resultados de las Tablas 11, y teniendo en cuenta que, en la 12, EM suma 13% mientras que la regresión suma 72%, parece claro que el mejor método para consumo sectorial es la *expectation maximization*.

Una vez definido el método de imputación apropiado para cada una de las categorías, se procede a realizar la verdadera imputación de los datos:

- *Consumer tracker* se imputa con el método de regresión en la herramienta SPSS, y con el objetivo de conseguir unos datos finales del año 2020 en frecuencia mensual.

- Consumo sectorial se imputa con el método de *expectation maximization* en la herramienta SPSS, para tener una base de datos final correspondiente a cada uno de los meses del año 2020.

### 2.3. Análisis Multivariante.

Teniendo las bases de datos con las que se quiere trabajar, la OECD (2008), establece que procede un análisis multivariante para determinar, fundamentalmente, las correlaciones que guardan las variables entre sí, así como las estructuras y patrones que las mismas siguen. Esto se realiza con un doble objetivo. Por un lado, conocer en mayor profundidad los datos con los que se va a trabajar. Por otro lado, evitar la selección de indicadores simples y determinación de pesos de una manera arbitraria. El Manual presenta, junto a otras técnicas más residuales, los siguientes tipos de análisis multivariante:

- Análisis de Componentes Principales, más conocido por sus siglas en inglés, *PCA* (*Principal Component Analysis*). Es un método de reducción de dimensiones que trata de explicar el mayor porcentaje de varianza posible de las variables. Las componentes principales son combinaciones lineales de las variables originales.

- Análisis Factorial o Factor Analysis (FA). Es un análisis muy similar al PCA, pero que difiere en varios puntos: por un lado, trata de explicar las covarianzas o correlaciones entre las variables originales. Por otro lado, las variables originales son combinaciones lineales de los factores. Tanto el *PCA* como el *FA* tienen como requisito que la base de datos original esté muy correlada.

- Coefficiente Alpha de Cronbach, o Cronbach Coefficient Alpha. Es un coeficiente que mide la correlación existente entre los indicadores individuales<sup>7</sup>.

- Análisis clúster, o cluster analysis. Se trata de la aplicación de distintos métodos de clúster jerárquico (por ejemplo, árbol) y no jerárquico (por ejemplo, *k-means*) para ver la estructura que siguen los datos.

Teniendo claras las técnicas de análisis multivariante, se toman las siguientes decisiones: siempre que sea posible (esto es, que el conjunto de datos esté muy correlado), se analizará su estructura con FA, ya que es de fácil aplicación y sus resultados proporcionan una visión sólida y global del patrón que sigue el conjunto de datos objeto de análisis. Para este caso concreto, se prefiere FA sobre PCA porque, en palabras de Arredondo (2019), aquél extrae factores que explican las variables (o indicadores base),

---

<sup>7</sup> Para más información, mírese OECD (2008), pp. 72-73.



MATRIZ DE CORRELACIONES PARA CONSUMER TRACKER														
	INC alimentación	INC electrónica	INC ent <sup>a</sup>	INC Internetmóvil	INC medicamentos	INC muebles	INC restauración	INC textil	INC TV	INC viajes	Preoc. No Pagos	Preoc. Salud	Preoc. Próx. Pagos	Retraso Compras
INC alimentación	1,00													
INC electrónica	0,24	1,00												
INC entretenimiento	0,64	0,51	1,00											
INC Internetmóvil	0,14	0,08	0,15	1,00										
INC medicamentos	0,36	-0,07	0,44	0,63	1,00									
INC muebles	-0,12	0,66	0,24	-0,48	-0,40	1,00								
INC restauración	-0,03	0,71	0,40	-0,12	-0,10	0,52	1,00							
INC textil	0,12	0,72	0,21	-0,13	-0,39	0,52	0,74	1,00						
INC TV	0,02	0,73	0,27	0,50	0,22	0,41	0,47	0,43	1,00					
INC viajes	-0,08	0,63	0,26	0,25	0,07	0,36	0,87	0,57	0,68	1,00				
Preoc. No Pagos	-0,05	-0,01	-0,19	0,50	0,14	-0,03	-0,13	-0,20	0,53	0,34	1,00			
Preoc. Salud	-0,03	-0,15	0,02	0,12	0,28	-0,43	-0,15	-0,48	-0,04	-0,15	-0,06	1,00		
Preoc. Próx. Pagos	-0,35	-0,14	-0,54	0,01	-0,06	-0,15	0,00	-0,18	0,22	0,21	0,47	0,45	1,00	
Retraso Compras	0,40	-0,38	0,43	0,21	0,43	-0,26	-0,46	-0,45	-0,13	-0,42	0,15	0,05	-0,33	1,00

**Tabla 15:** Matriz de correlaciones de los indicadores simples de la categoría *Consumer tracker*.

*Fuente:* Elaboración propia.

Viendo las matrices de correlación, la estructura de datos de las categorías consumo general y consumo sectorial (ambas con indicadores muy correlados) se estudiará mediante FA. En cambio, como *consumer tracker* resulta estar muy poco correlada, se llevará a cabo un análisis clúster. No obstante, procede una breve mención previa sobre la necesidad de eliminar ciertas variables por seguir un patrón prácticamente idéntico que otras. Resulta claro que *consumer tracker*, una categoría tan poco correlada, no precisa de eliminación de indicadores base. Para las categorías consumo general y sectorial, si bien presentan correlaciones altas, no se ha visto un patrón idéntico de ningún indicador entre sí ni en su relación con el resto de indicadores. Teniendo en cuenta, además, que uno de los requisitos para la aplicación del FA es que se trate de un conjunto de datos correlado, se concluye que no procede eliminación de ninguna variable.

Para la aplicación de FA se debe tomar una primera decisión importante, y es determinar el número de factores a extraer. Esto es, el número de “grupos” en los que se van a dividir las variables. Para tomar tal decisión, se puede seguir cualquiera de los siguientes criterios:

- Criterio de Kaiser. Según esta regla, se debe tomar únicamente aquellos factores con autovalores mayores a la unidad.

- Gráfico de sedimentación. Se plasman todos los autovalores en un gráfico. El comportamiento normal es que haya un primer descenso drástico en el valor de tales autovalores, y luego comiencen a descender de manera mucho más plana. Uno se debe quedar con el número de factores de aquel autovalor inmediatamente anterior a la estabilización del valor de los autovalores.

- Criterio de la varianza explicada. Indica que se debe quedar con un número de factores que expliquen un 80% o 90% de la varianza total.

- Criterio de Jolliffe. Muy similar al criterio Kaiser, pero descarta los autovalores menores a 0,7.

- Criterio lógico. Dejando de lado los criterios puramente matemáticos, este trata de encontrar coherencia en los resultados obtenidos, intentando entender el significado de las agrupaciones de variables obtenidas.

En el presente trabajo se va a determinar el número de factores a extraer analizando tres de estas reglas: el criterio Kaiser, el total de la varianza explicada por los factores, y el criterio lógico.

NÚMERO DE FACTORES A EXTRAER PARA CONSUMO GENERAL			
FACTOR	AUTOVALOR	% VAR. EXPLICADA	% VAR. ACUMULADA
1	10,2	73%	73%
2	2,1	15%	88%
3	0,8	6%	94%
4	0,3	2%	97%
5	0,3	2%	99%
...	...	...	...

**Tabla 16:** Análisis del número de factores a extraer para consumo general.

*Fuente: Elaboración propia.*

Analizando los resultados obtenidos en consumo general, según el criterio Kaiser el número de factores a extraer es dos, ya que el tercer factor ya presenta un autovalor inferior a la unidad. En cuanto a la varianza explicada, con dos factores ya se explica casi un 90% de la varianza. Si para el caso de los autovalores, en lugar de seguir el criterio Kaiser, se sigue el Jolliffe, el número de factores quedaría en tres. Por otro lado, si el porcentaje de varianza mínima que se busca explicar es un 90%, también nos tendríamos que quedar con tres factores. Ante esta dualidad de posibilidades, cobra sentido la aplicación del criterio lógico: en principio, la creación de un tercer factor no crea un tercer grupo notoriamente identificable, por lo que se decide la extracción de dos factores para el caso de consumo general.

NÚMERO DE FACTORES A EXTRAER PARA CONSUMO SECTORIAL			
FACTOR	AUTOVALOR	% VAR. EXPLICADA	% VAR. ACUMULADA
1	29,1	63%	63%
2	10,9	24%	87%
3	3,4	7%	94%
4	1,2	3%	97%
5	0,7	2%	99%
...	...	...	...

**Tabla 17:** Análisis del número de factores a extraer para consumo sectorial.

*Fuente: Elaboración propia.*

Para el caso del consumo sectorial, según el criterio Kaiser se debe extraer cuatro factores. Según el criterio Jolliffe, cinco. En función del porcentaje de varianza acumulada, con únicamente dos factores sería suficiente. Con tres ya se supera el umbral de explicación del 90% de la varianza, y añadiendo un cuarto factor, aumenta la varianza acumulada únicamente en tres puntos. Como este último cuarto factor, al igual que ocurre en consumo general, no extrae un cuarto grupo fácilmente identificable y separado del resto, se decide extraer tres factores para el consumo sectorial.

