

UNIVERSIDAD INTERNACIONAL DE LA RIOJA

PROGRAMA DE DOCTORADO EN ECONOMÍA Y EMPRESA

TESIS DOCTORAL

Aprendizaje automático (ML) en la economía del tabaco: el caso español.

Memoria presentada por

Miguel Ángel Del Arco Osuna
para optar al grado de Doctor
por la Universidad Internacional de La Rioja

Dirigida por los Doctores:

Juan Manuel Martín Álvarez
y
Alejandro Almeida Márquez

Logroño, 2024

Dedicatoria

A mi mujer Estefanía, mi hijo Miguel y a mis padres, que con su apoyo constante me han inspirado a seguir adelante, incluso en los momentos más bajos.

Y a todos aquellos que me han acompañado de una u otra manera en este proceso. Con su aliento, palabras de ánimo y confianza en mis capacidades, han sido un impulso fundamental en este camino.

Agradecimientos

Deseo expresar mi más sincero agradecimiento a mi Tutor, Codirector de Tesis y Amigo, Juan Manuel Martín Álvarez, por su orientación y apoyo a lo largo de este recorrido.

Su enorme profesionalidad, vastos conocimientos y gran paciencia han sido fundamentales para la realización de este trabajo. Gracias por ser un referente académico y profesional en todo lo que haces y por tu compromiso con mi crecimiento personal y profesional. Sin duda, este logro sin ti, sería imposible Amigo.

Resumen

Esta tesis analiza el impacto del consumo de tabaco en la salud pública y la economía, evaluando la efectividad de las políticas antitabaco, el uso de herramientas tecnológicas avanzadas como el aprendizaje automático (ML), y el efecto del desarrollo económico y la pandemia de COVID-19 en el mercado del tabaco en España. A través de un análisis exhaustivo, se concluye que las políticas de control del tabaco, como las leyes de 2005 y 2010, han tenido un efecto positivo en la reducción de las ventas de cigarrillos, aunque no han logrado una convergencia total en el mercado. El aprendizaje automático demostró ser una herramienta útil para predecir el impacto de la pandemia en las ventas de tabaco, revelando que el cierre de fronteras fue un factor clave en la disminución de las ventas en algunas regiones. Asimismo, el estudio revela una relación en forma de U invertida entre el desarrollo económico y el consumo de cigarrillos, mientras que productos alternativos como el tabaco para liar y el de pipa han funcionado como "productos refugio" durante las recesiones. Las conclusiones sugieren que las políticas deben adaptarse a las particularidades regionales y abarcar todos los productos de tabaco, con un enfoque en fortalecer las restricciones publicitarias y en monitorear los productos alternativos.

Palabras clave: Consumo de tabaco, políticas antitabaco, salud pública, aprendizaje automático (ML), COVID-19, desarrollo económico, ventas de cigarrillos, productos de tabaco alternativos, convergencia del mercado, impacto económico.

Abstract

This study analyzes the impact of tobacco consumption on public health and the economy, evaluating the effectiveness of anti-tobacco policies, the use of advanced technological tools such as machine learning (ML), and the effects of economic development and the COVID-19 pandemic on the tobacco market in Spain. Through a comprehensive analysis, it concludes that tobacco control policies, such as the 2005 and 2010 laws, have positively impacted cigarette sales reduction, though they have not achieved full market convergence. Machine learning proved to be a useful tool for predicting the pandemic's impact on tobacco sales, revealing that border closures were a key factor in sales declines in certain regions. Additionally, the study shows an inverted U-shaped relationship between economic development and cigarette consumption, while alternative products like roll-your-own (RYO) cigarettes and pipes served as "shelter products" during recessions. The conclusions suggest that policies should be adapted to regional particularities and encompass all tobacco products, with a focus on strengthening advertising restrictions and monitoring alternative products.

Keywords: Tobacco consumption, anti-tobacco policies, public health, machine learning (ML), COVID-19, economic development, cigarette sales, alternative tobacco products, market convergence, economic impact.

Índice

1. INTRODUCCIÓN Y OBJETIVOS	1
1.1 Introducción	1
1.2 Objetivos	4
1.3 Producción científica derivada de la tesis doctoral	4
2. UTILIZACIÓN DE LOS MODELOS ESTADÍSTICOS DE APRENDIZAJE AUTOMÁTICO ARIMA Y SARIMA PARA MEDIR EL IMPACTO DE LA COVID-19 EN LAS VENTAS PROVINCIALES DE CIGARRILLOS EN ESPAÑA	7
2.1 Introducción	7
2.2 Datos y metodología	10
2.3 Resultados	19
2.4 Conclusiones.....	26
3. CONVERGENCIA EN LAS VENTAS OFICIALES DE CIGARRILLOS EN ESPAÑA: UN ANÁLISIS DE CLÚSTERES DE 2005 A 2021	29
3.1 Introducción	29
3.2 Datos y metodología	31
3.3 Resultados	35
3.4 Conclusiones.....	40
4. EVOLUCIÓN SOCIOECONÓMICA Y VENTAS DE PRODUCTOS DEL TABACO EN ESPAÑA: UN ANÁLISIS A LARGO PLAZO DE LA CURVA DE KUZNETS DEL TABACO (TKC)	43
4.1 Introducción	43
4.2 Datos y Metodología	47
4.3 Resultados	51
4.4 Conclusiones.....	54
5. CONCLUSIONES INTEGRADORAS	57
6. REFERENCIAS	62

Capítulo 1

Introducción y objetivos

1.1 Introducción

El estudio del mercado del tabaco ha sido históricamente una cuestión clave debido a su gran impacto tanto en la salud pública como en las economías de los países. Existen dos razones estratégicas que explican la importancia de analizar el consumo de tabaco a lo largo del tiempo. En primer lugar, el tabaco es una sustancia adictiva que causa enfermedades graves como cánceres, afecciones cardiovasculares y respiratorias, entre otros problemas que provocan miles de muertes anualmente. Esto genera una carga económica considerable para los sistemas de salud. De hecho, se estima que los costes en salud relacionados con el consumo de tabaco representan un promedio del 11,5% del PIB en algunos países (Papanicolas et al., 2018). En segundo lugar, el tabaco es una fuente significativa de ingresos para los gobiernos a través de los impuestos especiales, siendo en España el producto que más contribuye a la recaudación fiscal de dichos impuestos.

Para combatir los efectos nocivos del tabaco, los gobiernos han implementado diversas políticas de control. España ha sido un ejemplo de éxito en la aplicación de leyes restrictivas para regular el consumo de tabaco, con medidas que afectan tanto a la oferta como a la demanda. Las leyes antitabaco de 2005 y 2010 introdujeron prohibiciones sobre la publicidad, promoción y patrocinio del tabaco, buscando reducir su atractivo, especialmente para los jóvenes. Estas leyes afectaron directamente a las

CAPÍTULO 1: INTRODUCCIÓN Y OBJETIVOS

estrategias comerciales de los fabricantes y minoristas, quienes se vieron obligados a mantener los precios sin recurrir a tácticas de promoción agresivas (Escario & Molina, 2004; Pinilla & Abásolo, 2017).

Diversos estudios han analizado los efectos de estas leyes en el mercado español del tabaco, concluyendo que tuvieron un impacto significativo en la reducción de las ventas y en la prevalencia de fumadores (Pinilla et al., 2019; Pons-Vigués et al., 2019). Además, se ha investigado el cumplimiento de las prohibiciones sobre publicidad, concluyendo que, aunque hubo desafíos en su implementación, las políticas tuvieron un efecto positivo en la reducción del consumo (Almeida et al., 2021).

El avance de la tecnología ha permitido utilizar herramientas analíticas más sofisticadas para investigar el mercado del tabaco. Una de las más destacadas es el aprendizaje automático (ML, por sus siglas en inglés). El ML es una rama de la inteligencia artificial que permite a las computadoras aprender de los datos sin ser explícitamente programadas (Murphy et al., 2012), y su uso se ha expandido rápidamente en la investigación del tabaco. Los enfoques más comunes dentro del ML son el aprendizaje supervisado y el no supervisado. Ambos ofrecen maneras únicas de abordar problemas relacionados con el consumo de tabaco.

El aprendizaje supervisado se utiliza principalmente para la predicción de fenómenos, como el riesgo de que un adolescente empiece a fumar, mientras que el aprendizaje no supervisado se centra en la identificación de patrones ocultos dentro de los datos, como la distribución de dependencia a la nicotina o el análisis de conversaciones en redes sociales sobre el tabaco (Fu et al., 2021).

Aunque el ML se ha empleado en diversas investigaciones sobre el tabaco, hasta ahora no se ha utilizado ampliamente para analizar la relación entre el tabaco y la pandemia de COVID-19. Durante la pandemia, se observaron cambios en el consumo de tabaco, con estudios que reportan tanto aumentos como disminuciones en el consumo (Saloojee & Mathee, 2021; Lee et al., 2021). Estas variaciones reflejan que la pandemia no afectó de manera uniforme a los hábitos de consumo, sugiriendo la necesidad de un análisis más profundo utilizando herramientas avanzadas como el ML para comprender mejor estas dinámicas.

La pandemia de COVID-19 representó un momento clave para el control del tabaquismo, afectando tanto la oferta como la demanda de productos de tabaco. Durante los confinamientos, las ventas de tabaco experimentaron comportamientos divergentes según diferentes estudios. Mientras algunos informaron de una disminución en el consumo debido a la prohibición de sitios públicos y al cierre de fronteras (Saloojee & Mathee, 2021), otros sugirieron que las ventas aumentaron, posiblemente debido al estrés y la ansiedad que generó la pandemia (Yach, 2020).

En España, aunque la aplicación del ML para analizar el impacto de la pandemia en las ventas de tabaco aún no ha sido ampliamente utilizada, diversos estudios indican que el confinamiento por COVID-19 tuvo un impacto en la reducción del consumo. Por ejemplo, estudios recientes muestran que la tasa de éxito en dejar de fumar aumentó

significativamente durante este periodo (Esplá et al., 2021). Sin embargo, otros trabajos sugieren que no hubo un efecto significativo de la pandemia en el consumo de tabaco, lo que refleja que el impacto puede depender de factores demográficos y económicos específicos (Estévez-Danta et al., 2022; Martínez-Cao et al., 2021).

El desarrollo económico es otro factor clave que influye en los patrones de consumo de tabaco. A medida que las economías crecen, se observa un aumento inicial en el consumo de bienes nocivos como el tabaco, debido a un mayor ingreso disponible y una menor conciencia sobre la salud pública. Sin embargo, en las fases más avanzadas de desarrollo, la educación y la concienciación sobre la salud juegan un papel crucial en la reducción del consumo (Cantarero-Prieto et al., 2019; Cantarero-Prieto et al., 2023).

Este fenómeno se ha estudiado bajo el marco de la Curva de Kuznets, que sugiere que el crecimiento económico inicialmente provoca un aumento en la desigualdad y el consumo de bienes nocivos, pero una vez que las economías alcanzan cierto nivel de desarrollo, estos efectos tienden a disminuir (Congregado et al., 2016; Iglesias et al., 2013; Robalino-López et al., 2015). En el caso del tabaco, se ha observado una relación similar, con un aumento en las ventas durante las primeras fases de desarrollo económico, seguido de una disminución cuando la conciencia socioeconómica y las políticas de salud pública se fortalecen (Martín-Álvarez et al., 2020a).

En el caso de España, el análisis de las ventas de cigarrillos a nivel provincial sugiere que las políticas de control del tabaco, como las leyes de 2005 y 2010, han tenido un impacto significativo en la reducción del consumo (Pinilla & Abásolo, 2017). Sin embargo, también se ha observado que la elasticidad de las ventas de tabaco ante cambios en los ingresos y el desempleo es considerablemente alta, especialmente durante las fases de recesión económica (Martín-Álvarez et al., 2020b). Esto indica que, aunque las políticas de control del tabaco son efectivas, los ciclos económicos también juegan un papel crucial en la determinación del consumo.

El estudio del mercado del tabaco requiere un enfoque multidimensional que tenga en cuenta factores de salud pública, desarrollo económico y la aplicación de herramientas analíticas avanzadas como el aprendizaje automático. Las políticas de control del tabaco en España han sido efectivas para reducir el consumo, pero el impacto de la pandemia de COVID-19 y las fluctuaciones económicas han mostrado que el consumo de tabaco sigue siendo un problema complejo. El uso de enfoques como la Curva de Kuznets y el aprendizaje automático ofrece nuevas oportunidades para analizar este problema desde diferentes perspectivas, proporcionando información valiosa para diseñar políticas más efectivas.

Es crucial que los futuros estudios sobre el mercado del tabaco incorporen el análisis simultáneo de múltiples factores socioeconómicos y utilicen herramientas tecnológicas avanzadas para lograr una comprensión más profunda de cómo se pueden reducir de manera efectiva las ventas de productos de tabaco. Con la continua evolución de la economía y la salud pública, se espera que nuevas estrategias, combinadas con políticas de control del tabaco, puedan lograr una reducción significativa en el consumo a largo plazo.

1.2 Objetivos

El objetivo fundamental de la tesis es analizar el impacto del consumo de tabaco en la salud pública y la economía, evaluando la efectividad de las políticas antitabaco, el uso de herramientas tecnológicas avanzadas como el aprendizaje automático, y el efecto del desarrollo económico y la pandemia de COVID-19 sobre el mercado del tabaco, con el fin de ofrecer recomendaciones para futuras investigaciones y políticas públicas más eficaces. En base a ese objetivo principal, se proponen los siguientes objetivos parciales:

- Analizar el impacto del tabaco en la salud pública y la economía
- Evaluar la efectividad de las políticas antitabaco.
- Explorar el uso de tecnologías avanzadas como el aprendizaje automático (ML) en la investigación del tabaco.
- Estudiar el impacto de la pandemia de COVID-19 en el consumo de tabaco.
- Investigar cómo los ciclos económicos, y otros factores socioeconómicos, afectan el consumo de tabaco en sus distintas formas.
- Ofrecer recomendaciones para futuras investigaciones y políticas.

1.3 Producción científica derivada de la tesis doctoral

En el desarrollo de esta tesis doctoral, los capítulos 2 y 3 han generado resultados científicos que se han concretado en dos publicaciones en revistas académicas especializadas, contribuyendo a los campos del aprendizaje automático (ML) aplicado a la economía del tabaco. El Capítulo 2 derivó en el artículo titulado "Using the Statistical Machine Learning Models ARIMA and SARIMA to Measure the Impact of COVID-19 on Official Provincial Sales of Cigarettes in Spain", publicado en la revista *International Journal of Interactive Multimedia and Artificial Intelligence* (<https://doi.org/10.9781/ijimai.2023.02.010>). En este trabajo, se aplican modelos de aprendizaje automático (ARIMA y SARIMA) para predecir el impacto de la pandemia de COVID-19 en las ventas de cigarrillos en las provincias españolas, abarcando un análisis profundo de las series temporales y sus implicaciones para la salud pública. La metodología desarrollada ha permitido identificar las variaciones en las ventas provocadas por las restricciones de movilidad y otras medidas relacionadas con la pandemia, resaltando el valor del uso de herramientas avanzadas de análisis estadístico para abordar problemas complejos en el campo de la salud y el control del tabaco.

1.3 Producción científica derivada de la tesis doctoral

Por otro lado, el Capítulo 3 dio lugar al artículo titulado "Impact of the Spanish Smoke-free Laws on Cigarette Sales by Brands, 2000–2021: Evidence from a Club Convergence Approach", publicado en la revista *Tobacco Induced Diseases* (<https://doi.org/10.18332/tid/174407>). Este estudio aborda el efecto de las leyes españolas de control del tabaco, implementadas en 2005 y 2010, sobre las ventas de cigarrillos por marcas, utilizando un enfoque de clubes de convergencia para evaluar la homogenización de las ventas. El análisis revela que la Ley 42/2010, que introdujo restricciones más estrictas sobre la publicidad, promoción y patrocinio del tabaco, generó una mayor homogenización en el comportamiento de las ventas de marcas, comparado con las disposiciones anteriores. Estos resultados proporcionan una perspectiva valiosa sobre la efectividad de las políticas de control del tabaco y su capacidad para reducir el atractivo del consumo, lo que refuerza la importancia de políticas más estrictas en la lucha contra el tabaquismo.

CAPÍTULO 1: INTRODUCCIÓN Y OBJETIVOS

Capítulo 2

Utilización de los modelos estadísticos de aprendizaje automático ARIMA y SARIMA para medir el impacto de la COVID-19 en las ventas provinciales de cigarrillos en España

2.1 Introducción

Hay dos razones estratégicas por las que el desarrollo del uso y comportamiento del tabaco en cualquier nación a través del tiempo es un tema pertinente. En primer lugar, fumar es adictivo por naturaleza y causa muchos tipos diferentes de cáncer, afecciones cardiovasculares y respiratorias, problemas reproductivos y otros impactos nocivos para la salud que provocan miles de muertes cada año. Como resultado, el sistema de salud soporta costes significativos relacionados con el daño causado por el consumo de tabaco: en promedio, el gasto en salud representa el 11,5% del PIB del país (Papanicolas et al., 2018). En segundo lugar, los presupuestos de los países de altos ingresos se ven

CAPÍTULO 2: Utilización de los modelos estadísticos de aprendizaje automático ARIMA y SARIMA para medir el impacto de la COVID-19 en las ventas provinciales de cigarrillos en España

significativamente afectados por los impuestos especiales recaudados sobre el tabaco y, en España, el tabaco es el producto que más aporta a la recaudación de impuestos especiales.

Además, un estudio reciente centrado en el mercado español demuestra que algunas provincias no cuentan con datos oficiales de ventas precisos que puedan utilizarse para evaluar las medidas de control del tabaquismo (Cadahia et al., 2021). En este sentido, las provincias fronterizas y turísticas son aquellas en las que las ventas se ven más afectadas, lo que respalda la prevalencia del comercio ilegal y transacciones transfronterizas sustanciales. Así, los hallazgos de este estudio demuestran la eficacia de políticas compartidas adoptadas por los gobiernos de naciones vecinas que preservan una pequeña diferencia de precios entre ellos.

Este escenario exige herramientas analíticas avanzadas para manejar grandes volúmenes de datos y llevar a cabo actividades altamente especializadas para erradicar los problemas de salud relacionados con el tabaco y los entornos cada vez más complicados de la investigación del tabaco. Debido a esto, algunas investigaciones ya han utilizado métodos de aprendizaje automático (en adelante, ML) para analizar datos relacionados con el mercado del tabaco (Fu et al, 2021). La definición de aprendizaje automático (ML) históricamente se ha descrito como "una rama de la investigación que ofrece a las computadoras la capacidad de aprender sin ser programadas explícitamente" para pronosticar datos futuros o tomar decisiones en situaciones inciertas (Murphy et al, 2012). El objetivo principal del ML es emplear "fuerza bruta" en lugar de supervisión humana al analizar datos. Debido a que el ML requiere mucha menos supervisión humana, puede considerarse como una extensión natural de las metodologías estadísticas convencionales (Beam & Kohane, 2018). El aprendizaje no supervisado y el aprendizaje supervisado son categorías que se encuentran dentro del aprendizaje automático. Cada uno de los dos conjuntos de enfoques de aprendizaje automático tiene cualidades distintivas que pueden ser de interés para los investigadores que estudian el tabaco. Cada uno de ellos está orientado a resolver una determinada dificultad. El enfoque del aprendizaje supervisado es la predicción. Para predecir los valores de una o más variables de salida o de respuesta para un conjunto específico de variables de entrada o predictoras, se debe entrenar y validar un modelo (Hastie et al., 2009). En este sentido, las técnicas de aprendizaje supervisado se utilizan cuando el objetivo es crear un modelo predictivo de alta precisión para datos futuros. Cualquier investigación del mercado del tabaco que requiera pronósticos extremadamente precisos, como la creación de un programa de vigilancia de la salud pública que prediga automáticamente la probabilidad de que los adolescentes comiencen a fumar, es útil para el aprendizaje supervisado (Fu et al., 2021). El aprendizaje no supervisado, por otro lado, no requiere una variable de salida porque su objetivo es determinar la distribución de probabilidad subyacente de los datos (también conocida como estimación de densidad) (Fu et al., 2021). Examinar las discusiones en las redes sociales relacionadas con el tabaco y la identificación de probables subtipos de dependencia de la nicotina mediante el examen de datos de resonancia magnética cerebral del paciente son dos ejemplos de aprendizaje no supervisado en la investigación del tabaco (Fu et al., 2021).