Habiendo tomado estas decisiones, se extraen los siguientes grupos, o subcategorías:

CONSUMO GENERAL	
FACTOR 1	FACTOR 2
BBVA_ATM	BBVA_Health
BBVA_Entº	BBVA_Hotels
BBVA_Food	ICC
BBVA_Online	
BBVA_Physical	
BBVA_Restaurants	
BBVA_Tot.Cons.	
BBVA_Transport	
ICPM	
Consumo eléctrico	

**Tabla 18:** Agrupación de indicadores simples en consumo general.

*Fuente: Elaboración propia.*

Al primer factor de la categoría consumo general se le podría llamar “consumo necesario”, pues, en función de las variables que lo componen, representa el consumo, o los gastos, en actividades o servicios fundamentales para una vida relativamente saludable -salud tanto física como mental- en el momento pandémico que se vivió en el año 2020, como es el entretenimiento, la comida, los servicios online, el gasto en restaurantes y consumo total, gasto en transporte, la evolución en el sector del comercio minorista, el gasto eléctrico, y la extracción de efectivo en los cajeros.

Al segundo factor de esta categoría se le llamará “positivos”, en tanto que está altamente relacionado con los casos que hubo de Covid-19 en aquel momento: numerosos hoteles quedaron destinados a albergar casos de Covid-19, lo que está directamente relacionado con el gasto en salud. Ello también influye en el índice de confianza de consumidor, medidor de la salud financiera y poder adquisitivo del ciudadano medio. Ya se adelantó en la introducción del presente trabajo que la pandemia del año 2020 no tuvo únicamente un impacto en salud física, sino también económica ya no solo entre los españoles, sino a nivel mundial.

CONSUMO SECTORIAL		
FACTOR 1	FACTOR 2	FACTOR 3
BellezaComplementos	RestauraciónRestaurantes	AlimentaciónHipermercados
HogarHogar	SaludSalud	AlimentaciónResto
HogarMascotas	SaludSaludDeporte	BellezaCalzado
OcioElectrónica	TransporteAvion	BellezaRopa
OcioEstudios	TransporteCoche	HogarBricolajeyMenaje
OcioOcio	TransporteMetro	HogarMuebles
OtrosDonaciones	TransporteOtros	HogarReparaciones
OtrosGasolineras	TransporteTaxi	HogarTelefonía
OtrosOtrosComercios	TurismoAgenciasdeViajes	OcioFotografía
OtrosParking	TurismoHoteles	OcioJuguetes
RestauraciónCafeterías		OtrosAutomovil
		OtrosBazares
		OtrosTributos
		RestauraciónBar
		SaludSaludBelleza
		TransporteTren

**Tabla 19:** Agrupación de indicadores simples en consumo sectorial.

*Fuente:* Elaboración propia.

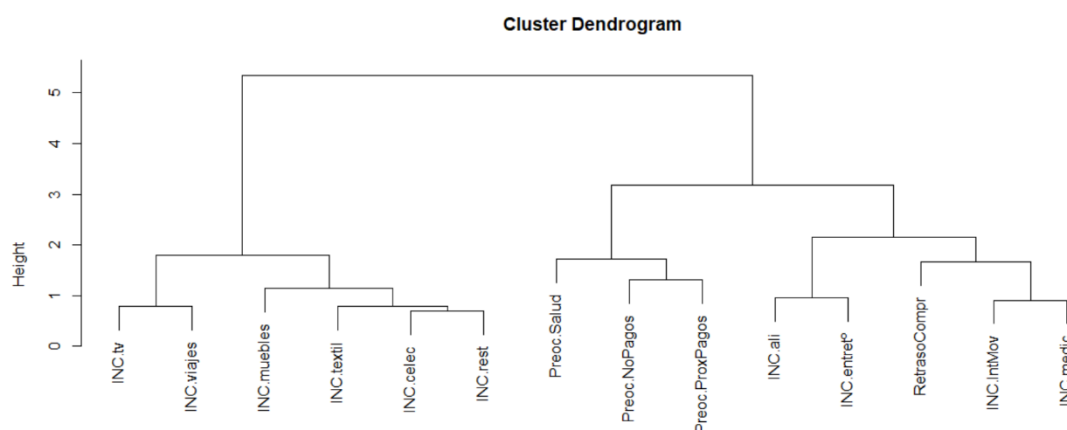
Al primer factor de la categoría consumo sectorial pasamos a denominarlo, al igual que al primer factor de la categoría ya analizada, “consumo necesario”, en tanto que representa el consumo, o gasto, necesario para poder vivir adecuadamente. Así, se incluyen gastos en transporte, salud, restaurantes, y otros. Al segundo factor se le llama “nuevas actividades”, en tanto que, en el momento pandémico y post-pandémico del año en cuestión emergieron muchas actividades que antes se practicaban con menos frecuencia. Por ejemplo, el gasto en alimentación se disparó. Como la gente pasó a estar más tiempo en casa, le dedicó más tiempo a tareas de hogar, entre ellas el bricolaje y menaje, reparaciones, telefonía de hogar, etc. Surgieron nuevas aficiones por actividades como la fotografía y los juguetes -por ejemplo, el gasto en juegos de mesa para disfrutar de la compañía en casa. Al grupo conformado por el tercer factor, se le denomina “ocio”,



ya que lo conforman las variables que determinan el consumo de actividades no fundamentales y más lúdicas.

Parece que los resultados extraídos del análisis factorial siguen un criterio, y que las agrupaciones tienen un sentido y razón subyacente.

El caso de *consumer tracker* el análisis es distinto, pues, como se trata de una categoría muy poco correlada, no se puede aplicar análisis factorial, teniendo que llevar a cabo un clustering. Para determinar la manera en que se agrupan las variables que componen esta categoría, se ha llevado a cabo un análisis de clúster con un árbol de clasificación (clustering jerárquico), y k-means (clustering no jerárquico), ambos llegando a conclusiones y resultados muy similares<sup>8</sup>:



**Figura 3:** Árbol jerárquico para *consumer tracker*.

**Fuente:** *Elaboración propia.*

Hay dos grupos claramente diferenciados, uno compuesto por todos los indicadores simples que miden la preocupación ciudadana, y otro que recoge el resto. No obstante, como esta categoría está muy poco correlada y la asignación de pesos se llevará a cabo mediante una técnica que requiere su división en grupos homogéneos constituidos por un número máximo de entre 7 y 8 indicadores simples (de esta técnica se hablará más adelante), procede, en este punto, extraer un mayor número de clústeres, para que sean de mayor utilidad en los pasos posteriores. Tales grupos quedan de la manera presentada en la siguiente Tabla:

<sup>8</sup> Se puede ver el árbol jerárquico y el *output* de k-means el Anexo

<b>CONSUMER TRACKER</b>			
CLÚSTER 1	CLÚSTER 2	CLÚSTER 3	CLÚSTER 4
INC electrónica	Preoc. no pagos	INC alimentación	INC Internetmóvil
INC muebles	Preoc. salud	INC entretenimiento	INC medicamentos
INC restauración	Preoc. próx. pagos		Retraso compras
INC textil			
INC TV			
INC viajes			

**Tabla 20:** Agrupación de indicadores simples en *consumer tracker*.

*Fuente:* Elaboración propia.

Al primer clúster se le podría llamar “INC nueva vida”, en tanto que representa la intención neta de consumo en las actividades que, para aquel entonces, comenzaron a ser relevantes: la electrónica, los muebles -por todo el tiempo que la gente pasó en casa- o la televisión, entre otros.

El segundo clúster parece claro que representa la “preocupación”, tanto por la propia salud, como por la situación económica que se empezó a vivir en tal momento: la gente estaba preocupada por no haber realizado pagos, y por no poder afrontar pagos futuros.

El tercer clúster lo conforma la “primera necesidad”, puesto engloba la intención neta de consumo en alimentación y en entretenimiento. Si bien esto último no es considerado primera necesidad, hay que pensar en tal variable en un momento como el vivido en el año 2020, en el que la gente sufrió un confinamiento y grandes restricciones en la vida cotidiana, y el entretenimiento, en un contexto de vida muy preocupante -a nivel económico y por la salud- y poco ocioso, pasó a ser fundamental.

El cuarto clúster es probablemente el más difícilmente relacionado, ya que engloba la intención neta de consumo de internet en el móvil, de medicamentos, y el retraso en las compras de gran volumen, por lo que se le puede denominar “otros”.

Todo este análisis multivariante nos ha permitido estudiar y conocer de forma más cercana el comportamiento de los datos, las correlaciones, y los grupos que tienden a formarse. Todo ello nos ayudará a tomar mejores decisiones en los puntos siguientes.