Como se ha indicado, el aprendizaje automático es una herramienta analítica muy poderosa para los investigadores del mercado del tabaco; los enfoques se pueden dividir en términos generales en aprendizaje supervisado y no supervisado. Sin embargo, además de esta clasificación de técnicas, los estudios que aplican el ML al análisis del mercado del tabaco también pueden clasificarse según los datos (entradas) utilizados. En este sentido, podemos encontrar estudios que analizan contenidos en redes sociales, textos de informes clínicos o datos administrativos. De hecho, varios artículos publicados que analizan el mercado del tabaco se centran en datos administrativos del análisis de (Reps et al., 2019; Mamoshina et al., 2019). Muchos de estos estudios aplican técnicas de aprendizaje supervisado para predecir un fenómeno binario relacionado con dejar de fumar, incluida la intención de dejar de fumar (Huda et al., 2010), la adherencia a las terapias para dejar de fumar (Kim et al., 2019) y el anhelo de fumar al máximo o al mínimo durante un intento de dejar de fumar (Dumortier et al., 2016). Sin embargo, pocos estudios han aplicado técnicas de aprendizaje supervisado con el objetivo de predecir variables continuas utilizando, por ejemplo, regresión o random forest (Fu et al., 2011; Coughlin et al., 2020; Davagdorj et al., 2019; Singh & Katyan, 2019).

Aunque el ML se ha aplicado al análisis de temas relacionados con el tabaco, hasta donde sabemos, el ML nunca se ha aplicado para estudiar la relación entre COVID-19 y el tabaco. La pandemia de COVID-19 ha planteado una oportunidad única para combatir el consumo de tabaco (Clancy et al., 2020). El consumo de tabaco y las prohibiciones de sitios, los cierres de fronteras y los confinamientos han tenido impactos tanto positivos como negativos en el control del tabaco. Un estudio reciente concluye que el consumo de cigarrillos disminuyó durante el confinamiento por la COVID-19 en 2020 (Saloojee & Mathee, 2021). Sin embargo, otros artículos concluyen lo contrario. En concreto, uno de los trabajos recientes concluye que la pandemia generó un aumento del 13% en las ventas de tabaco (Lee et al., 2021). Otro artículo indica que este aumento se debe a que los consumidores de nicotina utilizan el tabaco como principal mecanismo para afrontar el estrés y la ansiedad (Yach, 2020). Además, un artículo indica que la pandemia de COVID-19 está relacionada con mayores ventas de tabaco y sugiere investigar si los hábitos de fumar han cambiado desde los cierres pandémicos (Driezen et al., 2022). Respecto al uso del análisis de series temporales para analizar cambios en las ventas de cigarrillos, solo existe un estudio que aborda este problema y concluye que las ventas observadas durante la pandemia son superiores a las esperadas (Asare et al., 2022). A continuación, en relación con el aumento de las ventas de tabaco, otro estudio sugiere que la intención de dejar de fumar ha experimentado una disminución posterior a la pandemia de COVID-19 (Kim et al., 2022). Finalmente, otros trabajos que analizan el tabaquismo y el COVID-19 sugieren que la venta de tabaco debería haber sido prohibida durante la pandemia dada la gran oportunidad que presentó el COVID-19 para erradicar el tabaquismo (Ahluwalia et al., 2020; Hefler & Gartner, 2020; Burki 2021).

En España, aunque no existen trabajos en los que se aplique ML al mercado del tabaco para explicar el impacto del COVID-19 en las ventas de tabaco, sí hay trabajos que han analizado la influencia del COVID-19 en diferentes aspectos relacionados con el tabaco desde otra perspectiva. Alguna literatura indica que durante el confinamiento

CAPÍTULO 2: Utilización de los modelos estadísticos de aprendizaje automático ARIMA y SARIMA para medir el impacto de la COVID-19 en las ventas provinciales de cigarrillos en España

por COVID-19 en España, el consumo de tabaco disminuyó (Álvarez et al., 2021). En esta misma línea de menor prevalencia, otro trabajo indica que la tasa de éxito para dejar de fumar pasó del 25% al 35% (Esplá et al., 2021). Otro trabajo, que se centra en analizar la percepción de los fumadores sobre su exposición al virus, sugiere que muchos fumadores pueden haber cambiado sus patrones de tabaquismo y es posible que los que redujeron su consumo de tabaco superaron a los que aumentaron su consumo (Suelves et al., 2021). Otro estudio que analiza el impacto de la COVID-19 en el consumo de tabaco sugiere que en España no se observa ningún efecto significativo de la pandemia sobre el consumo de tabaco (Estévez-Danta et al., 2022). Finalmente, hay un grupo de trabajos que indican que el impacto que ha tenido el COVID-19 en el consumo de tabaco depende de cuestiones demográficas personales y que no todas las personas actuaron igual (Martínez-Cao et al., 2021). Además, este bloque incluye obras que advierten de la urgente necesidad de que los consumidores de tabaco dejen de fumar debido a los daños que provoca en la salud de los consumidores este nocivo producto (Vázquez & Redolar-Ripoll, 2020a; Vázquez & Redolar-Ripoll, 2020b).

Hasta donde sabemos, aún no se ha realizado ningún estudio sobre los efectos regionales que el COVID-19 ha tenido en el mercado tabacalero español. En este estudio, intentamos predecir cómo habría sido el mercado provincial del tabaco en ausencia de la pandemia de COVID-19. Luego, cuantificamos el impacto de la pandemia en las ventas de cigarrillos como la diferencia entre el pronóstico y los datos reales. Los datos utilizados en el presente estudio provienen del Ministerio de Hacienda y Función Pública y cubren el período comprendido entre enero de 2005 y diciembre de 2021 en términos de ventas de cigarrillos. El resto del capítulo está estructurado de la siguiente manera: una sección en la que proporciona una descripción de los datos y modelos estadísticos empleados, junto con información sobre las matemáticas que los sustentan, herramientas analíticas y medidas de evaluación. En otra sección se muestra los resultados análisis de series temporales para explorar en profundidad el impacto provincial de COVID-19 en el mercado del tabaco. Finalmente, se aportan las conclusiones a las que se llegó a partir de esta investigación.

2.2 Datos y metodología

Para lograr el objetivo marcado en este trabajo, generamos una estimación de las ventas de cigarrillos para las 48 provincias españolas desde enero de 2020 hasta diciembre de 2021 utilizando los modelos estadísticos ML ARIMA y SARIMA. Los modelos sugeridos se han optimizado eligiendo los parámetros más adecuados para cada provincia. Para garantizar que la serie temporal tenga la misma duración en todas las provincias, utilizamos el 1 de enero de 2005 como fecha de inicio para cada provincia. De acuerdo con la literatura previa, se necesita un tamaño de muestra mínimo de 30 observaciones para proporcionar un pronóstico estadísticamente significativo de los datos de series temporales (Yaffee & McGee, 2000). Dado que el

modelo de cada provincia se entrenó utilizando datos de enero de 2005 a diciembre de 2017 (168 observaciones), el tamaño de la muestra para estimar las ventas de cigarrillos es significativamente mayor que el umbral establecido en la literatura previa.

Para construir nuestra investigación empírica se utilizó un panel de datos mensuales de las provincias españolas desde enero de 2005 hasta diciembre de 2021. La sección de estadísticas de la página web del Comisionado para el Mercado de Tabacos (órgano dependiente del Ministerio de Hacienda y Función Pública) proporciona los datos de ventas de cigarrillos en euros y unidades. Además, se ha utilizado el repositorio del Instituto Nacional de Estadística para recopilar datos sobre la población mayor de 18 años con el fin de estimar las ventas provinciales per cápita.

Las series temporales son colecciones de valores numéricos, cada uno de los cuales tiene un componente periódico. Las series temporales se pueden dividir en dos grupos: series temporales estacionarias y series temporales no estacionarias, dependiendo de cómo se comporten los valores numéricos de la serie temporal. Las series temporales no estacionarias tienen patrones que impiden que la media y/o la varianza sean constantes, mientras que las series estacionarias no exhiben patrones en su media y/o varianza con respecto al tiempo. La estacionalidad o la tendencia pueden ser las culpables de que una serie temporal sea no estacionaria. Calcular la diferencia entre dos observaciones sucesivas puede hacer que las series temporales no estacionarias sean estacionarias. La tendencia y la estacionalidad se eliminan de la serie temporal mediante el método de diferenciación. La diferenciación de primer y segundo orden son los dos procedimientos que se emplean con mayor frecuencia; sus procesos de cálculo se describen en las ecuaciones (1) y (2):

$$y_t^{\prime} = y_t - y_{t-1} \quad (1)$$

$$y_t^{\prime\prime} = y_t - 2y_{t-1} + y_{t-2} \quad (2)$$

donde y_t son datos de series temporales no estacionarias, y_t^{\prime} es la serie después de la diferenciación de primer orden, $y_t^{\prime\prime}$ es la serie después de la diferenciación de segundo orden, y_{t-1} es la observación de la serie temporal en el período t-1, y_{t-2} es la observación de la serie temporal en el período t-2. Sólo cuando la serie temporal no es estacionaria después de la diferenciación de primer orden se requiere la diferenciación de segundo orden. También existe la opción de distinción estacional. En este caso, la distancia entre una observación y la observación idéntica del año anterior se utiliza para calcular la diferencia (o período). Esta ecuación proporciona una definición para el primer grado de diferenciación estacional (3).

$$y_s^{\prime} = y_t - y_{t-m} \quad (3)$$

donde y_s^{\prime} es la serie temporal después de la diferenciación estacional de primer orden, y_{t-m} es la observación del período t-m, m es el número de períodos que existen entre una observación y la misma en el período anterior. En este trabajo, las series

CAPÍTULO 2: Utilización de los modelos estadísticos de aprendizaje automático ARIMA y SARIMA para medir el impacto de la COVID-19 en las ventas provinciales de cigarrillos en España

temporales fueron sometidas a diferenciación para eliminar la estacionalidad y el conjunto de datos resultante es el que se utiliza para realizar las estimaciones. Además, se debe tener en cuenta que la estimación de los parámetros de los modelos ARIMA y SARIMA se realiza asumiendo 4 supuestos básicos: (i) las series temporales no contienen puntos atípicos, (ii) las series temporales están compuestas de una única variable que es la que, con sus valores pasados, ayuda a realizar las predicciones; (iii) las series temporales son estacionarias, (iv) los parámetros y errores del modelo son constantes durante todo el período de tiempo.

Box y Jenkin crearon el modelo ARIMA (p, d, q) en 1976 (Box-George et al., 1976) que puede utilizarse para predecir series temporales estacionarias sin estacionalidad. Tres términos (p, d y q) definen este modelo ARIMA. El orden del término de media móvil (MA) es q, el orden del término de autorregresión (AR) es p y el orden de diferenciación necesario para mantener estacionaria la serie temporal es d. La regresión de la variable contra sí misma para pronosticar su comportamiento futuro se conoce como autorregresión. Se trata de comparar el valor observado en un determinado punto con los valores de periodos anteriores. MA es un modelo similar a una regresión que pronostica una variable en una etapa posterior utilizando los errores de pronóstico de una etapa anterior. Las ecuaciones generalizadas para el modelo AR de orden p y el modelo MA de orden q se dan a continuación (ecuaciones (4) y (5), respectivamente).

$$y_t = C + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + \varepsilon_t \quad (4)$$

$$y_t = C + \varepsilon_t + \theta_1 \varepsilon_{t-1} + \theta_2 \varepsilon_{t-2} + \dots + \theta_q \varepsilon_{t-q} \quad (5)$$

El modelo AR (ecuación (4)), la integración (I) y el modelo MA (ecuación (5)) se combinan para crear modelos ARIMA en este estudio. Para crear el pronóstico, la integración (I) utiliza la diferenciación a la inversa. La formulación matemática del modelo ARIMA generalizado es la ecuación (6).

$$y_t = C + \phi_1 y + \phi_p y_{t-p} + \dots + \phi_n y_{t-n} + \theta_1 \varepsilon_{t-1} + \theta_q \varepsilon_{t-q} + \varepsilon_t \quad (6)$$

Donde C es el término independiente, ϕ_i ($i = 1, 2, \dots, p$) son los parámetros del modelo autorregresivo, θ_i ($i = 1, 2, \dots, q$) son los parámetros del modelo de media móvil, y_t es la serie temporal actual, $y_{t-1}, y_{t-2}, \dots, y_{t-p}$ son valores pasados y ε_t es el error aleatorio del período t y viene dado por la siguiente ecuación:

$$\varepsilon_t = y_t - y_{t-1} \quad (7)$$

Para tener en cuenta la estacionalidad de la serie temporal, el modelo ARIMA estacional (SARIMA) combina el ARIMA no estacional (p, d y q) con términos estacionales adicionales (P, D y Q). El término AR estacional, el término de promedio móvil estacional y el término de diferenciación estacional están representados,

respectivamente, por los términos P, Q y D. El modelo general SARIMA se representa matemáticamente de la siguiente manera:

$$\Phi_p(B^m)\phi_p(B)(1 - B^m)^D(1 - B)^d y_t = \Theta_Q(B^m)\theta_q(B)w_t \quad (8)$$

Donde y_t es la serie temporal no estacionaria, w_t es el proceso de ruido blanco gaussiano, $\phi(B)$ es un polinomio autorregresivo no estacional y $\theta(B)$ es un polinomio de media móvil no estacional, D es la diferenciación estacional (el término es igual a 1 o 2, etc.). Cuando D toma el valor 1 es suficiente para imponer estacionariedad a los datos. $\Phi(B^m)$ es un polinomio autorregresivo estacional y $\Theta(B^m)$ es un polinomio de media móvil estacional. Donde B se define como el operador de retroceso el cual se expresa de la siguiente manera:

$$B^k y_t = y_{t-k} \quad (9)$$

A continuación, se proporcionan las expresiones para el modelo autorregresivo no estacional (Ec. 10), el modelo de media móvil (Ec. 11), el modelo AR estacional (Ec. 12) y el modelo MA estacional (Ec. 13).

$$\varphi(B) = 1 - \varphi_1 B - \varphi_2 B^2 - \dots - \varphi_p B^p \quad (10)$$

$$\theta(B) = 1 + \theta_1 B + \theta_2 B^2 + \dots + \theta_q B^q \quad (11)$$

$$\Phi_p(B^m) = 1 - \Phi_1 B^m - \Phi_2 B^{2m} + \dots + \Phi_p B^{pm} \quad (12)$$

$$\Theta_Q(B^m) = 1 + \Theta_1 B^m + \Theta_2 B^{2m} + \dots + \Theta_Q B^{Qm} \quad (13)$$

Los indicadores se utilizan para juzgar la precisión del análisis de series temporales una vez que se han estimado los parámetros de los modelos ARIMA y SARIMA y se han producido las predicciones. Estos indicadores incluyen la función de autocorrelación parcial (PACF), el criterio de información de Akaike (AIC), la función de autocorrelación (ACF) y los criterios de información bayesianos (BIC). Estas métricas muestran cómo se relacionan entre sí las observaciones de las series temporales. Mientras que PACF correlaciona las series temporales con sus propios valores retrasados espaciados por unidades de tiempo específicas, ACF proporciona la correlación de los datos de las series temporales con sus datos de series anteriores. Los valores del criterio de probabilidad penalizada AIC y BIC están relacionados; cuanto más bajos sean, más probable será que el modelo sea aceptado como modelo genuino. Además, los criterios de evaluación de este estudio incluyen error medio (ME), error cuadrático medio (RMSE), error absoluto medio (MAE), error porcentual medio (MPE), error porcentual absoluto medio (MAPE) y error absoluto medio escalado (MASE).

CAPÍTULO 2: Utilización de los modelos estadísticos de aprendizaje automático ARIMA y SARIMA para medir el impacto de la COVID-19 en las ventas provinciales de cigarrillos en España

En una serie temporal, la autocorrelación es la relación entre la observación más reciente y las observaciones retardadas. El ACF describe la relación lineal entre la observación en el momento t y la observación en un momento anterior. El gráfico de autocorrelación es la representación de la serie temporal de la autocorrelación frente a los retrasos ($t-k$). A modo de ilustración, el ACF para la serie temporal y_t viene dado por:

$$ACF(y_t, y_{t-k}) = \frac{\text{Covariance}(y_t, y_{t-k})}{\text{variance}(y_t)} \quad (14)$$

donde k es el retraso y se define como la diferencia entre y_t e y_{t-k} . Por otro lado, en la autocorrelación parcial, las observaciones intermedias se consideran al calcular la correlación entre dos observaciones en momentos diferentes. Por ejemplo, considere que una serie temporal y_t , el PACF entre dos observaciones y_t e y_{t-2} (suponiendo $k = 2$) se puede escribir como se muestra en la ecuación.

$$PACF(y_t, y_{t-2}) = \frac{\text{Covariance}(y_t, y_{t-2} | y_{t-1})}{\sqrt{\text{variance}(y_t | y_{t-1})} \sqrt{\text{variance}(y_{t-2} | y_{t-1})}} \quad (15)$$

Es necesario probar los modelos creados para ver qué tal funcionan en términos de dilucidar las relaciones entre las variables. Hemos evaluado la capacidad de un modelo para explicar relaciones utilizando los criterios de información. AIC y BIC son dos medidas ampliamente utilizadas que evalúan la calidad de los modelos premiando a aquellos que tienen menos errores y penalizando a aquellos que tienen demasiados parámetros. Así es como se representa matemáticamente AIC:

$$AIC = -2\log L(\hat{\theta}) + 2K \quad (16)$$

Donde K es el número total de parámetros del modelo y $\log L(\theta)$ es la función de verosimilitud. BIC es un criterio de selección de modelo diferente, pero en una línea similar. En comparación con AIC, BIC impone una penalización menor en la cantidad de parámetros. El modelo con el valor de probabilidad más alto está representado por el valor más bajo tanto en la configuración AIC como en BIC. Como resultado, ayuda a los analistas de series temporales a seleccionar el modelo óptimo entre el número limitado de modelos alternativos generados. Así es como se representa matemáticamente BIC:

$$BIC = -2\log L(\hat{\theta}) + K \log N \quad (17)$$

Donde N es el número de observaciones.

MAE, RMSE, MAPE, MPE, ME y MASE se utilizan a menudo para evaluar la precisión de los modelos ML (Kasihmuddin et al., 2021; Gupta et al., 2021), que vienen dados por las siguientes ecuaciones:

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (18)$$

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (19)$$

$$MAPE = \frac{1}{n} \sum_{i=1}^n \left| \frac{y_i - \hat{y}_i}{y_i} \right| \quad (20)$$

$$MPE = \frac{1}{n} \sum_{i=1}^n \frac{y_i - \hat{y}_i}{y_i} \quad (21)$$

$$ME = \frac{1}{n} \sum_{i=1}^n y_i - \hat{y}_i \quad (22)$$

$$MASE = \frac{\frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|}{\frac{1}{T-1} \sum_{t=2}^T |y_t - y_{t-1}|} \quad (23)$$

Los scripts se crearon utilizando el lenguaje de programación R en el entorno RStudio, para lograr el objetivo mencionado en este capítulo (R Core Team, 2021; RStudio Team 2021). Las bibliotecas tidyverse y forecast también se han utilizado para limpiar los datos y hacer las estimaciones y representaciones gráficas (Wickham et al, 2019; Hyndman & Khandakar, 2008; Kyndman et al., 2022). Independientemente al proceso de limpieza de datos, en esta sección se detallan las acciones utilizadas por el algoritmo ML para lograr el objetivo indicado. Inicialmente, el algoritmo determina si cada serie temporal presenta no estacionariedad (si se hubiera hecho manualmente, esto se habría verificado mediante gráficos ACF y PACF). La serie temporal no es estacionaria si la autocorrelación disminuye sólo ligeramente a medida que aumenta el número de retrasos. A continuación, la técnica aplica diferencias antes de ejecutar el modelado ARIMA o SARIMA si hay evidencia de que la serie temporal no es estacionaria. Dependiendo de qué opción se ajuste mejor a la serie temporal, el algoritmo selecciona ARIMA o SARIMA. Dado el importante componente estacional presente en las series temporales de ventas de tabaco, en el estudio de caso de este artículo se utilizó el método SARIMA para todas las series. Los modelos SARIMA requieren un tiempo promedio de procesamiento de 7 segundos para completar cada simulación en la computadora local.