## 2.4. Normalización de los datos

Ya se adelantó en puntos anteriores la necesidad de normalizar los datos antes de que sean tratados. Por un lado, para que se encuentren en la misma unidad de medida. Por otro lado, para que los datos no presenten distintos rangos de variación (Bas, 2014). Si una categoría se compone de variables que están en distintas escalas, sus resultados podrían quedar desvirtuados, ya que el indicador final sería reflejo de aquellos valores excesivamente altos o bajos en la base de datos.

Hay varias técnicas de normalización de datos, pero dos de las más usadas son la estandarización y el mín-máx. Aquella escala las variables de tal manera que tengan media cero y desviación típica uno, y el resultado muestra el número de desviaciones que tal observación se encuentra por encima o por debajo de la media de la variable en cuestión. Toma la forma prevista en la ecuación (8).

$$I_i^t = \frac{x_i^t - x_i}{\sigma_i} \quad (8)$$

Donde  $I$  es la observación normalizada el indicador base  $i$  y el mes  $t$ ;  $x$  es el dato original,  $\bar{x}$  es la media del indicador en cuestión, y  $\sigma$  la desviación típica del mismo.

La normalización mín-máx, por otro lado, normaliza todas las variables en una escala del 0 al 1, teniendo en cuenta los valores mínimos y máximos de la variable en cuestión. De este modo, todas las variables contarán con una observación cuyo valor sea 0 -aquella observación con peor valor-, y otra con valor 1 -la mejor. Para la construcción del presente indicador, se ha escogido la normalización mín-máx, ya que proporciona el resultado de las variables entre dos números siempre idénticos, el 0 y el 1. Se expresa con la siguiente ecuación (9).

$$I_i^t = \frac{x_i^t - \min(x_i)}{\max(x_i) - \min(x_i)} \quad (9)$$

Otro factor a tener en cuenta a la hora de normalizar los datos es la interpretación que se le debe dar a cada variable. En función del significado de la misma, sus datos pueden venir definidos de manera que, a mayor valor del dato, mejor, o que cuanto mayor sea, peor es la situación que está representando. Véase el siguiente cuadro, que repasa las variables de cada categoría, qué representan, y por tanto su interpretación:

RESUMEN DE LOS KPIs DE CONSUMO GENERAL		
KPIs	SIGNIFICADO	INTERPRETACIÓN
BBVA_ATM	Gasto en cajeros ATM	A mayor, mejor
BBVA_Entertainment	Gasto en entretenimiento	A mayor, mejor
BBVA_Food	Gasto en alimentación	A mayor, mejor
BBVA_Health	Gasto en salud	<i>*A mayor, peor</i>
BBVA_Hotels	Gasto en hoteles	A mayor, mejor
BBVA_Online	Compras <i>online</i>	A mayor, mejor
BBVA_Physical	Gasto en compras físicas	A mayor, mejor
BBVA_Restaurants	Gasto en restaurantes	A mayor, mejor
BBVA_Total.Consumption	Gasto en consumo total	A mayor, mejor
BBVA_Transport	Gasto en transporte	A mayor, mejor
ICC	Intención de gasto	A mayor, mejor
ICPM	Empleo en comercio minorista	A mayor, mejor
Consumo eléctrico	Consumo de la electricidad	A mayor, mejor

**Tabla 21:** Interpretación de los indicadores simples en consumo general.

*Fuente: Elaboración propia.*

RESUMEN DE LOS KPIs DE CONSUMO SECTORIAL					
KPIs	SIGNIFICADO	INTERPRETACIÓN	KPIs	SIGNIFICADO	INTERPRETACIÓN
AlimentaciónHipermercados	Gasto en hipermercados	A mayor, mejor	OtrosOtros	Otros gastos	A mayor, mejor
AlimentaciónResto	Gasto en alimentación	A mayor, mejor	OtrosOtrosComercios	Gasto en otros comercios	A mayor, mejor
BellezaCalzado	Gasto en calzado	A mayor, mejor	OtrosParking	Gasto en parking	A mayor, mejor
BellezaComplementos	Gasto en complementos	A mayor, mejor	OtrosTributos	Gasto en tributos	A mayor, mejor
BellezaRopa	Gasto en ropa	A mayor, mejor	RestauraciónBar	Gasto en bares	A mayor, mejor
HogarBricolajeYmenaje	Gasto en bricolaje y menaje	A mayor, mejor	RestauraciónCafeterías	Gasto en cafeterías	A mayor, mejor
HogarHogar	Gasto para el hogar	A mayor, mejor	RestauraciónRestaurantes	Gasto en restaurantes	A mayor, mejor
HogarMascotas	Gasto para mascotas	A mayor, mejor	SaludSalud	Gasto en salud	<i>*A mayor, peor</i>
HogarMuebles	Gasto en muebles	A mayor, mejor	SaludBelleza	Gasto en belleza	A mayor, mejor
HogarReparaciones	Gasto en reparaciones	A mayor, mejor	SaludDeporte	Gasto en deporte	A mayor, mejor
HogarSeguros	Gasto en seguros	A mayor, mejor	TransporteAvion	Gasto en avión	A mayor, mejor
HogarTelefonía	Gasto en telefonía	A mayor, mejor	TransporteCoche	Gasto en coche	A mayor, mejor
OcioDeportes	Gasto en deportes	A mayor, mejor	TransporteMetro	Gasto en metro	A mayor, mejor
OcioElectrónica	Gasto en electrónica	A mayor, mejor	TransporteOtros	Gasto en otros	A mayor, mejor
OcioEstudios	Gasto en estudios	A mayor, mejor	TransporteTaxi	Gasto en taxi	A mayor, mejor
OcioFotografía	Gasto en fotografía	A mayor, mejor	TransporteTren	Gasto en tren	A mayor, mejor
OcioJuguetes	Gasto en juguetes	A mayor, mejor	TurismoAgenciasdeViajes	Gasto en agencias de viajes	A mayor, mejor
OcioLoterías	Gasto en loterías	A mayor, mejor	TurismoHoteles	Gasto en hoteles	A mayor, mejor
OcioOcio	Gasto en ocio	A mayor, mejor	TurismoViajes	Gasto en viajes	A mayor, mejor
OtrosAutomóvil	Gasto en automóvil	A mayor, mejor	M. V.	Compra de vehículos	A mayor, mejor
OtrosBazares	Gasto en bazares	A mayor, mejor	M. V. para empresas	Compra de vehículos	A mayor, mejor
OtrosDonaciones	Donaciones	A mayor, mejor	M. V. para particulares	Compra de vehículos	A mayor, mejor
OtrosGasolineras	Gasto en gasolineras	A mayor, mejor			

**Tabla 22:** Interpretación de los indicadores simples en consumo sectorial.

*Fuente: Elaboración propia.*

RESUMEN DE LOS KPIs DE CONSUMER TRACKER		
KPIs	SIGNIFICADO	INTERPRETACIÓN
INC alimentación	Intención de consumo en alimentación	A mayor, mejor
INC electrónica	Intención de consumo en electrónica	A mayor, mejor
INC entretenimiento	Intención de consumo en entretenimiento	A mayor, mejor
INC Internetmóvil	Intención de consumo en internet en el móvil	A mayor, mejor
INC medicamentos	Intención de consumo en medicamentos	<i>*A mayor, peor</i>
INC muebles	Intención de consumo en muebles	A mayor, mejor
INC restauración	Intención de consumo en restauración	A mayor, mejor
INC textil	Intención de consumo en textil	A mayor, mejor
INC TV	Intención de consumo en TV	A mayor, mejor
INC viajes	Intención de consumo en viajes	A mayor, mejor
Preoc. No Pagos	Preocupación por impagos	<i>*A mayor, peor</i>
Preoc. Salud	Preocupación por propia salud	<i>*A mayor, peor</i>
Preoc. Próx. Pagos	Preocupación por no poder afrontar próximos pagos	<i>*A mayor, peor</i>
Retraso Compras	Retraso en compras de gran volumen	<i>*A mayor, peor</i>

**Tabla 23:** Interpretación de los indicadores simples en *consumer tracker*.

*Fuente: Elaboración propia.*

Se puede ver que la mayoría de las variables de las que se dispone se interpretan de manera que cuanto mayor sea su valor, mejor resulta la realidad observada. En este caso, que se mide la recuperación del consumo en España tras el Covid-19, cuanto mayor sea la intención de gasto, o consumo, mejor, pues ello representa una mayor salud económica. No obstante, nos encontramos algunas variables cuya interpretación podría generar controversias, y son aquellas que miden el gasto o la intención de gasto en el ámbito médico. Desde un punto de vista puramente económico y centrándonos en lo que es ámbito de estudio del presente trabajo, un mayor gasto en materia médica implica un mayor índice de consumo, lo cual se podría considerar positivo. Desde un punto de vista más allá del puramente económico, se podría considerar que el gasto en el ámbito sanitario viene motivado por un problema de salud subyacente, por lo que un mayor gasto en este ámbito podría ser tomado como negativo y no como positivo. Como ambas líneas de pensamiento se podrían fundamentar rigurosamente, se plasmarán todos los escenarios posibles en el análisis de sensibilidad y robustez, si bien en las tablas que se muestren hasta entonces se va a enseñar los resultados de una interpretación de que a mayor en el ámbito sanitario, peor. Las siguientes tres variables representan el grado de preocupación poblacional en tres áreas distintas: por no haber realizado ciertos pagos, por la propia salud, y por no poder afrontar próximos pagos. A mayor nivel de preocupación en todos estos ámbitos, peor la situación reflejada. Por último, la pandemia mencionada afectó sobremanera a las compras y a su distribución, tanto a nivel nacional como internacional. A mayor retraso en compras de gran volumen, peor la situación pandémica y postpandemia vivida a nivel nacional y mundial.

La interpretación que se le da a las variables debe ser homogénea: al igual que no se puede mezclar variables con escalas distintas, tampoco se puede mezclar variables cuyo mayor valor es mejor, con aquellas cuyo menor valor es el mejor. Esto se trata en el paso de normalización de los datos. Para tratar el problema de la escala, se normalizarán los datos con la técnica mín-máx, que sigue la fórmula ya presentada en la ecuación (9).

La interpretación que se quiere dar al resultado es que a mayor valor, mejor, de tal manera que 1 represente el mejor valor posible de la variable en cuestión, y que 0 represente el peor. Para aquellas variables cuya interpretación sea la contraria, el resultado de la ecuación (9) se le resta a 1, de tal forma que se aplica la siguiente<sup>9</sup>:

---

<sup>9</sup> Mismos índices y subíndices que en la ecuación (9).

$$I_i^t = 1 - \frac{x_i^t - \min(x_i)}{\max(x_i) - \min} \quad (10)$$

Si se aplica esta última ecuación a las variables ya señaladas -por tener una interpretación de que a mayor valor, peor resultado-, y la primera al resto, ya se tiene todas las observaciones de la base de datos escaladas a la misma unidad, y con la misma interpretación, y se puede proceder a la construcción final del indicador sintético.

## 2.5. Pesos y agregaciones

Se podría decir que un indicador sintético resume, en un único dato, la información de muchos otros indicadores simples sobre una realidad multidimensional. La manera en que se logra la síntesis de los datos recogidos por varias variables en una sola es mediante la asignación de pesos y sus agregaciones.

Los pesos determinan, por tanto, la importancia relativa del indicador base sobre el indicador final, precisando en su construcción de dos fases en caso de que la base de datos se encuentre dividida por categorías (como ocurre con la base de datos aquí presente). En una primera fase, se establecen los pesos de los indicadores base sobre el total de su categoría y se agregan. En una segunda fase, se le asigna pesos a cada categoría y, mediante la agregación, se construye el indicador final. Así, para la creación del indicador final nos ayudamos de una estructura de agregación jerárquica, en la que los indicadores base se agregan en indicadores de categoría (en este caso, consumo general, consumo sectorial y *consumer tracker*) y estos, a su vez, en el indicador final.

Los pesos que se puede asignar a cada indicador base y de categoría pueden ser dados de dos maneras alternativas:

- Mediante la equiponderación, esto es, dar el mismo peso, a cada indicador en cuestión. Se suele acudir a este método de imputación por su simple construcción, cuando no existe una razón subyacente que demuestre una importancia relativa dispar de cada indicador simple, o cuando no se llega a un consenso al respecto (Schuschny & Soto de la Rosa 2009; Greco *et al* 2018). Siendo  $N$  el número de indicadores total en la categoría  $c$ , el peso relativo  $w$  de cada indicador  $i$  para tal categoría  $c$  se define de la forma ilustrada en la ecuación (11).

$$w_{ic} = \frac{1}{N_c} \quad (11)$$

Nótese que la asignación de pesos equitativa presume una importancia relativa idéntica de cada indicador base en su categoría (y no en el indicador final, pues los indicadores simples de una categoría con un menor número de los mismos tendrán una importancia relativa en el indicador compuesto mucho mayor a aquéllos pertenecientes a una categoría con un mayor número de indicadores). Por ello, se debe tener muy presente las correlaciones existentes entre los distintos indicadores base, ya que, si varios de ellos cuentan con una alta correlación y se le da a cada uno la misma importancia relativa que a otro indicador no correlado con ninguna otra variable, la realidad que representen aquéllos influirá en mayor medida al indicador final (Bas, 2014).

- Mediante la asignación de distintos pesos en función, normalmente, de la importancia relativa de cada indicador base, de la calidad estadística de cada uno de ellos, o por ambas razones (Bas, 2014). A su vez, estos métodos pueden ser estadísticos-matemáticos, o participativos (en los que el criterio experto adquiere mayor relevancia). Algunos ejemplos de estos últimos son<sup>10</sup>:

-- Budget allocation process (BAP), o “método de asignación presupuestaria”, que consiste en que expertos en el tema determinen, en función de sus conocimientos y de la estructura de los datos analizados, el peso correspondiente a cada variable.

-- Determinación de la importancia de cada variable en función de la opinión pública, a través de encuestas al público (OECD, 2008).

-- Analytic hierarchy process (AHP), o “proceso de jerarquía analítica”, en el que se compara, por parejas, los indicadores simples, y se determina cuál de los dos es más relevante, y por cuánto.

-- Conjoint analysis (CA), o “análisis conjunto”, que se podría decir que es lo contrario del método anterior. En este caso, se presenta una multitud de escenarios, y se pide que se escoja el favorito.

Dentro de los métodos matemáticos-estadísticos de asignación de pesos encontramos lo siguientes:

---

<sup>10</sup> Para un mayor desarrollo de cada uno de estos métodos, véase Bas (2014), pp.96-101.



-- Análisis de correlaciones, en el que se determinan los pesos en función de las correlaciones entre variables. Tales pesos se pueden asignar con base en la matriz de correlaciones, o en la “capacidad de la información” que recoge cada indicador (Hellwig 1969; Ray 1989, como se citó en Greco *et al* 2018).

-- Método de pesos basado en el análisis factorial o análisis de componentes principales (más conocidos por sus siglas en inglés, *FA: Factor Analysis* y *PCA: Principal Component Analysis*). Como ya se adelantó en el Análisis Multivariante, ambas son técnicas que explican la relación entre variables y su agrupación, que comparten núcleos matemáticos muy similares, con algoritmos análogos, si bien no idénticos (Gniazdowski 2021). PCA trata de reducir el número de variables originales teniendo en cuenta la varianza total. Así, extrae, en un número reducido de variables (no correlacionadas entre sí), la mayor información posible del conjunto de datos. Se utiliza en bases de datos con muchas observaciones. FA analiza las interrelaciones entre variables según su varianza “común” (o covarianza), de tal manera que busca la extracción de factores que expliquen, las correlaciones entre las variables, es decir, lo “común” a todas ellas. De esta manera, se podría decir que el PCA extrae componentes (o factores) que son combinación lineal de las variables originales (en palabras de Arredondo (2019. Introducción, último párrafo), “las variables explican los factores”), mientras que el FA extrae factores de manera que las variables originales sean combinaciones lineales de éstos (dicho de otro modo “los factores explican las variables” Arredondo (2019. Introducción, último párrafo)) (Arredondo, 2019). Un requisito de ambos métodos es que la base de datos original se encuentre altamente correlacionada.

-- Data envelopment Analysis (DEA), (en español, “análisis envolvente de datos”), un método que utiliza la programación lineal para establecer una frontera eficiente a partir de la cual se calcula el desempeño de un conjunto de unidades de análisis (Bas 2014).

-- Benefit of the doubt approach (BOD), o “método del beneficio de la duda”. Es el nombre que se le da a la aplicación del DEA a los indicadores compuestos.

-- Unobserved components model (UCM), (en español, modelo de los componentes no observables), y es un análisis regresivo donde se asume que los indicadores simples dependen de una variable que no ha sido observada más un error (OECD 2008).

Nótese que es altamente probable que la decisión que se tome en este punto afecte en gran medida al indicador compuesto. De todos los métodos expuestos, no hay ninguno



que se pueda definir objetivamente como “el mejor”, o el más adecuado para todos los conjuntos de datos. De esta manera, se debe, en función de la base de datos disponible, determinar cuál se considera el más adecuado, y justificar su elección de la manera más explícita y transparente posible (Schuschny & Soto de la Rosa 2009). Teniendo todo esto en cuenta, las técnicas utilizadas para la determinación de los pesos en el presente caso y sus justificaciones son las siguientes:

- Para hallar los pesos de los indicadores base se ha acudido a los métodos de asignación de pesos en función de la importancia relativa de indicador en la categoría a la que pertenece<sup>11</sup>. En concreto, los siguientes:

-- Para las categorías consumo general y consumo sectorial, se va a aplicar el análisis factorial, en tanto que se trata de una técnica fácilmente aplicable con la herramienta SPSS Statistics, y que es capaz de aportar resultados de interés, fácilmente analizables, y robustos. Se escoge el FA y no el PCA porque, a efectos de este trabajo, interesa más tener factores que expliquen las variables, y no a la inversa (Arredondo 2019). Recuérdese que un requisito fundamental para poder aplicar FA es que el conjunto de datos original se debe encontrar muy correlado, requisito que cumplen ambas categorías.