La selección manual del mejor parámetro (p, d, q) (P, D, Q) de los modelos ARIMA y SARIMA utilizando gráficos ACF y PACF puede llevar mucho tiempo, ya que los modelos han sido estimados para 48 provincias y 3 diferentes variables (euros, cajetillas totales y cajetillas per cápita). Para seleccionar la combinación adecuada de valores de parámetros del modelo, realizamos una búsqueda automática utilizando la librería forecast, como se indica en el párrafo anterior. Esta librería utiliza AIC como métrica de evaluación para elegir el mejor modelo entre varios modelos ARIMA y SARIMA. Dado

CAPÍTULO 2: Utilización de los modelos estadísticos de aprendizaje automático ARIMA y SARIMA para medir el impacto de la COVID-19 en las ventas provinciales de cigarrillos en España

que todas las series temporales utilizadas inician en enero de 2005, finalizan en diciembre de 2021 y las ventas de tabaco muestran una fuerte estacionalidad, el parámetro m tomó un valor de 12 en todos los casos.

Los datos de la serie temporal de las 48 provincias españolas se dividieron en dos partes: el conjunto de datos de entrenamiento seleccionado va de enero de 2005 a diciembre de 2017 y el conjunto de datos de validación va de enero de 2018 a diciembre de 2019. Utilizando el conjunto de datos de entrenamiento, se construye el modelo y el conjunto de datos de validación se utiliza para estimar el rendimiento del modelo. Se utilizaron las siguientes métricas de evaluación para evaluarlo: MAE, RMSE, MAPE, MPE, ME y MASE. El modelo entrenado se utilizó para pronosticar los valores de ventas de tabaco desde enero de 2020 hasta diciembre de 2021 (período en el que las ventas reales se ven alteradas por confinamientos, restricciones en la hostelería y cierre de fronteras en las 48 provincias españolas. Finalmente, para estimar el impacto que ha tenido la COVID-19 en las ventas de tabaco en España, se comparan las estimaciones realizadas por los modelos SARIMA con las ventas reales observadas desde enero de 2020 a diciembre de 2021.

Tabla 2.1. Modelos SARIMA seleccionados para la previsión en euros.

	SARIMA (p,d,q)(P,D,Q,m)	RMSE	MAE	MPE	MAPE	MASE
Alava	(1,1,4)(2,0,0,12)	4,84E+05	4,06E+05	-8,16E-01	7,60E+00	1,39E+00
Albacete	(2,1,2)(1,1,0,12)	7,41E+05	5,86E+05	-7,22E-01	7,63E+00	1,67E+00
Alicante	(1,0,2)(2,1,1,12)	5,40E+06	4,63E+06	-1,03E+00	1,05E+01	2,77E+00
Almería	(4,1,0)(2,1,2,12)	1,70E+06	1,37E+06	-1,11E+00	9,14E+00	3,16E+00
Asturias	(2,1,1)(2,1,2,12)	1,77E+06	1,39E+06	-7,72E-01	6,97E+00	2,14E+00
Ávila	(2,1,2)(1,1,0,12)	6,19E+05	4,91E+05	-3,01E+00	1,48E+01	2,81E+00
Badajoz	(2,1,2)(2,1,0,12)	1,12E+06	9,02E+05	-8,55E-01	7,44E+00	2,01E+00
Balears (Illes)	(3,0,1)(2,1,0,12)	4,06E+06	3,42E+06	-2,50E+00	1,22E+01	2,79E+00
Barcelona	(5,1,1)(2,0,0,12)	7,78E+06	6,42E+06	-6,72E-01	6,78E+00	1,81E+00
Burgos	(0,0,0)(2,0,0,12)	8,14E+05	6,53E+05	-1,44E+00	9,73E+00	2,94E+00
Cáceres	(2,1,2)(1,1,0,12)	9,77E+05	7,55E+05	-1,40E+00	9,32E+00	2,22E+00
Cádiz	(2,1,2)(2,1,2,12)	2,27E+06	1,76E+06	-1,88E+00	1,19E+01	1,83E+00
Cantabria	(2,1,2)(1,1,0,12)	1,40E+06	1,04E+06	-1,32E+00	8,78E+00	2,72E+00
Castellón	(2,1,2)(2,1,0,12)	1,73E+06	1,31E+06	-1,71E+00	1,07E+01	2,71E+00
Ciudad Real	(2,1,1)(2,1,0,12)	9,32E+05	7,02E+05	-9,58E-01	7,42E+00	2,58E+00
Córdoba	(5,1,3)(2,0,0,12)	9,79E+05	7,72E+05	-7,15E-01	6,68E+00	1,68E+00
Coruña (A)	(0,1,4)(2,0,0,12)	1,90E+06	1,53E+06	-8,53E-01	7,49E+00	2,52E+00
Cuenca	(2,1,2)(2,1,0,12)	5,92E+05	4,45E+05	-1,48E+00	1,00E+01	2,57E+00
Girona	(2,1,2)(2,1,0,12)	1,07E+07	8,59E+06	-6,38E+00	2,25E+01	2,39E+00
Granada	(2,1,2)(2,1,0,12)	1,59E+06	1,28E+06	-8,76E-01	7,50E+00	2,22E+00
Guadalajara	(2,1,2)(1,1,0,12)	4,46E+05	3,65E+05	-7,61E-01	7,84E+00	1,96E+00
Guipúzcoa	(2,1,2)(1,1,0,12)	2,62E+06	2,12E+06	-1,37E+00	1,07E+01	1,70E+00
Huelva	(2,1,2)(2,1,2,12)	1,31E+06	1,05E+06	-1,31E+00	1,04E+01	2,28E+00
Huesca	(2,1,2)(2,1,0,12)	8,19E+05	5,96E+05	-1,77E+00	1,08E+01	2,27E+00
Jaén	(2,1,2)(2,0,0,12)	8,11E+05	6,57E+05	-5,22E-01	5,86E+00	1,92E+00
León	(2,1,2)(1,1,0,12)	1,07E+06	8,18E+05	-1,37E+00	9,12E+00	2,54E+00
Lleida	(1,1,2)(1,0,0,12)	1,33E+06	1,16E+06	-1,48E+00	1,06E+01	1,18E+00
Lugo	(2,1,2)(1,1,0,12)	6,57E+05	5,12E+05	-1,14E+00	8,42E+00	2,03E+00
Madrid	(5,1,0)(2,0,0,12)	7,38E+06	6,14E+06	-4,82E-01	5,78E+00	1,61E+00
Málaga	(2,1,1)(2,1,2,12)	4,27E+06	3,71E+06	-1,34E+00	1,20E+01	2,76E+00
Murcia	(2,1,2)(2,1,2,12)	2,46E+06	2,02E+06	-6,98E-01	6,90E+00	1,76E+00
Navarra	(2,1,2)(2,1,0,12)	2,41E+06	1,93E+06	-1,41E+00	1,08E+01	2,72E+00
Ourense	(2,1,2)(1,1,0,12)	6,07E+05	4,55E+05	-1,08E+00	7,92E+00	2,25E+00
Palencia	(2,1,2)(0,0,2,12)	3,80E+05	2,95E+05	-1,10E+00	9,06E+00	1,57E+00
Pontevedra	(3,1,2)(2,1,2,12)	1,87E+06	1,41E+06	-1,34E+00	8,91E+00	2,28E+00
Rioja (La)	(2,1,1)(2,0,0,12)	5,70E+05	4,62E+05	-9,08E-01	7,74E+00	3,15E+00
Salamanca	(0,1,1)(0,0,2,12)	7,89E+05	6,23E+05	-1,48E+00	1,03E+01	1,91E+00
Segovia	(2,1,2)(1,1,0,12)	3,81E+05	2,89E+05	-1,64E+00	1,02E+01	2,25E+00
Sevilla	(2,1,2)(2,1,0,12)	2,00E+06	1,58E+06	-7,26E-01	6,71E+00	1,14E+00
Soria	(2,1,2)(1,1,0,12)	2,20E+05	1,75E+05	-1,24E+00	9,94E+00	1,57E+00
Tarragona	(1,0,0)(2,1,0,12)	3,09E+06	2,43E+06	-2,37E+00	1,37E+01	3,13E+00
Teruel	(2,1,2)(1,1,0,12)	4,25E+05	3,33E+05	-1,80E+00	1,15E+01	2,10E+00
Toledo	(2,1,2)(2,1,2,12)	1,10E+06	8,92E+05	-6,52E-01	7,16E+00	1,74E+00
Valencia	(4,1,1)(2,0,0,12)	3,72E+06	3,01E+06	-6,02E-01	6,31E+00	1,89E+00
Valladolid	(2,1,1)(2,0,0,12)	7,44E+05	6,17E+05	-6,92E-01	6,89E+00	1,60E+00
Vizcaya	(4,1,3)(2,0,0,12)	1,26E+06	1,09E+06	-4,04E-01	5,48E+00	1,50E+00
Zamora	(2,1,2)(2,1,0,12)	4,88E+05	3,61E+05	-2,01E+00	1,08E+01	2,23E+00
Zaragoza	(2,1,2)(2,0,0,12)	1,40E+06	1,17E+06	-5,70E-01	6,29E+00	1,39E+00

CAPÍTULO 2: Utilización de los modelos estadísticos de aprendizaje automático ARIMA y SARIMA para medir el impacto de la COVID-19 en las ventas provinciales de cigarrillos en España

Tabla 2.2. Modelos SARIMA seleccionados para la previsión en cajetillas.

	SARIMA (p,d,q)(P,D,Q,m)	RMSE	MAE	MPE	MAPE	MASE
Alava	(2,1,2)(2,0,1,12)	3,78E-01	3,19E-01	-8,26E-01	7,67E+00	1,35E+00
Albacete	(2,1,2)(2,1,2,12)	4,86E-01	3,85E-01	-7,16E-01	7,61E+00	1,69E+00
Alicante	(2,1,0)(2,1,0,12)	7,44E-01	6,36E-01	-1,01E+00	1,03E+01	2,97E+00
Almería	(3,1,2)(2,1,1,12)	6,42E-01	5,21E-01	-1,11E+00	9,19E+00	3,44E+00
Asturias	(2,1,2)(2,0,0,12)	4,12E-01	3,26E-01	-7,59E-01	6,94E+00	2,29E+00
Ávila	(5,1,1)(2,1,0,12)	9,38E-01	7,44E-01	-3,01E+00	1,48E+01	2,82E+00
Badajoz	(5,1,1)(2,1,1,12)	4,30E-01	3,46E-01	-8,53E-01	7,44E+00	1,99E+00
Balears (Illes)	(2,1,0)(2,1,0,12)	1,12E+00	9,20E-01	-2,33E+00	1,36E+01	2,83E+00
Barcelona	(2,1,2)(2,0,0,12)	3,62E-01	2,98E-01	-6,84E-01	6,82E+00	1,62E+00
Burgos	(2,1,1)(2,0,0,12)	5,57E-01	4,48E-01	-1,43E+00	9,71E+00	2,93E+00
Cáceres	(2,1,2)(2,1,1,12)	6,14E-01	4,77E-01	-1,38E+00	9,31E+00	2,27E+00
Cádiz	(2,1,2)(2,1,0,12)	4,81E-01	3,72E-01	-1,86E+00	1,18E+01	1,87E+00
Cantabria	(4,1,1)(2,1,2,12)	5,94E-01	4,44E-01	-1,31E+00	8,74E+00	2,64E+00
Castellón	(3,1,2)(2,1,1,12)	7,84E-01	5,94E-01	-1,68E+00	1,06E+01	2,69E+00
Ciudad Real	(2,1,2)(2,1,1,12)	4,83E-01	3,64E-01	-9,65E-01	7,45E+00	2,63E+00
Córdoba	(2,1,2)(2,0,0,12)	3,21E-01	2,53E-01	-7,14E-01	6,67E+00	1,70E+00
Coruña (A)	(2,1,2)(2,0,0,12)	4,13E-01	3,32E-01	-8,47E-01	7,46E+00	2,51E+00
Cuenca	(1,1,2)(2,1,1,12)	7,25E-01	5,44E-01	-1,47E+00	1,00E+01	2,58E+00
Girona	(2,1,2)(1,1,1,12)	3,63E+00	2,93E+00	-6,20E+00	2,21E+01	2,66E+00
Granada	(4,1,0)(2,1,2,12)	4,51E-01	3,58E-01	-8,75E-01	7,43E+00	2,23E+00
Guadalajara	(5,1,0)(2,1,1,12)	4,50E-01	3,66E-01	-7,54E-01	7,72E+00	1,99E+00
Guipúzcoa	(2,1,1)(1,1,1,12)	9,24E-01	7,42E-01	-1,35E+00	1,06E+01	1,80E+00
Huelva	(2,1,1)(2,1,0,12)	6,55E-01	5,25E-01	-1,29E+00	1,03E+01	2,48E+00
Huesca	(4,1,1)(2,1,1,12)	9,17E-01	6,69E-01	-1,74E+00	1,07E+01	2,29E+00
Jaén	(2,1,2)(2,0,0,12)	3,27E-01	2,65E-01	-5,16E-01	5,82E+00	1,97E+00
León	(4,1,3)(1,1,2,12)	5,40E-01	4,15E-01	-1,36E+00	9,10E+00	2,59E+00
Lleida	(2,1,1)(2,0,0,12)	7,98E-01	6,94E-01	-1,48E+00	1,07E+01	1,21E+00
Lugo	(4,1,1)(2,0,0,12)	4,61E-01	3,59E-01	-1,13E+00	8,41E+00	2,02E+00
Madrid	(2,1,4)(2,0,0,12)	2,93E-01	2,45E-01	-4,94E-01	5,88E+00	1,47E+00
Málaga	(2,1,2)(2,1,0,12)	6,66E-01	5,81E-01	-1,33E+00	1,20E+01	3,38E+00
Murcia	(5,1,0)(2,1,1,12)	4,44E-01	3,69E-01	-7,06E-01	6,99E+00	1,73E+00
Navarra	(2,1,2)(2,1,0,12)	9,57E-01	7,69E-01	-1,41E+00	1,07E+01	3,19E+00
Ourense	(2,1,2)(1,1,1,12)	4,52E-01	3,41E-01	-1,06E+00	7,91E+00	2,35E+00
Palencia	(2,1,1)(2,0,0,12)	5,60E-01	4,35E-01	-1,09E+00	9,00E+00	1,57E+00
Pontevedra	(4,1,0)(2,1,2,12)	4,93E-01	3,72E-01	-1,33E+00	8,84E+00	2,23E+00
Rioja (La)	(2,1,2)(2,0,0,12)	4,60E-01	3,75E-01	-9,08E-01	7,77E+00	2,79E+00
Salamanca	(0,1,2)(2,0,0,12)	5,65E-01	4,47E-01	-1,47E+00	1,03E+01	1,91E+00
Segovia	(5,1,0)(2,1,2,12)	6,05E-01	4,59E-01	-1,63E+00	1,01E+01	2,23E+00
Sevilla	(2,1,0)(2,1,0,12)	2,75E-01	2,17E-01	-7,33E-01	6,70E+00	1,16E+00
Soria	(1,0,0)(1,1,0,12)	5,82E-01	4,62E-01	-1,24E+00	9,90E+00	1,54E+00
Tarragona	(3,1,1)(2,1,0,12)	9,92E-01	7,81E-01	-2,32E+00	1,36E+01	3,16E+00
Teruel	(5,0,2)(2,1,0,12)	7,68E-01	6,02E-01	-1,78E+00	1,14E+01	2,08E+00
Toledo	(4,1,0)(2,1,2,12)	4,19E-01	3,36E-01	-6,47E-01	7,09E+00	1,82E+00
Valencia	(2,1,2)(2,0,0,12)	3,85E-01	3,13E-01	-5,98E-01	6,33E+00	1,98E+00
Valladolid	(2,1,2)(2,0,0,12)	3,56E-01	2,95E-01	-6,88E-01	6,88E+00	1,61E+00
Vizcaya	(5,1,4)(2,0,0,12)	2,76E-01	2,39E-01	-4,08E-01	5,53E+00	1,44E+00
Zamora	(1,1,3)(1,1,1,12)	6,40E-01	4,76E-01	-1,99E+00	1,08E+01	2,50E+00
Zaragoza	(2,1,2)(2,0,0,12)	3,72E-01	3,11E-01	-5,76E-01	6,35E+00	1,42E+00

2.3 Resultados

La tabla 2.3 muestra los resultados de la comparación entre las ventas reales observadas después del COVID-19 y las estimaciones realizadas por el modelo (de enero de 2020 a diciembre de 2021). En este sentido, se muestran los resultados de los gaps detectados en términos de ventas en euros, en cajetillas y en cajetillas per cápita. Los gaps positivos indican que las ventas observadas superan las estimaciones realizadas por el modelo, mientras que los gaps negativos indican que las ventas reales después del COVID-19 son inferiores a las estimaciones realizadas por los modelos SARIMA estimados. En la tabla se puede observar el mínimo, máximo y promedio de los gaps provinciales calculados. Además, en la Figura 2.1 se muestra gráficamente la dinámica de la serie temporal junto con la previsión realizada mediante la variable cajetillas per cápita.

Si nos centramos en el gap medio calculado, en algunas provincias el impacto del COVID-19 en las ventas de tabaco ha sido casi nulo. En concreto, en Almería, Ávila, Cantabria, Coruña (A), Valladolid y Zaragoza, el impacto de la COVID-19 en las cajetillas per cápita es inferior al 1% en valor absoluto. Ante esta situación, en la Figura 2.1 se puede observar cómo en estas provincias, en las que las ventas de tabaco no se vieron afectadas por el COVID-19, las líneas de pronóstico y las ventas reales post-COVID-19 se superponen. Sin embargo, en otras provincias el impacto del COVID-19 ha provocado un importante efecto negativo en las ventas per cápita de cajetillas que alcanza, de media, hasta el -25,72%. Las provincias en las que se observa esta situación son Alicante/Alacant, Baleares (Illes), Girona, Guipúzcoa, Lleida y Málaga, en las que el impacto medio del COVID-19 en las ventas mensuales de tabaco ha sido del -18,95%, -25,72%, -22,71%, -14,98%, -16,66% y -11,41%, respectivamente. En el caso de estas provincias, en la Figura 2.1 se observa que la línea de previsión supera la línea de ventas Post COVID-19 desde enero de 2020 hasta diciembre de 2021. En todos los casos, las provincias en las que se observan estos efectos son de zonas con alta afluencia de turistas y zonas fronterizas con Francia. Estos resultados están en línea con la literatura previa que indica que las ventas de tabaco en España están muy condicionadas por las ventas a turistas y residentes en Francia (Cadahia et al, 2021; Gomajee et al., 2021).