-- Para *consumer tracker*, categoría que no cumple con el requisito anterior, se debe acudir a un método participativo. En concreto, se ha escogido el BAP, pues no se tiene acceso a la opinión pública (por lo que se descarta tal técnica), y se requiere de gran experiencia para llevar a cabo un análisis tan pormenorizado como los requieren el AHP y el CA. Para la aplicación del BAP, se hará uso de los resultados obtenidos en el Análisis Multivariante, y se acudirá nuevamente a la herramienta RStudio para la implementación de otras técnicas que ayuden a acreditar la elección de unos pesos u otros.

Ya se ha adelantado que FA se ha llevado a cabo con la herramienta SPSS Statistics. Como se razonó en el apartado de Análisis Multivariante, el número de categorías escogido es 2 para consumo general y 3 para consumo sectorial. Se ha escogido opción de reducción de dimensiones factorial con rotación varimax, un método de rotación ortogonal que reduce el número de variables con cargas altas en cada factor, de manera que simplifica la interpretación de los factores, y permite asociar cada variable a un único factor (IBM 2021). El resultado obtenido de tal reducción de dimensiones con rotación

---

<sup>11</sup> Nótese que podría ser igualmente justificable que se hubiera decidido equiponderar los indicadores base, por lo que se añadirá este estudio al análisis de sensibilidad y robustez.

varimax se plasma en la matriz de componente rotado (en adelante, MCR), de la que se extraerá los pesos para cada uno de los indicadores base. La manera en que se obtiene tales pesos es escalando la MCR a la unidad, de la manera detallada en los pasos siguientes (OECD 2008).

Paso 1.- Siendo  $O$  la observación de la MCR para el indicador  $i$  y el factor  $f$ , se halla la varianza explicada  $VE$  por tal factor  $f$ . Véase la ecuación (12).

$$VE_f = \sum O_{if}^2 \quad (12)$$

Paso 2.- Se escala a unidad  $U$  todas las observaciones  $O$  de la MCR, dividiendo el cuadrado de cada observación para el indicador  $i$  y factor  $f$  entre la varianza explicada del factor en cuestión, tal y como se expone en la ecuación (13). Todo ello, para cada factor  $f$ , debe sumar 1. Además, en este paso, debemos asociar un indicador a un solo factor, lo que se hará seleccionando el mayor valor de cada indicador para cada factor.:

$$U_{if} = \frac{O_{if}^2}{VE_f} \quad (13)$$

Paso 3.- Se determina el porcentaje de varianza explicada  $PVE$  por cada factor  $f$  sobre la varianza total explicada con tantos factores como se haya utilizado para el análisis factorial. Véase la siguiente ecuación (14).

$$PVE_f = \frac{VE_f}{\sum VE_f} \quad (14)$$

Paso 4.- Se halla el peso  $P$  para el indicador base  $i$  en el factor  $f$  al que se ha asociado de la manera prevista en la ecuación (15).

$$w_{if} = U_{if} * PVE_f \quad (15)$$

Paso 5.- Se halla el peso final  $P$  para cada indicador  $i$  en la categoría  $c$  (pudiendo ésta ser consumo general o consumo sectorial) de la manera expuesta en la ecuación (16).

$$w_{ic} = \frac{w_{if}}{\sum w_{if}} \quad (16)$$

Con la aplicación de todas estas ecuaciones se ha llegado a los siguientes pesos:

ASIGNACIÓN PESOS CONSUMO GENERAL	
KPI	PESO
BBVA_ATM	0,10
BBVA_Entertainment	0,09
BBVA_Food	0,09
BBVA_Health	0,03
BBVA_Hotels	0,02
BBVA_Online	0,08
BBVA_Physical	0,09
BBVA_Restaurants	0,09
BBVA_Tot.Consumption	0,10
BBVA_Transport	0,09
ICC	0,07
ICPM	0,08
Consumo eléctrico	0,05

**Tabla 24:** Asignación de pesos a cada indicador simple de la categoría consumo general.  
*Fuente: Elaboración propia.*

ASIGNACIÓN PESOS CONSUMO SECTORIAL					
KPI	PESO	KPI	PESO	KPI	PESO
AlimentaciónHipermercados	0,01	OcioFotografía	0,03	SaludSalud	0,02
AlimentaciónResto	0,01	OcioJuguetes	0,03	SaludSaludBelleza	0,03
BellezaCalzado	0,02	OcioLoterías	0,01	SaludSaludDeporte	0,02
BellezaComplementos	0,02	OcioOcio	0,03	TransporteAvion	0,03
BellezaRopa	0,02	OtrosAutomovil	0,03	TransporteCoche	0,03
HogarBricolajeyMenaje	0,03	OtrosBazares	0,02	TransporteMetro	0,03
HogarHogar	0,02	OtrosDonaciones	0,02	TransporteOtros	0,02
HogarMascotas	0,03	OtrosGasolineras	0,03	TransporteTaxi	0,03
HogarMuebles	0,02	OtrosOtros	0,01	TransporteTren	0,02
HogarReparaciones	0,03	OtrosOtrosComercios	0,03	TurismoAgenciasdeViajes	0,03
HogarSeguros	0,01	OtrosParking	0,03	TurismoHoteles	0,03
HogarTelefonía	0,02	OtrosTributos	0,03	TurismoViajes	0,02
OcioDeportes	0,01	RestauraciónBar	0,02	Matriculación de vehículos	0,02
OcioElectrónica	0,03	RestauraciónCafeterías	0,02	M.V. para empresas	0,01
OcioEstudios	0,02	RestauraciónRestaurantes	0,02	M.V. para particulares	0,02

**Figura 25:** Asignación de pesos a cada indicador simple de la categoría consumo sectorial.

*Fuente: Elaboración propia.*

A *consumer tracker*, que se trata de una categoría muy poco correlada, no se le puede aplicar el análisis factorial, por lo que se han establecido los pesos con el método BAP. Así, en este caso, la asignación de pesos se puede llevar a cabo de dos formas distintas: una primera, equiponderar los pesos de toda la categoría, de manera que a cada variable se le asignaría un peso de 1/14. Una segunda, es tener en cuenta los cuatro grupos que se acaban de analizar. En este caso, se equiponderarían las variables correspondientes a cada grupo (por ejemplo, el clúster 3 lo componen dos variables, por lo que cada una tendría un peso de 0,5 en el clúster), y luego se pondera cada clúster (también equiponderado, de tal forma que cada uno pesa 0,25 ya que se cuenta con un total de 4 clústers). El peso final asignado a cada KPI sería el resultado de multiplicar el peso del KPI dentro del clúster

(en el ejemplo ilustrado es 0,5) por el peso del clúster en cuestión (que, en todos los casos es 0,25). Los resultados de ambos métodos (equiponderar los KPIs en el total categoría, o por clústers) se exponen a continuación, para elegir el método que más parece se asemeja con la realidad vivida en el año 2020:

ASIGNACIÓN PESOS <i>CONSUMER TRACKER</i>		
KPI	FORMA 1	FORMA 2
INC alimentación	0,07	0,13
INC electrónica	0,07	0,04
INC entretenimiento	0,07	0,13
INC Internetmóvil	0,07	0,08
INC medicamentos	0,07	0,08
INC muebles	0,07	0,04
INC restauración	0,07	0,04
INC textil	0,07	0,04
INC TV	0,07	0,04
INC viajes	0,07	0,04
Preoc. no pagos	0,07	0,08
Preoc. salud	0,07	0,08
Preoc. próx. pagos	0,07	0,08
Retraso compras	0,07	0,08

**Tabla 26:** Asignación de pesos a cada indicador simple de la categoría *consumer tracker*.

*Fuente:* Elaboración propia.

Parece coherente que se escoja el segundo método expuesto, ya que da más importancia a la intención neta de consumo de alimentación y entretenimiento -que, como ya se ha comentado, fueron aspectos muy relevantes en el año en cuestión-, seguido de la intención neta de consumo de internet en el móvil, de medicamentos, de la preocupación ciudadana por la salud y por lo económico y por el retraso en las compras de gran volumen, y finalmente, en un tercer grupo de nivel de importancia se encontraría la intención neta de consumo en electrónica, muebles, restauración, textil, TV y viajes. Efectivamente, si bien todos los aspectos mencionados eran relevantes en aquel entonces, también es cierto que estos últimos podrían ser, de todos ellos, los menos importantes en comparación con todas las variables en cuestión.

En cuanto a los pesos correspondientes a cada categoría (estas son, consumo general, consumo sectorial y *consumer tracker*), se le dará una similitud de pesos, pues al tratarse únicamente de tres categorías, el reparto de pesos de forma no equitativa tendría que estar razonado por una clara predominancia de una de las categorías, situación no contemplada en el caso actual. De esta manera, a cada categoría le corresponde un peso de 0,33.