Respecto al valor mínimo de los gaps provinciales calculados en ventas per cápita de cajetillas, en la Tabla 2.3 se muestra que el mayor impacto, en valor absoluto, de la COVID-19 en las ventas de tabaco se observó en Alicante/Alacant, Balears (Illes), Girona, Guipúzcoa, Lleida y Navarra, en las que el valor mínimo del impacto de la COVID-19 en las ventas mensuales de tabaco fue del -39,11%, -58,20%, -66,74%, -54,48%, -46,69% y -51,38%, respectivamente. En todos los casos, este valor mínimo se detectó en los meses de febrero y/o marzo de 2020, meses en los que las fronteras de España estuvieron cerradas debido a la pandemia de COVID-19. Es decir, el mayor impacto en valor absoluto de COVID-19 en las ventas de tabaco también se observa en las provincias limítrofes con Francia y en las provincias con alta afluencia de turistas. Por otro lado, las provincias en las que el impacto mínimo ha sido menor en valor absoluto son Cádiz y Sevilla, donde dicho impacto ha sido del -11,37% y -5,41%, respectivamente. Estos resultados también están en línea con la literatura previa que indica que las ventas en

CAPÍTULO 2: Utilización de los modelos estadísticos de aprendizaje automático ARIMA y SARIMA para medir el impacto de la COVID-19 en las ventas provinciales de cigarrillos en España

Cádiz y Sevilla se ven afectadas por la proximidad de estas provincias a Gibraltar, zona con la que existe un importante diferencial de precios (Cadañia et al., 2021).

Nuestros resultados indican que las restricciones implementadas por los gobiernos debido al COVID-19 han tenido un efecto significativo en las ventas provinciales de tabaco en España. En este sentido, encontramos que las provincias en las que las ventas se ven más afectadas son las fronterizas y turísticas, lo que parece indicar que, independientemente de la limitación del ocio, la restricción que más ha afectado a las ventas es el cierre de fronteras. Los resultados sugieren que, en zonas turísticas y fronterizas con Francia, la COVID-19 ha provocado un efecto negativo en las ventas de tabaco que en la mayoría de los casos aún no se había revertido hasta diciembre de 2021.

Tabla 2.3. Impacto provincial de la COVID-19 en el mercado tabacalera español.

	Gap in Euros (%)			Gap in Packs (%)			Gap in Per capita packs (%)		
	Min	Max	Mean	Min	Max	Mean	Min	Max	Mean
Alava	-18,01	22,02	0,26	-19,07	25,26	2,02	-19,00	27,09	2,87
Albacete	-21,68	12,96	-3,04	-20,27	15,10	0,44	-5,83	59,23	26,89
Alicante/Alacant	-38,13	11,16	-14,88	-38,96	8,41	-18,80	-39,11	8,52	-18,95
Almería	-23,63	10,38	-1,47	-21,72	14,00	0,53	-21,65	14,01	0,22
Asturias	-18,21	10,08	-1,15	-14,06	23,20	7,07	-15,70	19,49	5,17
Ávila	-26,73	18,05	-0,55	-24,05	17,17	1,57	-25,08	15,36	0,29
Badajoz	-21,31	11,16	-0,03	-20,98	14,49	2,34	-21,42	13,57	1,41
Balears (Illes)	-58,08	8,50	-26,50	-57,97	13,32	-24,93	-58,20	5,41	-25,72
Barcelona	-20,96	5,34	-5,50	-19,35	7,21	-3,87	-18,91	9,57	-2,35
Burgos	-35,41	16,68	-5,01	-31,13	32,86	6,75	-33,38	21,68	-2,26
Cáceres	-28,66	14,05	-2,10	-26,68	16,52	2,41	-26,76	16,16	2,08
Cádiz	-8,68	22,96	5,74	-10,12	17,92	2,93	-11,37	17,13	2,15
Cantabria	-20,81	12,65	-1,76	-18,90	15,03	0,73	-19,17	13,81	-0,09
Castellón/Castelló	-26,30	11,48	-3,67	-24,58	12,31	-2,25	-24,73	12,79	-3,13
Ciudad Real	-20,81	13,59	0,86	-19,02	18,36	4,46	-19,36	16,78	3,36
Córdoba	-15,72	11,83	0,89	-16,25	16,38	1,34	-16,35	17,01	1,72
Coruña (A)	-14,06	10,25	-1,17	-14,78	12,72	-1,10	-15,98	13,76	0,40
Cuenca	-22,24	15,22	2,42	-23,25	15,55	1,95	-23,45	14,14	2,13
Girona	-65,26	9,61	-17,54	-65,40	8,44	-18,21	-66,74	4,76	-22,71
Granada	-21,43	9,42	-3,10	-20,15	9,91	-1,33	-20,43	8,63	-2,02
Guadalajara	-20,37	14,15	1,05	-16,48	15,09	4,22	-17,27	14,58	3,00
Guipúzcoa	-53,43	22,25	-11,65	-54,48	20,29	-15,21	-54,48	20,10	-14,98
Huelva	-18,14	16,42	0,37	-18,38	3,75	-6,06	-18,38	2,95	-7,14
Huesca	-33,40	15,58	-5,25	-34,08	17,38	-5,34	-35,35	15,17	-6,60
Jaén	-25,28	6,99	-4,10	-23,38	12,25	-1,06	-23,43	12,57	-0,99
León	-22,22	15,01	-1,02	-16,70	22,72	5,32	-16,86	22,07	4,79
Lleida	-47,06	10,75	-16,42	-46,95	12,24	-16,76	-46,69	11,46	-16,66
Lugo	-16,04	20,94	-0,76	-12,10	23,77	4,23	-11,83	23,14	3,83
Madrid	-20,06	5,36	-4,66	-19,29	7,23	-4,27	-17,69	11,76	-1,33
Málaga	-29,07	8,13	-12,07	-29,23	9,88	-10,95	-29,53	9,74	-11,41
Murcia	-24,57	10,08	-4,24	-22,88	10,11	-2,77	-23,27	12,56	-2,45
Navarra	-50,82	32,66	-5,84	-51,48	31,02	-7,33	-51,38	30,92	-6,74
Ourense	-19,94	18,77	0,57	-15,10	14,68	0,23	-14,68	18,05	3,62
Palencia	-22,39	23,28	0,82	-15,74	34,22	8,02	-18,49	27,07	4,03
Pontevedra	-16,82	11,58	-0,36	-15,11	15,44	0,46	-15,88	13,09	-1,25
Rioja (La)	-24,21	12,10	-3,58	-23,49	14,67	-1,97	-23,41	15,79	-1,49
Salamanca	-34,98	18,00	-7,76	-30,38	27,67	-1,92	-30,80	25,68	-3,07
Segovia	-23,38	14,12	0,25	-21,04	13,92	1,75	-20,82	17,15	2,15
Sevilla	-3,79	15,71	4,32	-4,40	16,68	4,75	-5,41	15,82	3,79
Soria	-24,35	27,71	1,85	-23,83	28,97	2,99	-24,04	29,18	3,49
Tarragona	-33,09	8,87	-8,12	-31,95	12,31	-9,89	-30,66	9,99	-7,53
Teruel	-19,93	16,38	-2,85	-18,19	15,19	1,04	-18,87	15,81	-2,27
Toledo	-16,61	12,20	2,59	-15,71	12,32	1,92	-16,86	8,75	-3,70
Valencia/València	-20,97	9,35	-4,94	-22,09	4,35	-5,49	-22,27	4,08	-5,73
Valladolid	-26,57	6,10	-6,14	-22,55	16,02	0,25	-22,44	16,80	0,61
Vizcaya	-17,89	7,13	-3,47	-18,52	6,61	-3,40	-18,76	6,88	-3,43
Zamora	-25,08	24,26	1,72	-21,32	27,12	6,72	-22,25	23,76	4,65
Zaragoza	-23,41	5,65	-5,43	-21,09	8,57	-2,24	-20,92	11,84	-0,62

CAPÍTULO 2: Utilización de los modelos estadísticos de aprendizaje automático ARIMA y SARIMA para medir el impacto de la COVID-19 en las ventas provinciales de cigarrillos en España

Figura 2.1: Previsión de ventas de cajetillas en 2020 y 2021 en base a los mejores modelos SARIMA seleccionados.

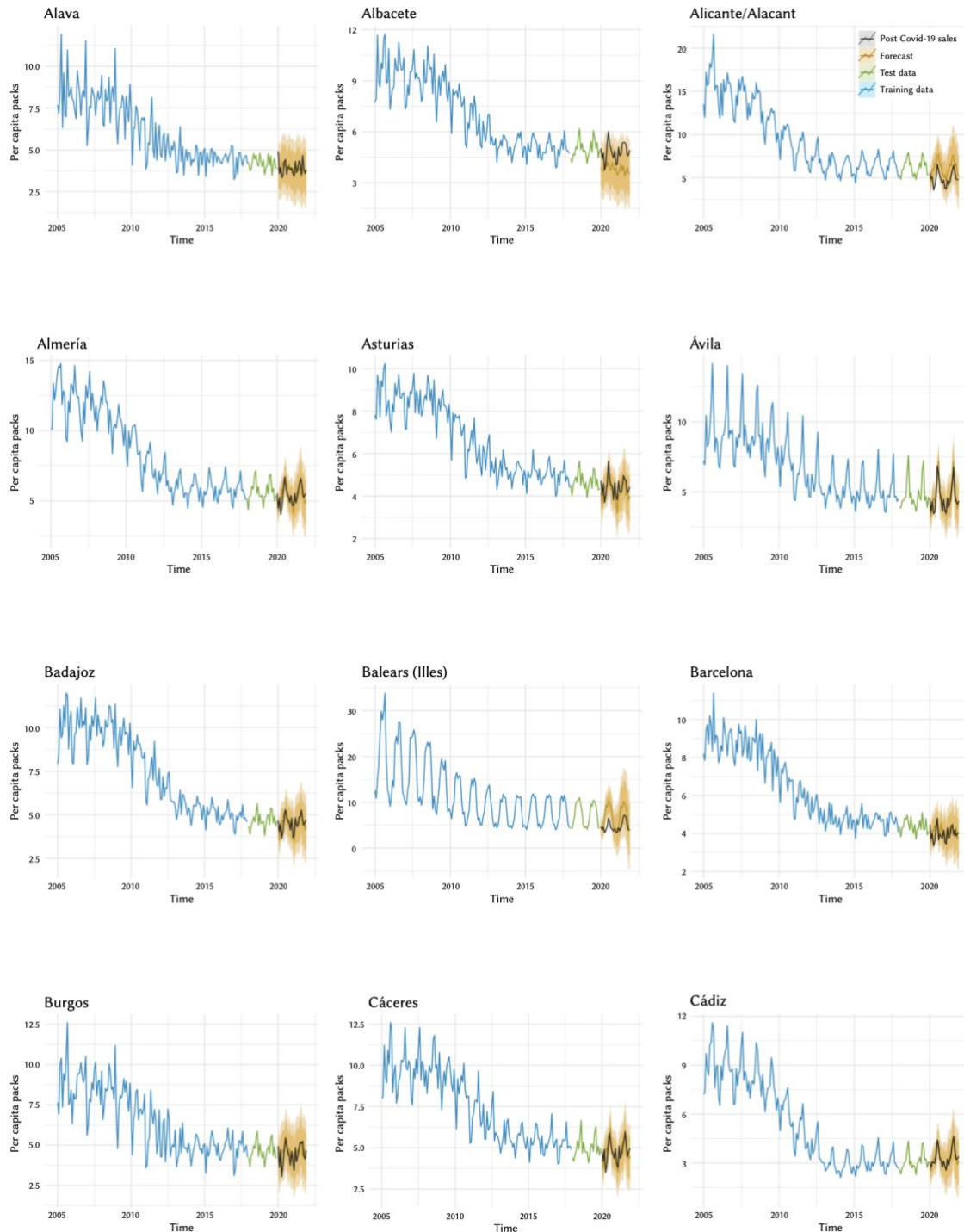
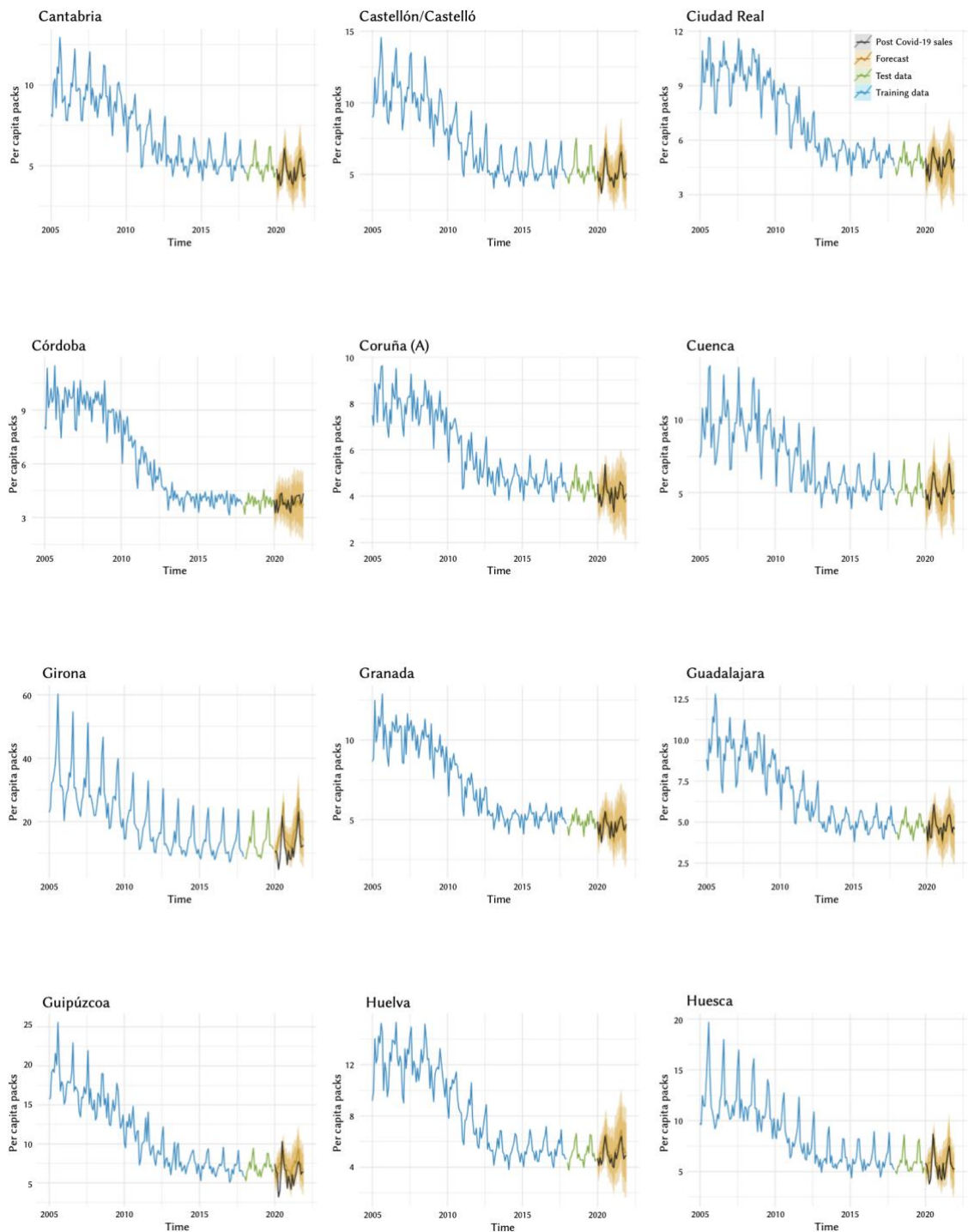


Figura 2.1: Previsión de ventas de cajetillas en 2020 y 2021 en base a los mejores modelos SARIMA seleccionados.



CAPÍTULO 2: Utilización de los modelos estadísticos de aprendizaje automático ARIMA y SARIMA para medir el impacto de la COVID-19 en las ventas provinciales de cigarrillos en España

Figura 2.1: Previsión de ventas de cajetillas en 2020 y 2021 en base a los mejores modelos SARIMA seleccionados.

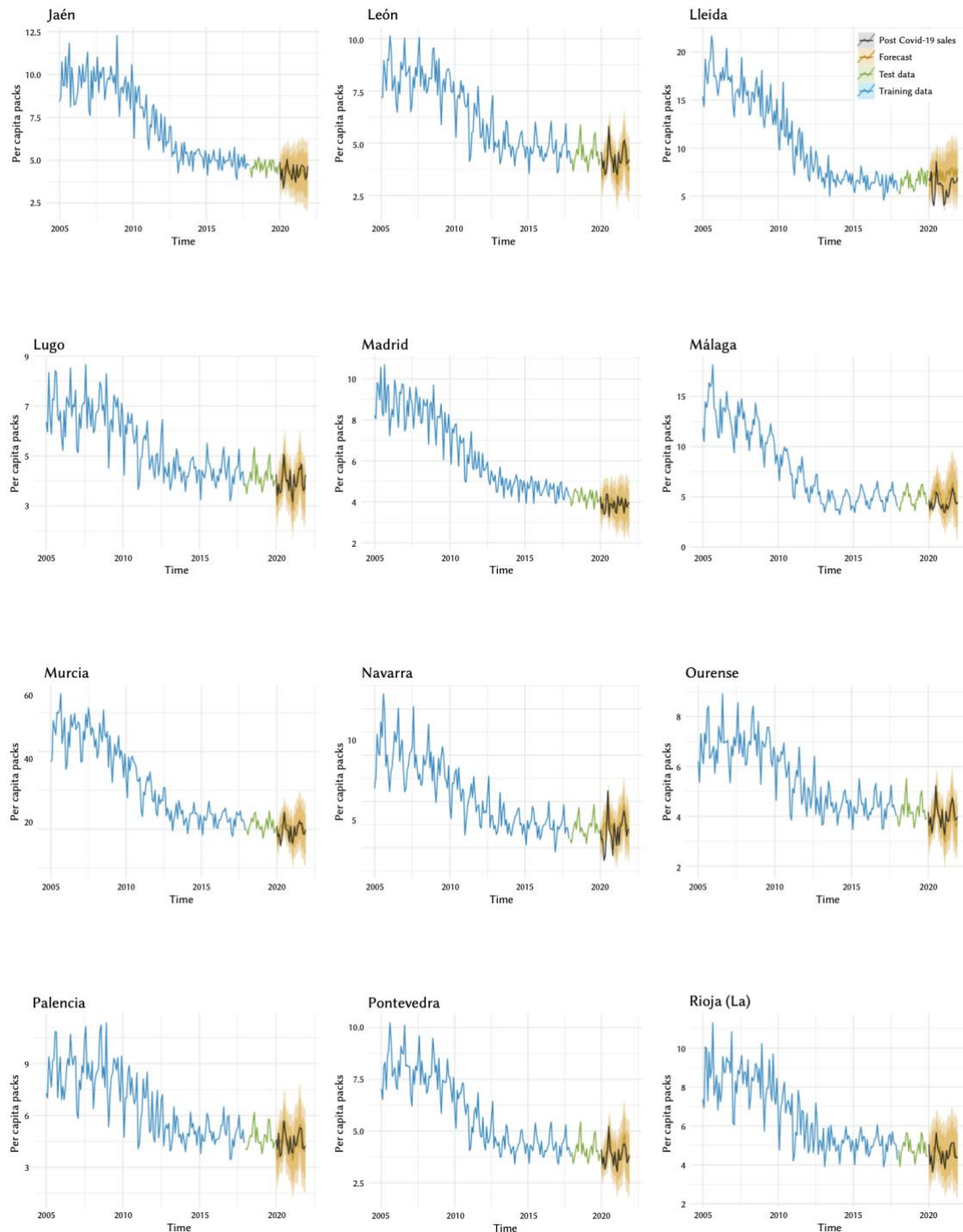
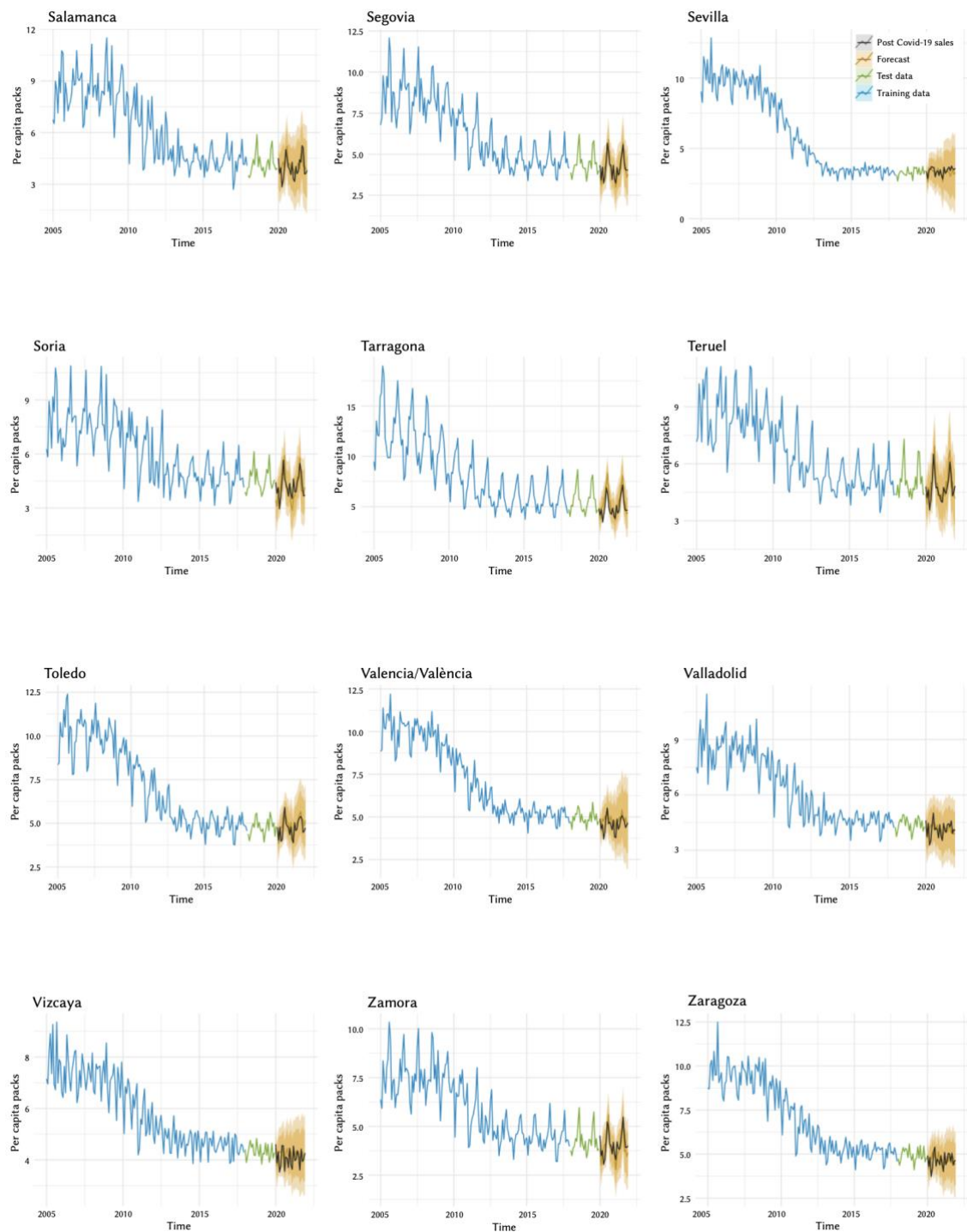