Una vez se tienen los pesos para cada indicador base, se procede a agregarlos. Los métodos de agregación más generalizados son la agregación aritmética, la geométrica, y la multicriterio (OECD 2008). El primero suma los valores por su peso correspondiente, de la forma expuesta en la ecuación (17).

$$IC_c = \sum w_i I_{ic} \quad (17)$$

Donde  $IC$  es el indicador de categoría  $c$ , y es resultado de la suma de los pesos  $w$  para los indicadores base  $i$  por el valor observado normalizado  $I$ .

Este método de agregación es el más utilizado en la construcción de indicadores compuestos (Bandura y Martín, 2006; Nardo *et al.*, 2008, citados en Bas, 2014). Aun así, presenta dos inconvenientes fundamentales: presupone la independencia de los indicadores entre sí, característica muy difícil de conseguir en la realidad práctica, y adquiere un carácter compensatorio, de manera que valores muy bajos en el conjunto de datos se compensan con otros altos (Greco *et al.*, 2018).

Este carácter compensatorio de la agregación aritmética se puede salvar con la aplicación de otros métodos. Uno de ellos es la agregación geométrica, que, si bien no radica del todo esta compensabilidad característica de los métodos de agregación lineal, la suaviza considerablemente. Ello lo consigue mediante la penalización de los valores bajos (Bas, 2014), y toma la siguiente forma:

$$IC_c = \prod (I_{ic})^{w_i} \quad (18)$$

La agregación geométrica sí que tiene en cuenta la dependencia entre variables, y es utilizada para agregar indicadores “heterogéneos”, cuando el análisis se debe centrar en cambios porcentuales y no absolutos (Lafortune *et al.*, 2018).

El tercer método de agregación es el análisis de decisión multicriterio, no aplicable en este caso por la necesaria disponibilidad de información de varios países, que además se deben numerar en forma de *ranking*<sup>12</sup>.

Recuérdese que hay dos fases de determinación de pesos y agregación. Una primera, sobre cada indicador base. En este caso, se optará por la agregación lineal, ya que la geométrica podría castigar excesivamente los valores bajos, y desvirtuar el resultado

---

<sup>12</sup> Para más información, véase OECD (2008), pp. 104 y siguientes.

final<sup>13</sup>. Una segunda, que es la ponderación y agregación de los indicadores para cada categoría (consumo general, consumo sectorial y *consumer tracker*). En este segundo caso, se considera que podría ser aplicable tanto la agregación aritmética como la geométrica, ya que ambas son fácilmente aplicables y analizables, y ambas presentan sus ventajas e inconvenientes. Por ello, si bien en un principio, se decide optar por la agregación lineal, se plasmará los resultados que se hubiera obtenido mediante la agregación geométrica en el análisis de robustez y sensibilidad.

Véase a continuación las tablas de indicadores para cada categoría, e indicadores finales.

CONSUMO GENERAL		CONSUMO SECTORIAL		CONSUMER TRACKER		INDICADOR SINTÉTICO		
						AG. LINEAL	AG. GEOM.	
ene-20	0,87	ene-20	0,80	ene-20	0,55	ene-20	0,74	0,72
feb-20	0,81	feb-20	0,78	feb-20	0,48	feb-20	0,69	0,67
mar-20	0,24	mar-20	0,50	mar-20	0,47	mar-20	0,40	0,38
abr-20	0,15	abr-20	0,19	abr-20	0,49	abr-20	0,28	0,24
may-20	0,34	may-20	0,44	may-20	0,74	may-20	0,51	0,48
jun-20	0,58	jun-20	0,64	jun-20	0,64	jun-20	0,62	0,62
jul-20	0,63	jul-20	0,49	jul-20	0,59	jul-20	0,57	0,57
ago-20	0,62	ago-20	0,44	ago-20	0,47	ago-20	0,51	0,50
sep-20	0,58	sep-20	0,42	sep-20	0,37	sep-20	0,46	0,45
oct-20	0,52	oct-20	0,39	oct-20	0,22	oct-20	0,38	0,35
nov-20	0,56	nov-20	0,36	nov-20	0,61	nov-20	0,51	0,50
dic-20	0,61	dic-20	0,34	dic-20	0,49	dic-20	0,48	0,47

**Tabla 27:** Datos finales correspondientes a cada categoría y al indicador final.

*Fuente: Elaboración propia.*



**Figura 4:** Representación gráfica en línea temporal del indicador sintético final, agregado lineal y geoméricamente.

*Fuente: Elaboración propia.*

Se puede ver que ambas agregaciones presentan una tendencia similar, si bien, tal y como se comentó anteriormente, la agregación geométrica penaliza los valores más

<sup>13</sup> Un ejemplo puramente matemático es que, como la agregación geométrica *multiplica* los valores por sus correspondientes pesos, si una sola observación (que, en el caso de normalización mín-máx siempre la hay) toma valor nulo, toda la agregación quedaría con valor 0.

pequeños, y en este caso no se ven compensados por valores más altos. El indicador sintético, pues, muestra el consumo en España en el año 2020. Parece, en principio, que los resultados guardan relación y pueden ser buena representación de lo que ocurrió en la práctica. Para ello, memoremos brevemente lo que ocurrió en ese año y cómo llegó el coronavirus a España.

El primer caso positivo de Covid-19 en Europa se confirmó el 27 de diciembre, en Francia. En España no se dio hasta el 31 de enero, caso de un turista alemán en la isla de La Gomera. El siguiente caso se confirmó el 10 de febrero en Mallorca, por un turista británico, y más tarde, el 25 del mismo mes, por un turista italiano en la isla de Tenerife. Estos son los primeros casos de coronavirus que se dieron en España, pero verdaderamente la primera ola no llegó hasta la semana del 9 al 15 de marzo, que es cuando se decretó el Estado de alarma, siendo éste último día el primero del confinamiento que a todos los españoles esperaba.

En cuanto a la relación de los resultados del indicador compuesto con estos datos, se aprecia claramente un alto nivel de consumo en el mes de enero, que minora drásticamente durante los meses de febrero y marzo, llegando a su pico más bajo en el mes de abril. El descenso apreciado en los dos primeros meses no parece que pueda ser consecuencia del Covid-19, sino más bien de una tendencia general a consumir menos tras el mes de enero, en el que el nivel de consumo aumenta considerablemente. El pico bajo del mes de abril, no obstante, sí que ya podría ser consecuencia económica de la pandemia. Tras dos meses de confinamiento, si bien aún en situaciones muy adversas, el confinamiento comienza a ser la nueva realidad de todos los españoles, y parece que se empieza a recuperar el nivel de consumo en el país. No obstante, con el impacto de la segunda ola (de junio a diciembre del 2020) y, principalmente, con la aprobación del toque de queda el 25 de octubre del mismo año -que desmotivó a muchos españoles a recuperar el estilo de vida anterior que, parecía, se estaba restableciendo tras la salida del confinamiento el 21 de junio-, el consumo se vuelve a ver enormemente afectado. En los meses siguientes, parece que se vuelve a recuperar, pero sin llegar a los niveles que había llegado a haber a principios del año 2020.

## 2.6. Análisis de la robustez y sensibilidad

A lo largo de la construcción del indicador sintético se ha ido tomando una serie de decisiones de importante talante subjetivo que pueden haber influido en gran medida al resultado obtenido. Para garantizar que las decisiones, si bien se han venido justificando a lo largo de todo el trabajo, están tomadas con cierto criterio, se precisa, en última instancia, de un análisis de la robustez y sensibilidad.

El conjunto de decisiones va a venir representado por  $X_n$ , donde  $n$  hace referencia al punto en que tal decisión se ha tomado. Cuando  $X_n$  toma como valor la unidad, se indica que es la decisión que se ha seguido en el presente trabajo. Por el contrario, cuando toma valor nulo, hace referencia a una decisión alternativa

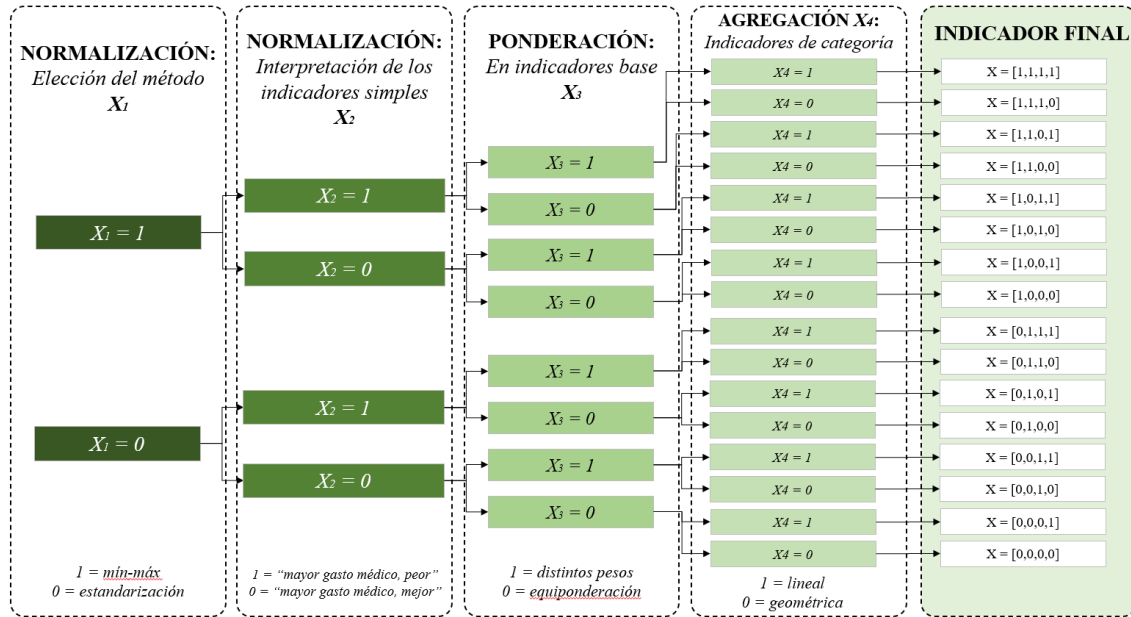
- Las dos primeras grandes decisiones se tomaron en el punto de normalización de los datos. Por un lado, en el método de normalización escogido ( $X_1$ ), y, por otro lado, en la interpretación que se dio a los indicadores medidores del gasto en el ámbito sanitario ( $X_2$ ). En lo que a  $X_1$  se refiere, se optó por la técnica mín-máx, si bien también se podría haber optado por la estandarización de los mismos. En  $X_2$ , de una manera mucho más dudosa, se optó la interpretación de que a mayor gasto en el ámbito sanitario, peor. En este punto se va a analizar precisamente qué hubiera pasado con el indicador final si se hubiera tomado las decisiones alternativas. Como se ha indicado *supra*, si  $X_1$  toma el valor 1, se estará indicando que se sigue una normalización mín-máx (esta es la que, en principio, se ha seguido). Si  $X_1$  toma el valor 0, se hace referencia a la estandarización. Lo mismo con  $X_2$ , que cuando toma valor 1, indica una interpretación de “a mayor, peor”, mientras que cuando toma valor 0, indicará una interpretación de “a mayor, mejor”.

- Una tercera decisión que puede haber afectado en gran medida a los resultados obtenidos, ha sido la de dar distintos pesos a los indicadores base indicador, y no equiponderarlos.  $X_3$ , por tanto, hace alusión a la decisión relativa al método con el que se ha determinado los pesos para cada indicador simple en cada una de las categorías.

- En el apartado de la agregación de los indicadores de categoría se indicó que, si bien se optaba por la agregación lineal, la geométrica también podría ser aplicable, por lo que  $X_4$  refleja esos escenarios.

El conjunto de decisiones tomadas y alternativas en cada fase de construcción del indicador compuesto se puede ver de manera visual en la Figura 5.





**Figura 5:** Esquema de las decisiones alternativas en cada paso de construcción del indicador compuesto.

*Fuente: Elaboración propia.*

La interpretación que se le debe dar a la Figura anterior es la siguiente: para el caso en que  $X = [1,1,1,1]$  representa el indicador con normalización mín-más, interpretación de “a mayor gasto médico, peor”, distintos pesos para los indicadores base de cada categoría, y agregación lineal de los indicadores de categoría.

Los resultados para cada uno de estos escenarios son los siguientes:



**Figura 6:** Escenarios de las distintas alternativas (I).

*Fuente: Elaboración propia.*

Viendo los resultados obtenidos en la Figura 6, se pueden extraer diversas conclusiones. La primera, y probablemente más importante, es que el comportamiento general del indicador es similar en todos los casos -con los matices que se expondrá a continuación: el pico más alto de nivel de consumo está en el mes de enero, viéndose altamente afectado de manera negativa durante los meses siguientes, llegando a su punto más bajo en abril. A partir de entonces, el consumo se recupera notoriamente, alcanzando su punto más alto en el mes de junio, que, no obstante, nunca llega a los niveles tan prominentes de finales del 2019 – principios del 2020. El segundo pico de bajo nivel de consumo se encuentra en el mes de octubre, para finalmente comenzar a estabilizarse en los últimos meses del año. Las razones que pueden motivar este comportamiento del indicador ya se comentaron en el apartado 2.5.

Si bien la estructura general del indicador es homogénea en prácticamente todos los casos, hay que matizar los resultados obtenidos. Los dos gráficos que se encuentran en la parte superior de la Figura 6 muestran los resultados de un indicador con normalización de datos mín-máx. En el gráfico de la izquierda se le ha dado una interpretación de que, a mayor nivel de consumo en asuntos médicos, peor, mientras que el gráfico derecho presenta un razonamiento a la inversa.

Los dos gráficos de la parte inferior de la figura muestran los resultados obtenidos a partir de una estandarización de los datos. En estos este caso, parece claro que la agregación que se debe seguir es la lineal, y no la geométrica. Como la estandarización escala todas las variables a la misma unidad -con media cero y desviación típica 1-, los datos normalizados no oscilan en un rango concreto, por lo que hay mucha presencia de observaciones negativas. Como la agregación geométrica multiplica los indicadores simples, muchos valores negativos se ven transformados en positivos (menos por menos, es más), de manera que el resultado final del indicador queda muy desvirtuado, mostrando valores positivos -véase los meses de abril y mayo- cuando verdaderamente son positivos. Descartando estas dos opciones, las otras ocho presentan resultados más homogéneos, si bien cuando se interpreta de que a mayor gasto en asuntos médicos mejor, la figura aparece más plana.

Los resultados de la agregación geométrica, aplicable a los dos gráficos superiores, son ligeramente inferiores a aquellos de la agregación lineal, por lo que se comentó en

este documento respecto de la compensación y que la agregación geométrica penaliza en mayor medida los resultados negativos o muy bajos.

En cuanto a la manera de otorgar los pesos, resulta sorprendente que no hay mucha diferencia entre los resultados obtenidos con la asignación de pesos estadísticos y aquellos obtenidos mediante equiponderación. Probablemente esto se debe a que los pesos dados a cada indicador no difieren mucho de un método u otro, o, siendo cada uno de distinta importancia, se compensan. Si bien se esperaba una mayor discrepancia entre los resultados de pesos estadísticos y equiponderación, parece que los resultados en todos los casos son robustos y poco sensibles a pequeños cambios, lo cual indica que el indicador es fiable y, en la medida de lo posible, objetivo.

### **3. CONCLUSIONES**

La cantidad de datos creciente en las últimas décadas ha propagado sobremanera la capacidad de información y conocimiento sobre cualquier tipo de materia. Se podría decir, que hace unos años se buscaba una mayor información, pues no se tenía toda la suficiente, mientras que hoy día se busca su síntesis y resumen. ¿Cómo explicar tantas observaciones de una manera sencilla e inteligible? Ese es precisamente el objetivo buscado por los indicadores sintéticos.

Si bien se trata de una herramienta muy útil para el análisis de datos y distintas realidades que está adquiriendo una mayor importancia, presenta ciertas desventajas, ya nombradas en este documento, que no pueden quedar olvidadas. Como se ha ido viendo a lo largo de la construcción del indicador sintético medidor del consumo en España durante el año 2020, el número de decisiones que debe tomar el autor del mismo son numerosas, incluso dentro de cada etapa de construcción del indicador. De esta manera, si se consigue que los indicadores sintéticos sean robustos y tales decisiones estén debidamente justificadas -y, a ser posible, respaldadas por datos empíricos-, es indudable que los indicadores sintéticos serán -si bien ya lo son- una herramienta muy útil para su aplicación en distintos ámbitos.

En lo que al presente trabajo se refiere, se ha logrado la construcción de un indicador sintético medidor del consumo en España en el año 2020, que ayuda a analizar en qué medida y qué patrón ha seguido la recuperación del país en este ámbito. El análisis de sensibilidad y robustez ha demostrado que los resultados obtenidos son robustos. Si tales

resultados se comparan, además, con la Figura 1, que presenta en un gráfico de línea temporal el PIB de España en una franja temporal más amplia que la del indicador aquí creado, pero que incluye tales meses, se puede ver que los datos son coincidentes, lo cual considero que es prueba de que el indicador construido es reflejo de la realidad vivida en el año en cuestión.

Expuesto todo lo anterior, también cabe mencionar que en numerosas decisiones para cada etapa, se ha tenido que descartar técnicas complejas pero muy potentes, bien porque los datos disponibles no cumplían con ciertas características, bien por la falta de recursos de este autor, y cuya aplicación habría sido de gran interés para el presente trabajo, y lo habría dotado de un mayor rigor científico.

#### 4. BIBLIOGRAFÍA Y REFERENCIAS.

Andrino, B., Grasso, D. & Llaneras, K. (2021). *Los datos de una pandemia en tres olas.*

El País. Disponible en: <https://elpais.com/sociedad/2021-03-10/los-datos-de-una-pandemia-en-tres-olas.html>

Arredondo Sánchez, A.E. (2019). *Proyecto de reducción de dimensionalidades: Análisis de Componentes Principales (PCA) y Análisis Factorial (FA).* RPubs by RStudio.