Figura 2.1: Previsión de ventas de cajetillas en 2020 y 2021 en base a los mejores modelos SARIMA seleccionados.



2.4 Conclusiones

En este estudio hemos predicho el impacto que ha tenido la COVID-19 en las ventas de tabaco en España (en euros, en cajetillas y en cajetillas per cápita) desde enero de 2020 hasta diciembre de 2021, utilizando modelos estadísticos ARIMA y SARIMA de ML. Nuestras estimaciones indican que el mayor impacto del COVID-19 en las ventas de cigarrillos se observa en las provincias turísticas y fronterizas con Francia, donde, en los meses de cierre de fronteras, las ventas fueron hasta un 66,74% inferiores a las previsiones realizadas. Por el contrario, en las provincias limítrofes con Gibraltar el impacto de la COVID-19 fue muy leve (5,41%). Los motivos por los que el COVID-19 puede afectar a las ventas de tabaco pueden ser la sensibilización pública, las restricciones de ocio, el cierre de fronteras, etc. Sin embargo, parece que el mayor impacto del COVID-19 ha sido provocado por el cierre de fronteras.

En esta línea, en provincias como Alicante/Alacant, Baleares (Illes), Girona, Guipúzcoa, Lleida, Málaga y Navarra se ha observado un fuerte impacto de la COVID-19 en las ventas de tabaco. Además, el menor impacto se ha observado en Cádiz y Sevilla. Si se observa el impacto medio nacional, en España la COVID-19 casi no ha tenido efecto. En concreto, el impacto provincial medio en España se acerca al -2%. Esto se debe a que las previsiones realizadas con los modelos SARIMA y las ventas Post COVID-19 son casi las mismas en la mayoría de las provincias españolas.

Los resultados parecen mostrar que el cierre de fronteras ha tenido un marcado impacto en las ventas provinciales de tabaco en España. Por tanto, parece que el efecto del turismo y las compras transfronterizas entre España y Francia y España y Gibraltar se han visto alterados por las restricciones fronterizas provocadas por la COVID-19. Basándonos en nuestras predicciones y pronósticos, los responsables de la formulación de políticas deben tomar las decisiones correctas sobre las diferencias de precios del tabaco observadas entre los países europeos donde hay un movimiento transfronterizo constante y abundante. Para mantener el tabaquismo bajo control, todos los países deben tomar decisiones armonizadas.

Este trabajo no está exento de limitaciones. Un trabajo reciente revela que Philip Morris International, el principal fabricante de tabaco del mundo, está utilizando productos de tabaco calentado (PTC) para sustituir el cigarrillo tradicional. Los resultados obtenidos pueden verse influenciados por este fenómeno (Golpe et al., 2021). Además, un estudio reciente también indica que la asequibilidad de los cigarrillos es un factor clave para su demanda en España. Por este motivo, parte del “no loss” en Sevilla y Cádiz puede estar motivado por el efecto asequibilidad (Cadahia et al., 2022).

Dadas las limitaciones señaladas, las líneas de investigación futuras se pueden resumir en tres. En primer lugar, es interesante analizar si los PTC están provocando parte de los gaps detectados en este trabajo. En segundo lugar, sería importante analizar el papel que juega la asequibilidad en dichos gaps. Finalmente, se debe analizar

el comportamiento de los productos sustitutos para saber si parte de los efectos detectados en este trabajo pueden deberse al consumo de otros productos alternativos.

CAPÍTULO 2: Utilización de los modelos estadísticos de aprendizaje automático ARIMA y SARIMA para medir el impacto de la COVID-19 en las ventas provinciales de cigarrillos en España

Capítulo 3

Convergencia en las ventas oficiales de cigarrillos en España: un análisis de clústeres de 2005 a 2021

3.1 Introducción

Es bien sabido y está ampliamente estudiado que el tabaquismo provoca enfermedades agudas y crónicas (Martín-Álvarez et al., 2020a). Para minimizar este efecto nocivo sobre la salud de la población, el Convenio Marco de la Organización Mundial de la Salud para el Control del Tabaco (CMCT-OMS) (WHO,2003) es el primer tratado sanitario internacional que obliga a los países a aplicar medidas de control del tabaco. En España, el CMCT-OMS fue ratificado el 11 de abril de 2005. En 2008, la OMS desarrolló el paquete MPOWER (WHO, 2008) para ayudar a los países a cumplir sus compromisos con el tratado. Uno de los seis componentes de MPOWER es hacer cumplir las prohibiciones sobre publicidad, promoción y patrocinio del tabaco.

Para minimizar el impacto que el consumo de tabaco tiene en las sociedades, los gobiernos han tomado medidas que afectan tanto a la demanda como a la oferta. Por el lado de la demanda, se han aprobado leyes que prohíben fumar o que tienen en cuenta el impacto de los impuestos en los precios (Contreary et al., 2015). Por el lado de la oferta, los fabricantes y minoristas de productos del tabaco se han limitado a mantener

CAPÍTULO 3: CONVERGENCIA en las ventas oficiales de cigarrillos en España: un análisis de clústeres de 2005 a 2021

los precios para promocionar los productos del tabaco y mantener alta la demanda (Escario & Molina, 2004; Pinilla & Abásolo, 2017). Estas medidas surgen porque la CMCT-OMS ha acelerado la aplicación de medidas en varios ámbitos políticos cruciales, uno de los cuales es la prohibición de la publicidad, la promoción y el patrocinio del tabaco (Almeida et al., 2021).

En el caso de España, en los últimos años el Gobierno ha aplicado con éxito diferentes leyes restrictivas para controlar la venta legal de cigarrillos. En concreto, las leyes antitabaco españolas de 2005 y 2010 afectaron al consumo al introducir restricciones por el lado de la demanda y la oferta con la Ley 28/2005 (26 de diciembre) y la Ley 42/2010 (30 de diciembre), de medidas sanitarias frente al tabaquismo y reguladora de la venta, el suministro, el consumo y la publicidad de los productos del tabaco. Ambas leyes introducen una dura restricción a las estrategias de comercialización, ya que no se permite la publicidad, promoción y patrocinio del tabaco fuera de un establecimiento autorizado para la venta al por menor de tabaco en régimen de monopolio. En concreto, el texto de la Ley 28/2005 indica que "se prohíbe el patrocinio de los productos del tabaco, así como todo tipo de publicidad, y la promoción de los citados productos en todos los medios y soportes, incluidas las máquinas expendedoras y los servicios de la sociedad de la información". Estas medidas buscan homogeneizar el mercado y hacer menos atractivo el tabaco, especialmente para la población joven. Aunque la publicidad, la promoción y el patrocinio del tabaco estaban prohibidos por la Ley 28/2005, la Ley 42/2010 establece una concreción más restrictiva de la prohibición (Kasza et al., 2011; Pinilla et al., 2019).

El efecto que las leyes han tenido sobre el mercado español del tabaco desde 2005 ha sido analizado desde diferentes perspectivas por los académicos. Por un lado, hay trabajos que exponen los efectos sobre la prevalencia de fumadores y las ventas de tabaco (Pinilla & Abásolo, 2017; Pinilla et al., 2019; Pons-Vigués et al., 2019). Por otro lado, otros estudios se han centrado en la exposición al humo ambiental del tabaco (Fernández et al., 2015; Galán et al., 2007). Sin embargo, pocos de ellos se centran en analizar el cumplimiento de las prohibiciones sobre publicidad, promoción y patrocinio del tabaco (Almeida et al., 2021).

En este estudio, investigamos las tendencias de ventas de cigarrillos por marcas tras las dos importantes políticas antitabaco (2005, 2010) en España, que incluyen prohibiciones de publicidad, promoción y patrocinio del tabaco. Analizamos cómo se comportaron las diferentes marcas en relación con el descenso de las ventas de cigarrillos tras la legislación integral. Aunque, hasta donde sabemos, este análisis nunca se ha realizado desde la perspectiva de las marcas, un estudio reciente ha analizado el efecto de las leyes antitabaco desde la perspectiva de los clubes de convergencia (Tirgil et al., 2021). Para alcanzar el objetivo planteado, utilizamos el concepto de "club de convergencia". La premisa de un "club de convergencia" es que los grupos económicos con características iniciales similares, como las distintas marcas de cigarrillos que se venden en España, pueden alcanzar un estado estacionario de equilibrio mediante una senda de desarrollo equilibrado (Ben-David, 1998). La idea se basa en los modelos neoclásicos de crecimiento (Solow, 1956), en los que la hipótesis de convergencia β absoluta o incondicional predice que los países convergen a un estado estacionario

común independientemente de las condiciones iniciales, mientras que la hipótesis de convergencia condicional representa la convergencia a un estado estacionario común, independiente de las condiciones iniciales y de las propiedades estructurales comunes (Rodríguez-Andrés & Amavilah, 2022). Por lo tanto, este trabajo pretende identificar si ha habido convergencia total, divergencia o ambas en las ventas de cigarrillos en España desde 2005 hasta 2021. Utilizamos una novedosa prueba de convergencia log t, que permite estimar las trayectorias y velocidades de convergencia de acuerdo con Phillips & Sul (2007) y Phillips & Sul (2009).

3.2 Datos y metodología

Para este trabajo se han obtenido los datos mensuales de ventas de productos de tabaco en España publicados por el Comisionado para el Comercio de Tabacos de España (órgano dependiente del Ministerio de Hacienda y Función Pública) desde enero de 2005 hasta diciembre de 2021. Además, se ha realizado un preprocesado de los datos para detectar las marcas que han estado presentes en el mercado durante los tres subperiodos 2005-2021, 2005- 2010 y 2011-2021. Así, el panel de datos 2005-2021 cuenta con 8.160 observaciones (40 marcas con 204 meses), el de 2005-2010 tiene 5.328 observaciones (74 marcas con 72 meses), y de 2011-2021 tiene 8.844 observaciones (67 marcas con 132 meses).

Para el desarrollo de este trabajo se utilizará la metodología propuesta por Phillips y Sul (2007) y Phillips y Sul (2009), cuyo modelo factorial no lineal y variable en el tiempo es el modelo de detección de clubes más avanzado hasta la fecha. Esto permite diversas trayectorias temporales de los objetos estudiados y su heterogeneidad individual. Esta novedosa forma de definir los clubes de convergencia permite determinar, a través de las características intrínsecas de cada objeto, su pertenencia a un determinado club, lo que, por tanto, representa una gran ventaja a la hora de obtener resultados. Otros métodos anteriores requerían la definición de un objeto de referencia a priori, y de forma totalmente exógena, para clasificar al resto en función de su posición, lo que los hace menos eficientes y extrapolables.

En este caso, la formación de clubes de convergencia se da a partir de las características endógenas de cada objeto y por tanto la fiabilidad, así como su extrapolación, es mayor (Sichera & Pizzuto, 2019).

Phillips y Sul (2007) y Phillips y Sul (2009) proponen una modificación en la descomposición de los datos de la variable en estudio. Si los datos se descomponen habitualmente de la siguiente manera donde la variable objeto de estudio es X:

$$X_{it} = g_{it} + a_{it}$$

CAPÍTULO 3: CONVERGENCIA en las ventas oficiales de cigarrillos en España: un análisis de clústeres de 2005 a 2021

se define g_{it} como la parte sistémica que afecta a toda la variable e incluye el componente común, y a_{it} como el componente transitorio en el tiempo. Si se quiere tener la heterogeneidad del tiempo en la variable en estudio, la ecuación resultante es:

$$X_{it} = \left(\frac{g_{it} + a_{it}}{\mu_t} \right) \mu_t = b_{it} \mu_t,$$

Donde b_{it} es el elemento que considera el tiempo y cambia con él, incluye un componente aleatorio que absorbe a_{it} , y μ_t es el factor común a ellos. Con esta propuesta de formulación factorial que separa las partes propias del objeto de las comunes, b_{it} es el camino hacia un estado común, determinado por μ_t .

La estimación del b_{it} es esencial para comprobar que los diferentes objetos convergen; sin embargo, esta estimación no es posible sin establecer previamente algunas restricciones adicionales. De esta forma, Phillips y Sul (2007) y Phillips y Sul (2009) establecen la siguiente formulación que será la ruta de transición relativa adecuada para calcular directamente con los datos y sin el establecimiento de restricciones o supuestos estructurales:

$$h_{it} = \frac{X_{it}}{N^{-1} \sum_{i=1}^N X_{it}} = \frac{b_{it}}{N^{-1} \sum_{i=1}^N b_{it}}$$

Con la aplicación de la fórmula anterior, los autores consiguen trazar una trayectoria individual para cada objeto sin depender de la tendencia común de todos los objetos. Así, será posible observar la tendencia exacta de un determinado objeto dentro de la tendencia global μ_t .

Adicionalmente para esta trayectoria de transición, en caso de que exista convergencia, debe existir un límite común para cada caso estudiado. De esta forma, el coeficiente de acierto tenderá a la unidad para cada objeto a medida que avance el tiempo ($h \rightarrow 1$ para todos los objetos $i = 1, 2, \dots, N$ cuando $t \rightarrow \infty$).

Además, la distancia cuadrática media para el panel desde el límite común, la variación transversal H_{it} , debe converger a 0, de modo que:

$$H_t = N^{-1} \sum_{i=1}^N (h_{it} - 1)^2 \rightarrow 0 \quad \text{when } t \rightarrow \infty$$

A su vez, para construir un test estadístico de convergencia, Phillips & Sul (2007) asumen la siguiente estimación paramétrica de b_{it} :

$$b_{it} = b_i + \frac{\sigma_i \xi_{it}}{L(t)t^\alpha},$$

Con b_i como valor fijo, invariante en el tiempo, el parámetro ξ_{it} , como variables aleatorias independientes idénticamente distribuidas de $N(0, 1)$ a lo largo de i pero dependientes de t , $L(t)$ es una función lentamente variable en el tiempo que se aproxima a infinito a medida que el tiempo se aproxima a infinito ($L(t) \rightarrow \infty$ cuando $t \rightarrow \infty$), y el coeficiente de convergencia.

La prueba de hipótesis para la convergencia sería:

$$H_0: b_i = b$$

$$H_1: b_i \neq b$$

Si no se descarta la hipótesis nula, puede haber trayectorias diferentes para los objetos, incluida la divergencia.

Adicionalmente, Phillips y Sul (2007) y Phillips y Sul (2009) proponen estudiar la existencia de convergencia entre objetos estimando el siguiente modelo aplicando el método de mínimos cuadrados ordinarios:

$$\log \frac{H_1}{H_t} = -2 \log(\log t) = \alpha + \beta \log t + u_t, \quad \text{para } t = [rT], [rT] + 1, \dots, T.$$

La hipótesis nula de convergencia se contrasta aplicando la prueba t de una cola a un nivel de significación del 5%, rechazándose a este nivel si $t < -1,65$. La prueba es robusta a heteroscedasticidad y autocorrelación (HAC) a desigualdad $\alpha > 0$ (utilizando la estimación $\beta=2$) La prueba log- t servirá para contrastar la convergencia de los clubes para toda la muestra. Si esta se rechaza, Phillips y Sul (2007) y Phillips y Sul (2009) proponen repetir el procedimiento aplicando una agrupación en cuatro etapas.

Si se rechaza la prueba log- t para toda la muestra, repetimos la prueba con la metodología siguiente:

1. Ordenamos las unidades en orden descendente según la última observación del periodo.

2. Formaremos un grupo central ejecutando la regresión log- t para las primeras K unidades, donde $2 < k < N$, maximizándolo con la condición de que $t > -1,65$. Podemos establecer el tamaño del grupo central, k^* , como sigue:

$$k^* = \arg \max [t_k] \text{ s.a. } \min [t_k] > -1.65$$

En caso de que $k > 1,65$ no se cumpla para las dos primeras unidades, $k=2$, eliminaremos la primera unidad y repetiremos el procedimiento. Si ocurre que $k > 1,65$ no se cumple para ninguna de las unidades elegidas, toda la muestra diverge.

CAPÍTULO 3: CONVERGENCIA en las ventas oficiales de cigarrillos en España: un análisis de clústeres de 2005 a 2021

3. Selección de datos pertenecientes al club. Tras detectar el grupo * ejecutaremos la regresión log- t para añadir uno a uno los objetos que no pertenecen al grupo central. Si es k es mayor que el valor crítico c^* esta unidad se añadirá al club de convergencia. Todas estas unidades, incluidas las pertenecientes al club central, formarán la primera convergencia club.

4. Regla de repetición del proceso y parada. Si hay objetos para los que falla la condición anterior, hay que agruparlos todos y volver a ejecutar la prueba log-t, para comprobar si en ellos se cumple la condición $k > 1,65$. Si se cumple, tendremos u nuevo club de convergencia. Si no es así, se repetirán los pasos anteriores para identificar posibles subgrupos que formen grupos de convergencia. Si la condición $k > 1,65$ no se cumple para estos objetos en el paso 2, concluiremos que divergen. Los autores insisten en que la condición $k > 1,65$ debe cumplirse para los clubes. Si los clubes no la cumplen, podemos aumentar el valor de C^* hasta que se cumpla.