Disponible en: <https://rpubs.com/andrearredondo/477949>

Bas Cerdá, M.C. (2014). *Estrategias metodológicas para la construcción de indicadores compuestos en la gestión universitaria.* Universidad Politécnica de Valencia.

Disponible en: <https://riunet.upv.es/handle/10251/35330#>

Blancas Peral, F. J., Contreras Rubio, I. & Ramírez Hurtado, J. M. (2011). *Construcción de indicadores sintéticos: una aproximación para maximizar la discriminación.* Universidad Pablo de Olavide.

Cabrera Valverde, D. G., Rosero Oliveros, T. M. & Riascos, J. C. (2019). *Principales determinantes económicos del consumo.* Disponible en:

<http://www.scielo.org.co/pdf/tend/v20n1/2539-0554-tend-20-01-00077.pdf>

Chen, Y. & Shao, J. (1999). *Inference with survey data imputed by hot deck when imputed values are nonidentifiable.* Statistica Sinica. Disponible en:

<https://www3.stat.sinica.edu.tw/statistica/oldpdf/A9n24.pdf>

- Culebras, J.M., San Mauro Martín, I. & Vicente-Vacas, L. (2020). *Covid-19 y otras pandemias*. Disponible en: <https://scielo.isciii.es/pdf/jonnpr/v5n6/2529-850X-jonnpr-5-06-644.pdf>
- Durán Durán, Y. (2022). *Cómo medir el desarrollo de los países más allá del crecimiento económico*, *El Economista*, 14 de febrero [online]. Disponible en: <https://www.economista.com.mx/arteseideas/Como-medir-el-desarrollo-de-los-paises-mas-alla-del-crecimiento-economico-20220214-0034.html>
- Editorial Grudemi. (2018). *Producto interno bruto*, actualizado en junio de 2022. Recuperado de Enciclopedia Económica: <https://enciclopediaeconomica.com/pbi/>
- Editorial Grudemi. (2021). *Consumo*, actualizado en junio de 2022. Recuperado de Enciclopedia Económica: <https://enciclopediaeconomica.com/consumo/>
- Freudenberg, M. (2003). *Composite Indicators of Country Performance: A Critical Assessment*, *OECD Science, Technology and Industry Working Papers*, 2003/16, publicación de la OECD. doi:10.1787/405566708255
- Gastón Lorente, L. (2020). *Cómo calcular el PIB: Tres métodos*, BBVA, actualizado el 28 de agosto [online]. Disponible en: <https://www.bbva.com/es/tres-metodos-calcular-pib/>
- Gniazdowski, Z. (2021). *Principal Component Analysis versus Factor Analysis*. Warsaw School of Computer Science. Disponible en: [https://www.researchgate.net/publication/355207581\\_Principal\\_Component\\_Analysis\\_versus\\_Factor\\_Analysis](https://www.researchgate.net/publication/355207581_Principal_Component_Analysis_versus_Factor_Analysis)
- Greco, S., Ishizaka, A., Tasiou, M. & Torrisi, G. (2018). *On the Methodological Framework of Composite Indices: A Review of the Issues of Wighting, Aggregation, and Robustness*. *Social Indicators Research*.
- Guías Jurídicas (s.f.). *Consumo*. Disponible en: <https://guiasjuridicas.wolterskluwer.es/Content/Documento.aspx?params=H4sIAAA AAAAEAMtMSbF1jTAAASMjCzMjtbLUouLM DxbIwMDS0NDA1OQQGZapUt = ckhIQaptWmJOcSoA91YqlIDUAAAA=WKE#:~:text=Se%20llama%20funci%C3%B3n%20de%20consumo,renta%20disponible%20y%20la%20riqueza.&text=El%20c>

[onsumo%20es%20el%20gasto,considera%20un%20gasto%20de%20inversi%C3%B3n.](#)

IBM. (2019). *IBM SPSS – Valores perdidos 26*. Disponible en: [https://www.ibm.com/docs/en/SSLVMB\\_26.0.0/pdf/es/IBM\\_SPSS\\_Missing\\_Values.pdf](https://www.ibm.com/docs/en/SSLVMB_26.0.0/pdf/es/IBM_SPSS_Missing_Values.pdf)

IBM. (2021). *Análisis factorial explicativo: Rotación*. IBM SPSS Statistics. Disponible en: <https://www.ibm.com/docs/es/spss-statistics/beta?topic=analysis-exploratory-factor-rotation>

Instituto Nacional de Estadística (INE). (2022). *Contabilidad Nacional (PIB)*. Sección prensa. Disponible en: [https://www.ine.es/prensa/pib\\_prensa.htm](https://www.ine.es/prensa/pib_prensa.htm)

KPMG en España. (noviembre de 2020). *Sexto y último sondeo: Consumidores y nueva realidad: 4 tendencias clave*. KPMG. Disponible en: <https://home.kpmg/es/es/home/tendencias/2020/07/consumidores-nueva-realidad.html>

Lafortune, G., Fuller, G., Moreno, J., Schmidt-Traub, G. & Kroll, C. (2018). *SDG Index and Dashboards. Detailed Methodological paper*. Global responsibilities: implementing the goals. Disponible en <https://www.sdgindex.org/reports/sdg-index-and-dashboards-2018/>

MathWorks España - MATLAB. (2022). *Fill missing values*. Disponible en: <https://www.mathworks.com/help/matlab/ref/fillmissing.html>

Medina, F. & Galván, M. (2007). *Imputación de datos: teoría y práctica*. CEPAL – Serie de Estudios estadísticos y prospectivos n°54. Disponible en: [https://repositorio.cepal.org/bitstream/handle/11362/4755/1/S0700590\\_es.pdf](https://repositorio.cepal.org/bitstream/handle/11362/4755/1/S0700590_es.pdf)

Morettini, M. (2022). *Principales teorías macroeconómicas sobre el consumo*. Universidad Nacional de Mar de Plata. Disponible en: <http://nulan.mdp.edu.ar/1887/1/01486.pdf>

Muñoz Rosas, J. F., Álvarez Verdejo, E. (2009). *Métodos de imputación para el tratamiento de datos faltantes: aplicación mediante R/Splus*. Revista de Métodos Cuantitativos para la Economía y la Empresa. Disponible en: <https://www.redalyc.org/articulo.oa?id=233117228001>

- OECD. (2008). *Handbook on Constructing Composite Indicators: Methodology and User Guide*. Publicación de la OECD. Disponible en: <https://www.oecd.org/sdd/42495745.pdf>
- Otero García, D. (2011). *Imputación de datos faltantes en un Sistema de Información sobre Conductas de Riesgo*. Disponible en: [http://eio.usc.es/pub/mte/descargas/ProyectosFinMaster/Proyecto\\_616.pdf](http://eio.usc.es/pub/mte/descargas/ProyectosFinMaster/Proyecto_616.pdf)
- Pretel, E.A., Maza, C., Brandoli, J. & Barragán, C. (2020) *¿Cuándo llegó el coronavirus a Europa? Una pregunta geopolítica para cambiar la historia*. El Confidencial. Disponible en: [https://www.elconfidencial.com/mundo/europa/2020-05-05/coronavirus-europa-llegada-cronologia-china\\_2579515/#:~:text=En%20realidad%2C%20recuerda%20el%20experto,Mallorca%20el%2010%20de%20febrero.](https://www.elconfidencial.com/mundo/europa/2020-05-05/coronavirus-europa-llegada-cronologia-china_2579515/#:~:text=En%20realidad%2C%20recuerda%20el%20experto,Mallorca%20el%2010%20de%20febrero.)
- Primorac, C.R., La Red Martínez, D. L., Giovannini, M.E. (2020). *Metodología de evaluación del desempeño de métodos de imputación mediante una métrica tradicional complementada con un nuevo indicador*. European Scientific Journal. Disponible en: <https://eujournal.org/index.php/esj/article/view/13062>
- PwC. (junio 2021). *El consumidor ha cambiado para siempre: Global Consumer Insights Pulse Survey 2021*. PwC. Disponible en: <https://www.pwc.es/es/sala-prensa/notas-prensa/2021/pandemia-acelera-cambio-historico.html>
- Saisana, M., Tarantola, S. (2002). *State-of-the-art report on current methodologies and practices for composite indicator development*. Joint Research Centre of the European Commission.
- Saltelli, A. (2007). *Composite indicators between analysis and advocacy*. Social Indicators Research.
- Sandoval, W. (2021). *Imputación. Manejo de datos perdidos en series temporales*. R Pubs by RStudio. Disponible en: <https://rpubs.com/wilsonsr/780105>
- Schuschny, A.R., Soto de la Rosa, H. (2009). *Guía metodológica: Diseño de indicadores compuestos de desarrollo sostenible*. Disponible en: <https://repositorio.cepal.org/handle/11362/3661>

Gobierno de España: Ministerio de Asuntos Económicos y Transformación Digital.

*Síntesis de indicadores económicos.* Disponible en: [SIE: Tablas \(mineco.gob.es\)](https://sistema.inec.mineco.gob.es/)