Dado que el número de clubes que el algoritmo podrá identificar está supeditado a la constitución de un primer grupo central y que de la posición de éste se derivan las siguientes, el valor crítico que fijamos cuando nos referimos a C^* debe ser meditado previamente. Este valor reside en el nivel y tipo de error que estamos dispuestos a asumir. Por tanto: Un nivel alto de C^* será útil si no queremos equivocarnos al clasificar objetos en clubes a los que probablemente no pertenezcan. Establecer un nivel alto de C^* hará que se formen más grupos, pudiendo incluso parecer más de los que realmente hay. Un nivel bajo de C^* será útil si perseguimos el objetivo de tener menos clubes y así tener una mayor interpretabilidad y una menor complejidad a la hora de definir los grupos. Phillips y Sul (2007) y Phillips y Sul (2009) recomiendan que, en muestras pequeñas ($T < 50$), se utilice un $C^* = 0$.

En cuanto al número de clubes resultantes y como solución a este problema, Phillips y Sul desarrollan un algoritmo para fusionar los clústeres formados con el algoritmo anterior. Este será capaz de detectar clubes cercanos entre sí ejecutando la prueba log-t comparando los clubes uno a uno con el resto de los clubes.

Siguiendo el mismo criterio anterior, si el estadístico t toma un valor inferior a $-1,65$, estos clubes no convergerán. Si, por el contrario, el estadístico es superior a $-1,65$, los clubes convergerán y formarán un nuevo club.

3.3 Resultados

En esta sección, de acuerdo con la estrategia econométrica descrita, presentamos los resultados de las estimaciones para investigar la convergencia absoluta y relativa en las ventas de cigarrillos en España desde 2005 hasta 2021. Para alcanzar el objetivo propuesto, se ha utilizado la versión 4.0 de R y el paquete *CovergenceClubs* (Sichera & Pizzuto, 2019). En un primer paso, se ha realizado el test estadístico $\text{Log}(t)$ para comprobar si existe convergencia global en los tres periodos seleccionados en función de los cambios regulatorios que se han producido en España. En segundo lugar, una vez rechazada la convergencia absoluta, aplicamos un algoritmo de agrupación para comprobar la existencia de convergencia de subgrupos en los tres periodos seleccionados.

Tabla 3.1. Estadísticas de la prueba $\text{Log}(t)$ para las ventas de cigarrillos por marcas, 2005-2021.

Periodo	Nº de marcas	Prueba	β	s.e.	Estadístico t	p-valor
2005-2021	40	log t	-0.747	0.150	-4.988	0.000
2005-2010	74	log t	-0.725	0.026	-28.041	0.000
2011-2021	67	log t	-0.793	0.139	-5.702	0.000

Los resultados de los estadísticos de la prueba $\text{Log}(t)$ se presentan en la tabla 3.1, que muestra la información necesaria para comprobar si se puede rechazar la hipótesis nula de convergencia global en los tres periodos seleccionados. Si nos centramos en las estimaciones realizadas con las cajetillas vendidas, la tabla 3.1 muestra que el coeficiente de la pendiente es igual a -0,747, -0,725 y -0,793 con un error típico de heteroscedasticidad y autocorrelación (HAC) de 0,150, 0,026 y 0.139. Además, obtenemos unos estadísticos t de -4,988, -28,041 y -5,702 (por debajo del valor crítico de -1,65). Así pues, se rechaza la hipótesis nula de convergencia global al nivel de significación del 5%. Los resultados sugieren que las ventas de cigarrillos en España por marcas divergen o sólo convergen entre subgrupos.

A continuación, aplicamos un algoritmo de agrupación para comprobar la existencia de convergencia de subgrupos. Los resultados (tablas 3.2-3.4) identifican diferentes clubes, y los correspondientes β estimados, donde $\beta = 2\alpha$ es el coeficiente escalado de la velocidad de convergencia del club, y α es la velocidad de convergencia estimada para cualquier club (Philips & Sul, 2007). Como puede observarse, el análisis de los subgrupos de convergencia se ha realizado para los periodos 2005-2021, 2005-2010 y 2011-2021. En todos los casos el estadístico $t > -1,65$ (valor crítico), lo que implica convergencia dentro del club. Este resultado se mantiene tanto para los años anteriores a 2011 como para los posteriores. Sin embargo, en el periodo 2005-2010 se observaron 5 clubes de convergencia, mientras que en el periodo 2011-2021 las marcas se agruparon en 3 clubes.

CAPÍTULO 3: CONVERGENCIA en las ventas oficiales de cigarrillos en España: un análisis de clústeres de 2005 a 2021

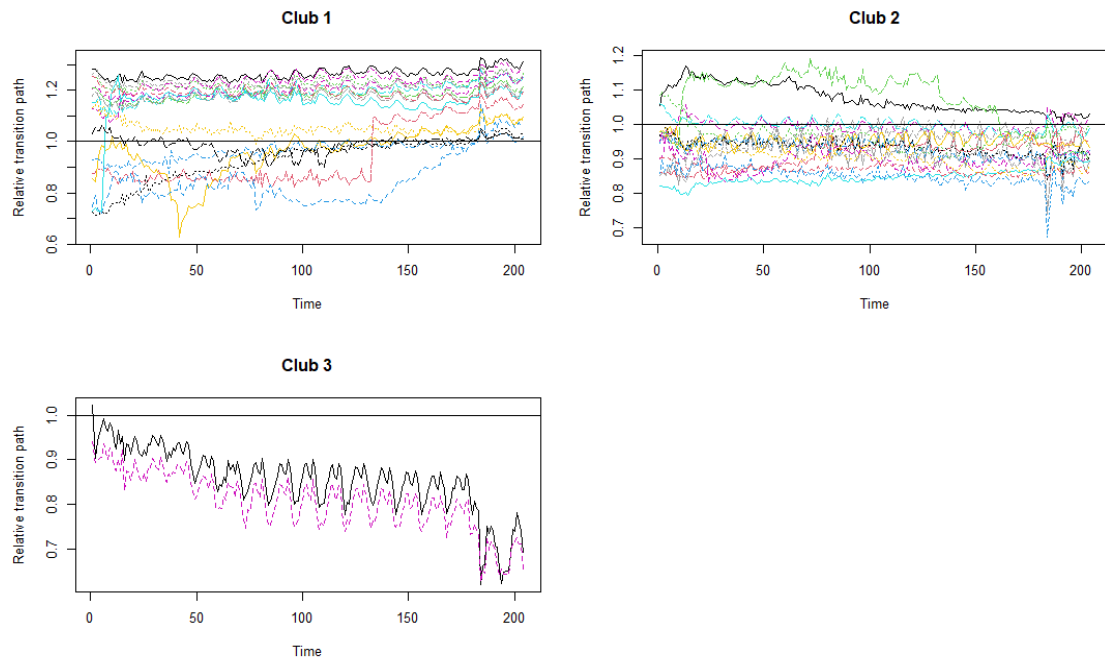
Analizando el periodo completo (2005-2021) también se encuentran 3 clubes. Parece que en el periodo 2011- 2021 la venta de cigarrillos tiene un comportamiento más homogéneo (3 clubes) que en el periodo 2005- 2010 (5 clubes y 2 marcas divergentes), lo que apoya lo establecido en la literatura previa que indica que la Ley 42/2010 proporcionó una especificación más restrictiva de la prohibición que la Ley 28/2005.

La Tabla 3.2 y la Figura 3.1 muestran los resultados de la aplicación del algoritmo de agrupación para comprobar la existencia de convergencia de subgrupos para el periodo total analizado (2005-2021). Como puede observarse, la aplicación del algoritmo genera 3 clubes y 5 marcas como divergentes. De las 40 marcas de cigarrillos que se han comercializado durante este periodo, 17 y 16 pertenecen a los clubes 1 y 2, respectivamente, mientras que el club 3 sólo incluye 2 marcas. La trayectoria de transición del club 1 es creciente, mientras que en el caso del club 2 se observa una trayectoria de transición mantenida con una ligera pendiente decreciente. El club 3 sólo incluye 2 marcas con un alto componente estacional debido a su vinculación con el turismo.

Tabla 3.2. Proceso del club de convergencia en las ventas de cigarrillos por marca (en cajetillas), 2005-2021.

Club	Número de marcas	β	s.e.	t-value	Marcas
1	17	0.793	0.262	3.026	Marlboro, Camel, Chesterfield, Fortuna, Lucky Strike, Rothmans, Austin, Ducal, Karelia, Winston, L&M, Nobel, Ducados Rubio, Ducados Negro, Elixir, R1, Benson & Hedges
2	16	0.264	0.081	3.251	John Player SP., Bullbrand, Excite, Gauloises Rubio, BN, Vogue, Silk Cut, Winfield, Peter Stuyvesant, Burton, Pall Mall, News, Davidoff Rubio, Dunhill, Royal Crown, Gitanes
3	2	0.576	0.433	1.332	Regal, Embassy
Divergentes	5				Rex, Lambert & Butler, Mayfair, Superkings, Royals

Figura 3.1. Caminos y velocidad de transición en las marcas de cigarrillos, 2005–2021.



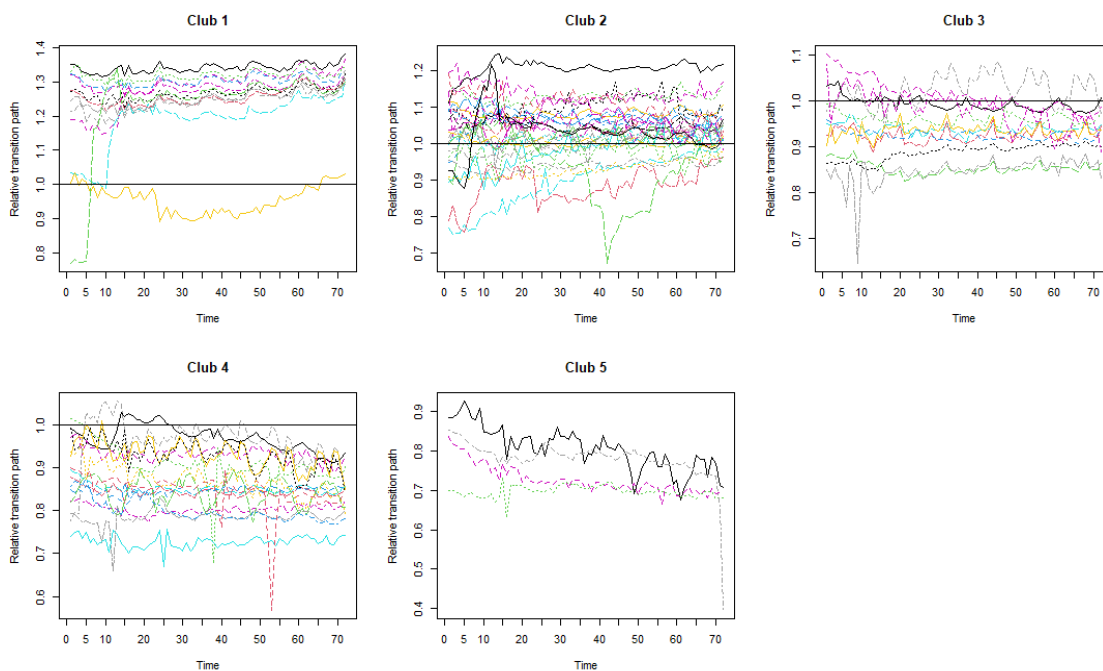
Por otro lado, la Tabla 3.3 y la Figura 3.2 muestran los resultados de aplicar el algoritmo de clustering para comprobar la existencia de convergencia de subgrupos para el primer subperiodo analizado (2005- 2010). Como puede observarse, la aplicación del algoritmo genera 5 clubes y 2 marcas como divergentes. De las 73 marcas de cigarrillos que se han comercializado en este periodo, 18, 28, 12 y 16 pertenecen a los clubes 1, 2, 3 y 4 respectivamente, mientras que el club 5 sólo incluye 4 marcas. Es en este subperiodo donde se observa una mayor heterogeneidad, algo que parece razonable de acuerdo con estudios previos sobre la efectividad de las leyes antitabaco en España, que indican que la Ley 42/2010 fue más efectiva que la Ley 28/2005. La trayectoria de transición del club 1 es creciente, mientras que en el caso de los clubes 2 y 3 se observa una trayectoria de transición mantenida. Los clubes 4 y 5 muestran una tendencia decreciente en las trayectorias de transición.

CAPÍTULO 3: CONVERGENCIA en las ventas oficiales de cigarrillos en España: un análisis de clústeres de 2005 a 2021

Tabla 3.3. Proceso del club de convergencia en las ventas de cigarrillos por marca (en cajetillas), 2005-2010.

Club	Número de marcas	β	s.e.	t-value	Marcas
1	12	0.034	0.19	0.18	Marlboro, Winston, Fortuna, Lucky Strike, Pall Mall, L&M, Burton, Chesterfield, Camel, Ducados Negro, Ducados Rubio, Nobel
2	28	0.051	0.242	0.212	John Player Sp., Elixyr, Gold Coast, R 1, Karelia, Ducal, Bn, Bullbrand, Philip 'Morris, Excite, Austin, Gauloises Rubio, News, Lambert & Butler, Vogue, Habanos, Popular, West, Silk Cut, Peter Stuyvesant, Winfield, Golden American, Coronas Negro, Benson & Hedges, Next, Superkings, Royals, Lark
3	12	0.077	0.094	0.818	Royal Crown, Celtas, Rex, Mayfair, Dunhill, Rothmans, Gitanes, Condal, Davidoff Rubio, Regal, Kool, Fine 120
4	16	0.108	0.151	0.712	Reales, Brooklyn, Bisonte, Sovereign, Sombra, Boncalo, Berkeley, More, Embassy, Craven A, Prince, Partagas, Romeo Y Julieta, Piper, Belga, Diana
5	4	0.239	0.252	0.948	Kensitas Club, Gold Leaf, Salem, Viceroy
Divergentes	2				Coronas Rubio, Bastos

Figura 3.2. Caminos y velocidad de transición en las marcas de cigarrillos, 2005–2010.



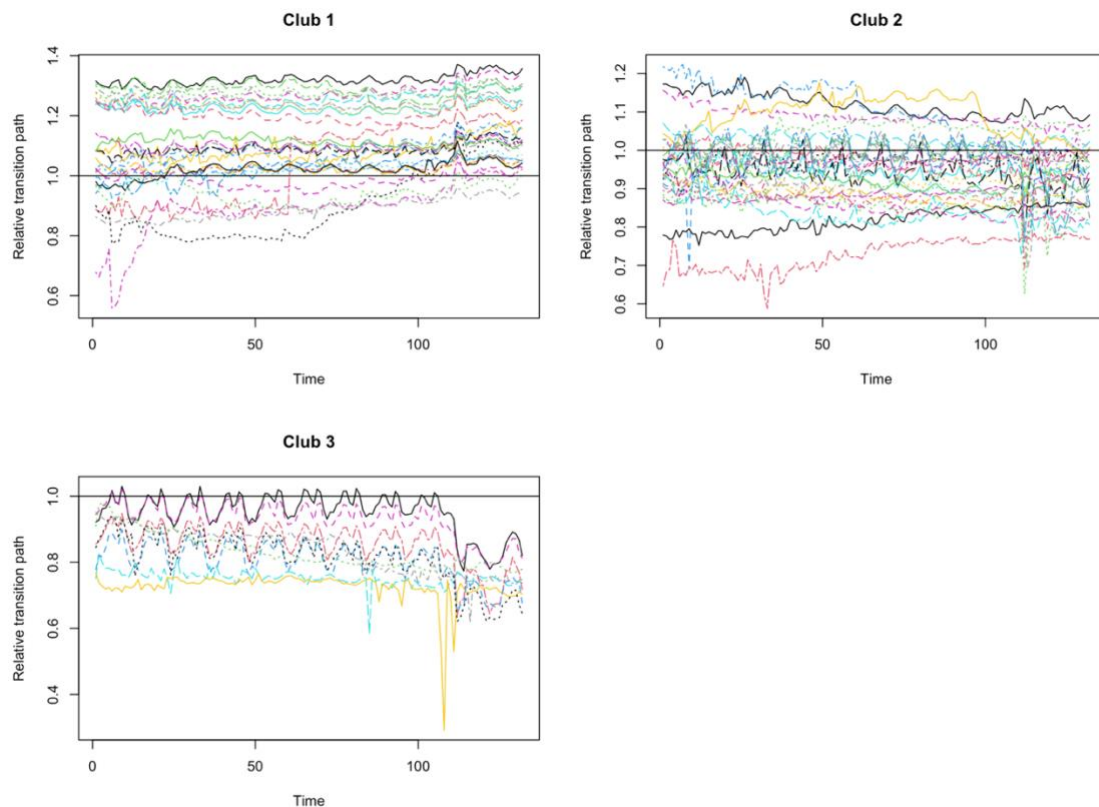
Por último, la Tabla 3.4 y la Figura 3.3 muestran los resultados de aplicar el algoritmo de agrupación para comprobar la existencia de convergencia de subgrupos para el segundo subperiodo analizado (2011-2021). En este caso, la aplicación del algoritmo genera 3 clubes. De las 67 marcas de cigarrillos que se han comercializado en este periodo, 28 y 30 pertenecen a los clubes 1 y 2, respectivamente, mientras que el club 3 incluye 9 marcas. La trayectoria de transición del club 1 es creciente, mientras que en el caso de los clubes 2 y 3 se observa una trayectoria de transición decreciente.

Tabla 3.4. Proceso del club de convergencia en las ventas de cigarrillos por marca (en cajetillas), 2011-2021.

Club	Número de marcas	β	s.e.	t-value	Marcas
1	28	0.244	0.029	8.479	Marlboro, Camel, Chesterfield, Fortuna, Lucky Strike, Rothmans, Philip Morris K/S, Austin, Ducal, American Legend, Winston, L&M, Nobel, Ducados Rubio, Ducados Negro, Nobel Style, Elixyr, Marlboro Pocket, Fortuna Red Line, Lucky Strike Blando, Karelia, R 1, Bullbrand, Benson & Hedges, Nobel Blando, Excite, Denim, Manitou
2	30	-0.08	0.0584	-1.364	Ducados Rubio Blando, John Player Sp., Benson & Hedges American, Gauloises Rubio, Bn, John Player Sp. Am.100's, Ducados Rubio 100's, Pueblo, Bravo, Vogue, John Player Sp. Black/Blue, Silk Cut, Winfield, Peter Stuyvesant, Burton, Pall Mall, News, Davidoff Rubio, Royal Crown, Dunhill, Desert Gold, Gitanes, Rex, Lambert & Butler, Natural American, Pepe, Gauloises Negro, Mayfair, Richmond, Superkings, Mohawk, John Player Sp. American, Natural American Spirit, Latino
3	9	-0.443	0.287	-1.544	Richmond, Superkings, John Player Sp.American, Black Devil, Latino, Regal, Mecanicos, Embassy, Royals

Así pues, las ventas de cigarrillos por marcas no son un club de convergencia homogéneo, sino que constituyen clubes heterogéneos de distintos tamaños, con trayectorias y velocidades de transición diferentes. Se trata de un dato no trivial, ya que sugiere que, mientras que se rechazó la hipótesis nula de convergencia global, dentro de este mismo grupo existen pruebas tanto de convergencia como de no convergencia.

Figura 3: Caminos y velocidad de transición en las marcas de cigarrillos, 2011–2021.



3.4 Conclusiones

En los últimos años, ha crecido el interés de académicos, profesionales y responsables políticos por encontrar herramientas de política económica para restringir el consumo de tabaco. Ese control es crucial no sólo por el gran impacto que el consumo de tabaco produce en la salud, sino también por sus efectos en los presupuestos de los países a través de la recaudación de impuestos y los costes sanitarios. La publicidad, la promoción y el patrocinio del tabaco pueden influir en el consumo de tabaco. En este contexto, la literatura empírica dedicada al análisis de la eficacia de las herramientas de política económica es heterogénea en cuanto al efecto de la prohibición de la publicidad, la promoción y el patrocinio sobre el consumo de tabaco. Además, pocos estudios analizan el efecto de dos leyes que restringen la publicidad, la promoción y el patrocinio del tabaco desde la perspectiva de las ventas de cigarrillos por marca. En este artículo, se ha explorado el efecto de las restricciones de 2005 y 2010 sobre la evolución de las ventas de cigarrillos por marca de una forma novedosa que permite analizar la convergencia en las ventas de todas las marcas comercializadas en España.

En nuestro trabajo, aplicando técnicas que comprueban la existencia de clubes de convergencia, revelamos que las leyes de 2005 y 2010 no han conseguido homogeneizar

el mercado promoviendo la convergencia absoluta. Aunque se rechaza la existencia de convergencia absoluta para todos los periodos estudiados, encontramos efectos diferentes de las leyes de 2005 y 2010. Nuestros resultados sugieren que, en España, de 2005 a 2010 las ventas de marcas se comportaron de forma más heterogénea que a partir de 2010, donde, aunque no existe convergencia absoluta, sólo se encuentran 3 clubes. Este hallazgo apoya lo establecido en la literatura previa que indica que la Ley 42/2010 proporcionó una especificación más restrictiva de la prohibición que la Ley 28/2005. Sin embargo, parece que no se ha conseguido la convergencia o neutralidad buscada por el Gobierno, algo coherente con un estudio reciente que indica que "las empresas tabaqueras han aprovechado los vacíos legales para seguir promocionando sus productos" (Moodie et al., 2022).

Aunque la prohibición de la publicidad, promoción y patrocinio del tabaco en España no ha generado una homogeneización o neutralidad absoluta, sí se observa un comportamiento heterogéneo en los distintos subperiodos. Por un lado, la Ley 28/2005 no impedía a las tabaqueras diferenciar sus marcas y, por tanto, el comportamiento de las ventas era muy heterogéneo. Por otro lado, la ley de 2011, que incluye una especificación de la prohibición más restrictiva que la Ley 28/2005, aunque no ha logrado una convergencia absoluta, ha producido una mayor homogeneización. Esto parece demostrar la eficacia de las restricciones a la publicidad, promoción y patrocinio del tabaco. De acuerdo con estos resultados, parece que se cumplen dos hipótesis: el punto de vista de la regulación española y la teoría económica asociada a la regulación de bienes que causan daño a la salud. Por un lado, la ley de 2011 ha sido más eficaz que la Ley 28/2005. Por otro lado, los resultados sugieren que cuanto más restrictiva es la regulación sobre publicidad, promoción y patrocinio del tabaco, más homogeneización o neutralidad se consigue.

Para resumir las implicaciones políticas prácticas que los resultados de este trabajo tienen para los responsables políticos, éstas pueden resumirse en tres. La primera implicación es que los gobiernos tienen en las leyes un instrumento para controlar las ventas legales de cigarrillos. Cuanto más restrictivas sean las limitaciones a la publicidad, la promoción y el patrocinio del tabaco, más limitada será la capacidad de los fabricantes de tabaco para diferenciarse en el mercado. En segundo lugar, los responsables políticos deberían tener en cuenta que las restricciones a la publicidad, la promoción y el patrocinio del tabaco, aunque limitan la capacidad de los fabricantes para diferenciarse, no provocan una convergencia total en la que todas las marcas sean de "talla única" para los consumidores. Por último, parece que las leyes que incluyen limitaciones a la publicidad, la promoción y el patrocinio del tabaco generan una especie de "efecto de congelación" que proporciona ventajas a las marcas mejor posicionadas justo en el momento en que las leyes entran en vigor.

Una extensión natural de este trabajo para futuras investigaciones sería comprobar la convergencia de los productos sustitutivos de los cigarrillos, lo que nos daría respuesta a si la mayor convergencia observada a partir de 2010 está relacionada con el comportamiento de otros productos sustitutivos, que incluyen productos del tabaco legales, como la picadura fina o el tabaco de pipa, e incluso productos ilegales. Además, otra ampliación puede ser analizar si la no convergencia observada en este trabajo es

CAPÍTULO 3: CONVERGENCIA en las ventas oficiales de cigarrillos en España: un análisis de clústeres de 2005 a 2021

diferente en función del grupo de interés (como jóvenes, hombres, mujeres o desempleados, por ejemplo). Otra línea futura puede ser incluir el efecto del comercio transfronterizo y el contrabando. Por último, un análisis espacial que permita estudiar la convergencia por regiones también puede añadir valor a la literatura existente.

Capítulo 4

Evolución socioeconómica y ventas de productos del tabaco en España: un análisis a largo plazo de la Curva de Kuznets del Tabaco (TKC)

4.1 Introducción

Comprender la relación entre el desarrollo económico y la salud pública es crucial para crear políticas eficaces, especialmente aquellas centradas en reducir el uso de sustancias nocivas como el tabaco. Estas políticas pueden implicar estrategias como aumentar impuestos y precios (Chaloupka, 1999; Bader et al., 2011; Martín-Álvarez et al., 2020a), implementar regulaciones para lugares de trabajo libres de humo (Fichtenberg y Glantz, 2002; Pinilla et al., 2019; Del Arco-Osuna et al., 2023), realizar campañas de educación sanitaria y hacer cumplir restricciones a la publicidad (Kasza et al., 2011; Almeida et al., 2021; Golpe et al., 2022). La situación económica de las personas también influye en el comportamiento de fumar (Hiscock et al., 2011), por lo que examinar cómo el crecimiento económico afecta el consumo de tabaco puede ayudar a los formuladores de políticas a crear intervenciones que sean más efectivas en diferentes etapas del desarrollo económico.

CAPÍTULO 4: EVOLUCIÓN socioeconómica y ventas de productos del tabaco en España: un análisis a largo plazo de la Curva de Kuznets del Tabaco (TKC)

La curva de Kuznets (Kuznets, 1955) se formuló originalmente para describir la relación entre la desigualdad del ingreso y el crecimiento económico. Sugiere que la desigualdad inicialmente aumenta y luego disminuye a medida que crecen las economías. Posteriormente, este concepto se amplió a otras asociaciones entre diferentes variables, como la progresión de las tasas de contaminación. Esta adaptación empírica, conocida como “curva ambiental de Kuznets” (EKC), fue aplicada por primera vez por el Banco Mundial (1992) (World Bank, 1992) y por Grossman y Krueger (1995). Su idea central es que a medida que las economías avanzan, contribuyen a la contaminación atmosférica, pero a medida que aumentan los niveles de ingresos, la contaminación tiende a disminuir (Congregado et al, 2016; Iglesias et al., 2013; Robalino-López et al., 2015). De manera similar, la curva de Kuznets se ha utilizado para analizar la relación entre el desarrollo económico y cuestiones sociales como la desigualdad de género (Duflo, 2012), las tasas de criminalidad (Fajnzylber, et al. 2002) y el acceso a la educación (Castello-Climent y Domenech, 2014). Estas aplicaciones resaltan la versatilidad de la curva de Kuznets para comprender cómo el crecimiento económico se cruza con diversas dimensiones del progreso social, proporcionando un marco valioso para los responsables de la formulación de políticas en diferentes ámbitos.

Estudios recientes han ampliado este marco al consumo de bienes con importantes implicaciones para la salud pública, como el tabaco y el alcohol (Grecu y Rotthoff, 2015; Cantarero-Prieto et al., 2019; Cantarero-Prieto et al., 2023). Sin embargo, a medida que los ingresos siguen aumentando y la salud se convierte en un bien normal, la gente tiende a orientar su consumo hacia alimentos más saludables. Al reflexionar sobre la relación entre salud, desigualdad y desarrollo económico, Deaton (2003) concluyó que es crucial centrarse en el papel del ingreso o factores relacionados, como la educación, la riqueza o el estatus social, en la promoción de la salud. Las aplicaciones contemporáneas de la curva de Kuznets a la salud siguen siendo valiosas. Por ejemplo, Costa-Font et al. (2018) exploraron la existencia de una curva de Kuznets en salud y descubrieron que las desigualdades en salud relacionadas con los ingresos aumentan con el PIB per cápita, pero eventualmente disminuyen una vez que se alcanza un cierto nivel de desarrollo económico. Una revisión reciente de la literatura realizada por Patterson (2023) revela un importante conjunto de investigaciones multidisciplinarias que exploran la conexión entre el crecimiento económico y la salud de la población. La revisión señala que muchos investigadores encuentran que la relación crecimiento-salud varía dependiendo de varios factores: (i) el aspecto específico de la "salud" que se examina, (ii) la naturaleza de la relación, ya sea positivamente lineal o logarítmica, (iii) consideraciones de tiempo, como si el crecimiento se observa a corto o largo plazo, (iv) un enfoque en las desigualdades en salud versus los promedios de la población general, y (v) relaciones multivariadas que involucran factores adicionales. El estudio concluye que el uso de enfoques multivariados puede proporcionar una comprensión más completa de cómo se vincula el crecimiento económico con la salud. Por lo tanto, la revisión sugiere que avanzar en este campo de investigación puede beneficiarse del análisis simultáneo de múltiples factores.

Las investigaciones más avanzadas actuales sugieren una relación positiva entre ingresos y salud tanto dentro como entre países (Marmot 1999), que abarca diferentes grupos de edad (Case et al., 2002) y tasas de mortalidad (Preston 1975). Además, Ruhm (2000) descubrió que las recesiones económicas podrían mejorar los resultados de salud. Ruhm analizó los efectos de las recesiones económicas en la curva de Kuznets considerando las tasas de desempleo. Sin embargo, en un estudio posterior, encontró que estos resultados variaban en diferentes períodos de tiempo (Ruhm 2015). En el caso de España, si bien se observa una relación positiva entre ingresos y ventas de tabaco, existe una fuerte asimetría en periodos de expansiones y recesiones (Martín-Álvarez et al., 2020a). Además, mientras en las fases de expansión las ventas de tabaco se ven afectadas por los ingresos, en las recesiones es el desempleo el indicador que provoca las caídas en las ventas de tabaco (Martín-Álvarez et al., 2020b). Por este motivo, parece imprescindible utilizar el PIB y el desempleo como indicadores del ciclo económico a la hora de analizar el comportamiento de las ventas de tabaco.

Muchos académicos han explorado la conexión entre ingresos y salud utilizando el marco de la curva de Kuznets, lo que ha dado lugar a una extensa literatura. Por ejemplo, Marmot (1999) concluyó que un nivel socioeconómico más bajo se asocia con una salud más deficiente, y viceversa. Case et al. (2002) demostraron que los ingresos del hogar desempeñan un papel crucial en la salud de los niños, mostrando una relación positiva. Grecu y Rotthoff (2015) examinaron el vínculo entre los ingresos y la obesidad a través de una curva de Kuznets, denominada "Curva de Kuznets de la obesidad". Además, Fotourehchi y Çalışkan (2018) investigaron la correlación entre el crecimiento económico y los resultados de salud en 60 países en desarrollo y no encontraron una curva de Kuznets específica para los resultados de salud. Esto sugiere que las políticas de salud deben adaptarse a cada indicador de salud específico. En este contexto, parece que cada indicador de salud puede presentar comportamientos heterogéneos, por lo que analizar el mercado del tabaco únicamente a partir de la venta de cigarrillos puede arrojar resultados sesgados.

En el contexto específico del consumo de tabaco, la aplicación del marco de la Curva de Kuznets nos permite investigar cómo el desarrollo económico impacta las conductas de fumar y las ventas de tabaco. Las primeras etapas del crecimiento económico podrían conducir a un mayor consumo de tabaco debido a un mayor ingreso disponible y una conciencia limitada sobre la salud pública. Sin embargo, a medida que las economías continúan desarrollándose, una mayor educación, políticas de salud pública más sólidas y una mayor conciencia sobre la salud podrían conducir a una disminución en el consumo de tabaco. Este estudio busca explorar si las ventas de tabaco siguen este patrón hipotético en forma de U invertida, analizando cómo diversos factores socioeconómicos influyen en esta relación considerando diferentes tipos de productos de tabaco como cigarrillos, cigarrillos para liar (RYO), tabaco de pipa y puros.

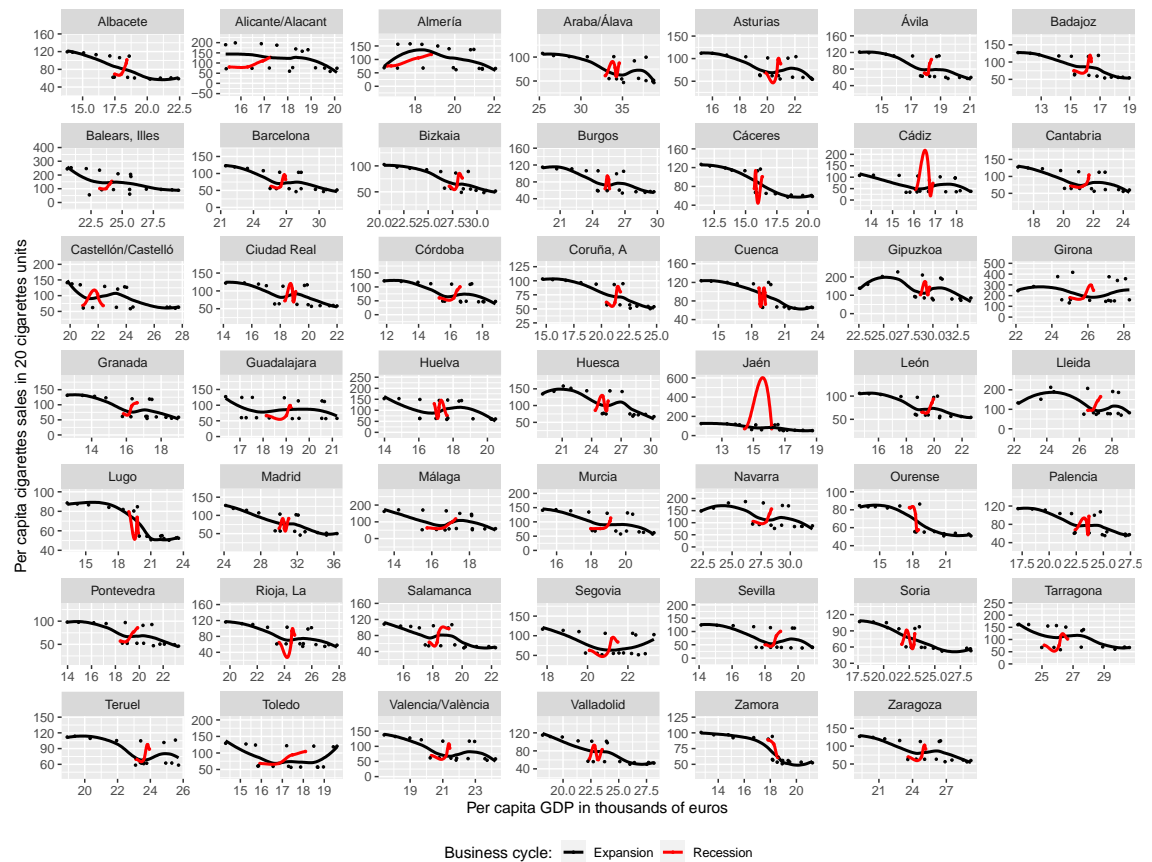
En este contexto, planteamos la hipótesis de que las ventas de tabaco aumentan con el crecimiento económico durante las primeras etapas de desarrollo, alcanzan su punto máximo y luego disminuyen a medida que se alcanzan niveles más altos de conciencia socioeconómica. El campo de la economía de la salud proporciona un telón de fondo crucial para comprender estas dinámicas, ya que explora los comportamientos

CAPÍTULO 4: EVOLUCIÓN socioeconómica y ventas de productos del tabaco en España: un análisis a largo plazo de la Curva de Kuznets del Tabaco (TKC)

económicos que sustentan los resultados de salud, incluidos los patrones de consumo y la efectividad de las intervenciones de salud pública. La economía de la salud ha reconocido desde hace mucho tiempo que los incentivos económicos, las estructuras de mercado y los entornos regulatorios tienen un impacto significativo en la salud pública (Glied y Smith, 2011). Por ejemplo, a medida que los países se desarrollan económicamente, el aumento del ingreso disponible puede conducir inicialmente a un mayor consumo de bienes nocivos como el tabaco. Sin embargo, con el tiempo, a medida que aumenta el estatus de la sociedad y las iniciativas de salud pública se vuelven más prevalentes, el comportamiento del consumidor tiende a cambiar, y las personas se vuelven más conscientes de su salud y reducen su consumo de tabaco. Esta perspectiva económica es esencial para analizar cómo las diferentes etapas del desarrollo económico interactúan con las medidas de salud pública para influir en las ventas de tabaco. En el caso de las provincias españolas, el Figura 4.1 muestra la relación entre el PIB y las ventas de cigarrillos de 2002 a 2021, observándose una relación negativa entre ambas variables en las expansiones económicas. Esto puede ser un indicador de que se alcanzan niveles más altos de conciencia socioeconómica.

Investigaciones anteriores de Grecu y Rotthoff (2015) y Cantarero-Prieto et al. (2019, 2023) ha demostrado patrones similares para el consumo de alcohol, lo que sugiere que el desarrollo económico y social altera significativamente las conductas de consumo. Sin embargo, la aplicación de la curva de Kuznets al consumo de tabaco sigue siendo relativamente inexplorada, particularmente cuando se considera una amplia gama de productos de tabaco. Este estudio pretende llenar este vacío realizando un análisis exhaustivo utilizando un panel de datos de las provincias españolas.

Figura 4.1. Relación entre el PIB y las ventas de cigarrillos por provincia.



Hasta donde sabemos, no hay muchos estudios que examinen los vínculos entre los ingresos y los estilos de vida poco saludables, como el consumo de tabaco, utilizando múltiples factores socioeconómicos simultáneamente, incluyendo el PIB y el desempleo como indicadores del ciclo económico y analizando diferentes tipos de productos de tabaco. Los hallazgos de esta investigación contribuirán a una comprensión más profunda de los factores socioeconómicos que impulsan los patrones de consumo de tabaco, ofreciendo información valiosa tanto para la teoría económica como para las políticas de salud pública. Al dilucidar las condiciones bajo las cuales las ventas de tabaco podrían disminuir a medida que crecen las economías, este estudio tiene como objetivo informar el diseño de intervenciones más efectivas para reducir el consumo de tabaco y sus cargas de salud asociadas.

4.2 Datos y Metodología

Los datos de ventas de productos del tabaco proceden del Ministerio de Hacienda y Función Pública de España, mientras que los datos de las variables explicativas socioeconómicas han sido extraídos del Instituto Nacional de Estadística de España. Los datos han sido observados de las 48 provincias que componen el territorio español

CAPÍTULO 4: EVOLUCIÓN socioeconómica y ventas de productos del tabaco en España: un análisis a largo plazo de la Curva de Kuznets del Tabaco (TKC)

donde se comercializa tabaco en el mercado regulado durante el periodo 2002-2021 (960 observaciones). En la tabla 4.1 se muestran los estadísticos descriptivos de las variables utilizadas.

Tabla 4.1. Estadísticos descriptivos de los datos utilizados.

Variable	Media	SD	Min	Max	Q1	Q2	Q3
Cigarettes	90.0	43.0	33.6	416.0	59.3	76.9	111.0
RYO	0.0037	0.0022	0.0008	0.0166	0.0021	0.0033	0.0046
Pipe	0.0005	0.0002	0.0001	0.0019	0.0003	0.0004	0.0006
Cigars	0.2050	0.0889	0.0560	0.6120	0.1340	0.1950	0.2570
GDP	21.5	4.92	11.4	38.8	17.9	20.5	24.8
Unemployment rate	15.6	7.48	2.87	42.3	9.95	14.2	20.1
Life expectancy	82.0	1.41	77.7	84.9	81.0	82.2	83.0
Aging ratio	1.51	0.563	0.718	3.40	1.10	1.35	1.82
Youth ratio	0.209	0.0274	0.143	0.267	0.189	0.211	0.230

Para responder a las preguntas de investigación propuestas, utilizamos un método de datos de panel para estimar la relación entre variables sociodemográficas y las ventas de productos de tabaco. Para ello, seguimos la especificación del modelo propuesto en la literatura previa de Economía de la Salud donde se analiza la curva de Kuznets para el consumo de tabaco y alcohol (Grecu & Rotthoff, 2015; Cantarero-Prieto et al., 2019; Cantarero-Prieto et al., 2023):

$$Tobacco\ sales_{st} = \beta_1 income_{st} + \beta_2 (income_{st})^2 + X' \delta + \gamma_s + \tau_t + \epsilon_{st}$$

Nos centramos en el impacto de los ingresos β_1 y β_2 . X' es un vector de factores socioeconómicos de los resultados de salud que varían en el tiempo. Para controlar las diferencias en las ventas de tabaco que son comunes a personas de la misma región, incluimos los efectos fijos regionales, γ_s . Los efectos fijos anuales, τ_t , absorben cualquier variación temporal en las ventas de tabaco, por ejemplo, cambios en las políticas de salud (Del Arco-Osuna, et al., 2023).

Una vez establecida la especificación del modelo a utilizar, en la Tabla 4.2 se muestra la estrategia empírica seguida para asegurar la robustez de los resultados. Para ejecutar el framework computacional se ha utilizado el lenguaje R, así como algunas librerías específicas para aplicar este tipo de metodologías (Croissant & Millo G, 2008; Hyndman & Khandakar, 2008; Pfaff, 2008; R Core Team, 2022; Trapletti & Hornik, 2020; Wickham et al., 2019; Wickham & Bryan, 2023).

Tabla 4.2. Estrategia empírica utilizada.

	Proceso	Hipótesis
Paso 1	Pesaran CIPS test	¿Son estacionarias las series temporales utilizadas?
Paso 2	Test de cointegración de Pedroni	¿Están cointegradas las series temporales utilizadas?
Paso 3	Hausman test	¿Cuál es el modelo más apropiado dadas las características de los datos utilizados? (Pooled Ordinary Least Squares, Fixed Effects, or Random Effects)
Paso 4	Modelo de datos de panel	¿Se cumple la Kurva de Kuznets para las ventas de todos los productos del tabaco analizados? (Cigarrillos, cigarrillos de liar, pipa y puros)
Paso 5	Modelo de datos de panel	¿Hay relación entre las tasas de desempleo y las ventas de todos los productos del tabaco analizados? (Cigarrillos, cigarrillos de liar, pipa y puros)

El primer paso a realizar es el test CIPS (Pesaran, 2007) para comprobar si las series temporales a utilizar son estacionarias. Esto es importante porque las estimaciones de regresiones con variables no estacionarias son espurias a menos que estén cointegradas. La prueba transversal Im-Pesaran-Shin (CIPS) de Pesaran se utiliza para determinar la presencia de raíces unitarias en series de datos de panel teniendo en cuenta la dependencia transversal. La prueba es una extensión de la prueba Im-Pesaran-Shin (IPS), modificada para acomodar posibles dependencias transversales restando la media transversal. Considere un modelo de datos de panel de la forma:

$$y_{it} = \rho_i y_{it-1} + \sum_{j=1}^p \phi_{ij} \Delta y_{it-j} + \epsilon_{it}$$

donde: y_{it} es el valor de la variable para la entidad i en el momento t , ρ_i es el coeficiente autorregresivo, ϕ_{ij} son los coeficientes de las diferencias rezagadas, ϵ_{it} es el término de error, que se supone es i.i.d. La estadística de prueba CIPS se calcula promediando las estadísticas transversales del ADF:

$$CIPS = \frac{1}{N} \sum_{i=1}^N CADF_i$$

donde $CADF_i$ es el estadístico ADF transversal para la entidad i :

$$CADF_i = \frac{y_{it} - \bar{y}_t}{\sigma_y}$$

La hipótesis nula de la prueba CIPS es que todos los paneles contienen una raíz unitaria, mientras que la hipótesis alternativa es que al menos un panel es estacionario.

CAPÍTULO 4: EVOLUCIÓN socioeconómica y ventas de productos del tabaco en España: un análisis a largo plazo de la Curva de Kuznets del Tabaco (TKC)

Si el estadístico CIPS es menor que el valor crítico en un nivel de significación elegido, se rechaza la hipótesis nula, lo que indica la presencia de estacionariedad en al menos un panel. Si el estadístico CIPS es mayor que el valor crítico, no se puede rechazar la hipótesis nula, lo que sugiere la presencia de raíces unitarias en todos los paneles.

Si no todas las series temporales utilizadas son estacionarias, el segundo paso consiste en realizar el test de cointegración de datos de panel de Pedroni (1999, 2004). Este método se usa ampliamente para probar la cointegración en paneles heterogéneos, donde tanto los coeficientes de intersección como de pendiente pueden variar entre unidades transversales individuales. El marco de Pedroni permite múltiples formas de heterogeneidad y dependencia transversal, lo que lo hace particularmente sólido en aplicaciones empíricas con datos de panel.

La prueba de Pedroni comprende siete estadísticas de prueba diferentes, agrupadas en dos categorías: pruebas de panel y pruebas de grupo. Las pruebas basadas en panel suponen homogeneidad entre las unidades de la estructura autorregresiva, mientras que las pruebas basadas en grupos permiten heterogeneidad en la dinámica de las unidades individuales. Estas siete estadísticas se basan en diferentes estimadores de los residuos de la relación de cointegración hipotética: (i) Estadístico de panel v (nipanel): prueba la varianza de los residuos en el panel combinado, similar a la prueba de Phillips-Perron. (ii) Estadístico ρ del panel (rhopanel): basado en el parámetro autorregresivo del panel. (iii) Estadístico t no paramétrico de panel (tpanelnonpar): estadístico t para los residuos agrupados sin supuestos paramétricos. (iv) Estadístico t paramétrico de panel (tpanelpar): un estadístico t para los residuos agrupados suponiendo una estructura AR(1). (v) Estadístico ρ de grupo (rogrupo): basado en el parámetro autorregresivo para cada unidad transversal individualmente. (vi) Estadístico t no paramétrico de grupo (tgroupnonpar): estadístico t no paramétrico calculado para unidades individuales. (vii) Estadístico t paramétrico de grupo (tgrouppar): un estadístico t paramétrico calculado para cada unidad de sección transversal.

La hipótesis nula en todas las pruebas es que no hay cointegración (es decir, no hay relación de largo plazo) entre las variables. La hipótesis alternativa difiere ligeramente entre los dos grupos de pruebas: (i) para las pruebas de panel, la alternativa es que las variables estén cointegradas para todas las unidades transversales; (ii) para las pruebas basadas en grupos, la alternativa es que exista cointegración para al menos una de las unidades transversales.

Cada una de los estadísticos de prueba se evalúa comparando los valores estandarizados con los valores críticos de la distribución normal estándar. Para los estadísticos t de panel y de grupo, el rechazo de la hipótesis nula se produce si el valor estandarizado es significativamente diferente de cero, normalmente utilizando un nivel de significación del 5% (es decir, $\pm 1,96$ para una prueba de dos colas).

Al aplicar estos test, podemos evaluar de manera sólida si las variables en nuestros datos de panel están cointegradas, incluso en presencia de heterogeneidad o dependencia transversal.

Si las series son estacionarias o integradas, el tercer paso es elegir qué método de estimación es más apropiado en este caso (Mínimos Cuadrados Ordinarios Agrupados, Efectos Fijos o Efectos Aleatorios). Para ello utilizaremos el test de Hausman (Hausman, 1978). La prueba de Hausman es un procedimiento estadístico utilizado para evaluar la consistencia de estimadores en modelos de datos de panel. Específicamente, se utiliza para comparar los estimadores de efectos fijos (FE) y efectos aleatorios (RE) para determinar qué modelo es más apropiado. La prueba evalúa si los errores únicos (efectos aleatorios) están correlacionados con los regresores, lo que violaría los supuestos del modelo de efectos aleatorios. La prueba de Hausman evalúa la hipótesis nula de que el modelo de efectos aleatorios es apropiado (es decir, los efectos aleatorios no están correlacionados con los regresores) frente a la alternativa de que el modelo de efectos fijos es preferible debido a la correlación. El estadístico de la prueba de Hausman se basa en la diferencia entre los estimadores FE y RE:

$$H = (\widehat{\beta}_{RE} - \widehat{\beta}_{FE})' [\text{Var}(\widehat{\beta}_{FE}) - \text{Var}(\widehat{\beta}_{RE})]^{-1} (\widehat{\beta}_{RE} - \widehat{\beta}_{FE})$$

donde β_{RE} y β_{FE} son los estimadores de efectos aleatorios y efectos fijos, respectivamente, $\text{Var}(\beta_{FE})$ y $\text{Var}(\beta_{RE})$ son las matrices de varianza-covarianza correspondientes de los estimadores.

El estadístico H sigue una distribución chi-cuadrado con k grados de libertad, donde k es el número de regresores.

4.3 Resultados

Siguiendo la estrategia econométrica descrita anteriormente, en esta sección presentamos los resultados del análisis de la Curva de Kuznes del Tabaco. En primer lugar, la Tabla 4.3 muestra los resultados de la prueba CPIS para las series temporales utilizadas. En segundo lugar, la Tabla 4.4 muestra los resultados de las pruebas de cointegración de Pedroni. Finalmente, la Tabla 4.5 muestra los resultados del modelo de efectos fijos para medir el impacto de los factores económicos y demográficos en las ventas de tabaco.

Para determinar si las series temporales están cointegradas se utiliza el criterio del valor p. Sin embargo, el criterio para determinar si las series están cointegradas es la comparación del estadístico de prueba con los valores críticos de Bootstrap. Los resultados de la Tabla 4.3, que presenta la prueba CIPS para raíces unitarias en varias variables, indican propiedades de estacionariedad mixtas. Las variables RYO, cigarrillos, PIB, esperanza de vida, índice de envejecimiento e índice de juventud no logran rechazar la hipótesis nula de raíces unitarias, lo que sugiere que son estacionarias, a pesar de que los valores p rondan el nivel del 10%. Por el contrario, se considera que cigarrillos, tabaco de pipa y tasa de desempleo no son estacionarias, dados sus valores p significativos y sus estadísticas de prueba. La Tabla 4.4 presenta los resultados de la

CAPÍTULO 4: EVOLUCIÓN socioeconómica y ventas de productos del tabaco en España: un análisis a largo plazo de la Curva de Kuznets del Tabaco (TKC)

prueba de cointegración de Pedroni, que rechaza consistentemente la hipótesis nula de ausencia de copointegración. Este fuerte rechazo sugiere la presencia de relación de cointegración entre las variables, lo que indica conexiones de equilibrio a largo plazo en el conjunto de datos. En conjunto, los hallazgos de ambas tablas resaltan problemas de estacionariedad a corto plazo en algunas variables, mientras que la prueba de cointegración sugiere relaciones sólidas a largo plazo.

Tabla 4.3. Resultados de la prueba CIPS para raíces unitarias en las variables utilizadas.

Variable	Estadístico del CIPS test	p-valor	Retardo	Conclusión
Cigarettes	-3.310	0.099	2	No estacionaria
RYO	-2.315	0.112	2	Estacionaria
Pipe	-2.617	0.064	2	No estacionaria
Cigars	-2.531	0.120	2	Estacionaria
GDP	-2.248	0.114	2	Estacionaria
Unemployment rate	-2.680	0.040	2	No estacionaria
Life expectancy	-2.429	0.105	2	Estacionaria
Aging ratio	-1.022	0.114	2	Estacionaria
Youth ratio	-2.539	0.119	2	Estacionaria

Tabla 4.4. Resultados de la prueba de cointegración de Pedroni.

Test	Empirical Statistic	Standardized Statistic	Bootstrap critical values			Conclusion
			10%	5%	1%	
Panel v -statistic	590.28	31.45	-2.58	-1.96	-1.28	Reject H_0
Panel ρ -statistic	-0.98	15.94	-2.58	-1.96	-1.28	Reject H_0
Panel non-parametric t-statistic	1.51	11.51	-2.58	-1.96	-1.28	Reject H_0
Panel parametric t-statistic	0.11	29.35	-2.58	-1.96	-1.28	Reject H_0
Group ρ -statistic	-1.15	27.65	-2.58	-1.96	-1.28	Reject H_0
Group non-parametric t-statistic	0.37	21.43	-2.58	-1.96	-1.28	Reject H_0
Group parametric t-statistic	-0.13	20.90	-2.58	-1.96	-1.28	Reject H_0

Tabla 4.5. Resultados de las estimaciones de los modelos de efectos fijos.

Variables independientes	Variables dependientes				
	Cigarettes	RYO	Pipe	Cigars	
GDP	-1.137* (0.506)				
GDP ²	0.572* (0.254)				
Unemployment rate		-0.247*** (0.015)	0.243*** (0.015)	0.243*** (0.016)	0.244*** (0.015)
Life expectancy	-1.404*** (0.415)	-0.900*** (0.033)	-1.124*** (0.032)	-1.125*** (0.033)	-1.05*** (0.032)
Aging ratio	-1.141*** (0.078)	-1.607*** (0.079)	1.275*** (0.076)	0.275*** (0.077)	0.276*** (0.077)
Youth ratio	-0.893*** (0.143)	-0.952*** (0.0105)	0.740*** (0.102)	1.741*** (0.103)	0.741*** (0.102)
Observations	960	960	960	960	960
R ²	0.812	0.843	0.601	0.624	0.604

Notas: ***, ** y * indican significatividad al 1%, 5% y 10%, respectivamente.

Nuestra estrategia econométrica se aplica para encontrar evidencia de la Curva de Kuznets del Tabaco en las ventas de los diferentes productos del tabaco comercializados en España, que es una de las principales aportaciones de nuestro trabajo. En este sentido, analizamos la posibilidad de que la relación entre las ventas de productos del tabaco y el PIB tenga forma de U invertida. Además, nos interesa analizar el efecto del desempleo como variable explicativa de las ventas de productos del tabaco. Los resultados de la estimación de efectos fijos revelan relaciones significativas entre las variables independientes y los diferentes tipos de consumo de tabaco (cigarrillos, tabaco para liar, tabaco de pipa y puros). El PIB y su término cuadrático (PIB²) muestran un efecto significativo y no lineal sobre el consumo de cigarrillos, con un coeficiente negativo para el PIB (-1,137, significativo al 10%) y un coeficiente positivo para el PIB² (0,572, significativo al 10%). lo que sugiere que la curva de Kuznets del tabaco se cumple en el caso de los cigarrillos.

La tasa de desempleo tiene un efecto muy significativo y positivo en todos los tipos de tabaco, excepto en los cigarrillos. Para RYO, pipa y cigarros, el coeficiente de la tasa de desempleo es positivo y significativo del 1%, lo que indica que un mayor desempleo conduce a un aumento en estas formas de consumo de tabaco. Estos resultados sugieren que en las recesiones económicas se utilizan productos alternativos a los cigarrillos como productos sustitutos.

La esperanza de vida muestra sistemáticamente un efecto negativo y muy significativo en todos los tipos de consumo de tabaco, lo que sugiere que una mayor esperanza de vida se asocia con un menor consumo en todas las categorías. De manera similar, la proporción de envejecimiento se asocia negativamente con el consumo de cigarrillos y de ryo, pero positivamente con el consumo de pipa y cigarros, lo que indica diferentes patrones de consumo de tabaco entre las poblaciones de mayor edad.

CAPÍTULO 4: EVOLUCIÓN socioeconómica y ventas de productos del tabaco en España: un análisis a largo plazo de la Curva de Kuznets del Tabaco (TKC)

Finalmente, la proporción de jóvenes tiene un efecto negativo significativo en el consumo de cigarrillos y de tabaco licuado, pero un impacto positivo significativo en el uso de tabaco de pipa y cigarros, lo que indica que las poblaciones más jóvenes tienden a consumir formas menos tradicionales de tabaco, pero consumen alternativas más modernas como RYO. El alto R^2 , que oscila entre 0,601 y 0,843, sugiere un buen ajuste para los modelos, especialmente para cigarrillos y RYO.

En resumen, nuestros resultados garantizan que la Curva de Kuznets del Tabaco sólo se observa en el caso de los cigarrillos, lo que está en línea con la literatura previa en la que se afirma que el crecimiento del PIB provoca un "efecto de concienciación" en la población fumadora (Martín-Álvarez et al., 2020b). Por otro lado, el resto de los productos del tabaco son utilizados por los consumidores como productos sustitutivos en fases de recesión económica. Esto también está en línea con la literatura previa, que indica que en las recesiones económicas es el desempleo el que provoca una disminución en las ventas de cigarrillos (Martín-Álvarez et al., 2020b). Es decir, en las recesiones económicas los consumidores recurren a alternativas más baratas y con falsas connotaciones "saludables" (Martín-Álvarez et al., 2023).

4.4 Conclusiones

En los últimos años, ha aumentado el interés entre académicos y responsables de políticas por identificar medidas económicas que reduzcan el consumo de tabaco. Este control es crucial, no sólo por el importante impacto del tabaco en la salud pública, sino también por sus implicaciones en los presupuestos nacionales, tanto en términos de ingresos fiscales como de costes sanitarios. Factores como el desarrollo económico, medido a través del PIB per cápita, y el nivel de desempleo, pueden influir en el consumo de tabaco. Sin embargo, la evidencia empírica disponible no llega a una conclusión definitiva sobre la relación entre los ciclos económicos y el consumo de tabaco. Además, hasta ahora muy pocos artículos se han preocupado por analizar conjuntamente el efecto del PIB, el desempleo y otros factores socioeconómicos sobre las ventas de diferentes productos del tabaco. Este artículo ofrece un enfoque novedoso al analizar cómo el PIB, el desempleo y otros factores socioeconómicos afectan las ventas de los diferentes productos del tabaco que existen.

En este artículo, aplicando el enfoque de datos de panel, mostramos que sólo la relación entre las ventas de cigarrillos y el PIB muestra una forma de U invertida, aceptando en ese caso la Curva de Kuznets del Tabaco. En el caso de el tabaco de liar, el tabaco de pipa y los cigarros, existen hallazgos relevantes respecto de la relación entre las ventas de estos productos de tabaco y las tasas de desempleo. Los resultados sugieren que estos tres tipos de productos de tabaco se utilizan como "productos refugio" en tiempos de recesión económica. Así, los hallazgos sugieren que, en España, las ventas de cigarrillos disminuyen a medida que se alcanzan mayores niveles de

conciencia socioeconómica, pero el efecto en la salud pública no es el deseado porque el tabaco de liar, el tabaco de pipa y los cigarrillos actúan como alternativas baratas y con falsas connotaciones “saludables”.

Aunque unos niveles más altos de conciencia socioeconómica tienen el efecto deseado sobre las ventas de cigarrillos en España, que se espera según la mayor parte de la literatura, este efecto se anula en parte por el efecto de otros productos de tabaco que actúan como impulsores del consumo del mismo. Por un lado, la población percibe que el cigarrillo es perjudicial para la salud y el desarrollo socioeconómico provoca una disminución en su consumo. Por otro lado, otros productos del tabaco existentes, que tienen un precio más bajo y connotaciones “menos nocivas” para la salud, actúan como elementos impulsores de la proliferación del tabaquismo.

Según estos resultados, parece que se cumplen dos hipótesis: el punto de vista epidemiológico y la teoría económica asociada al consumo de bienes normales. Por un lado, un aumento del PIB parece estar asociado con una mayor conciencia general sobre los efectos nocivos del tabaquismo para la salud. Por otro lado, los resultados sugieren que cuando se permite asociar productos de tabaco con connotaciones “poco nocivas” para la salud, aumenta la proliferación del tabaquismo, tal y como establece la literatura sobre Economía del Tabaco.

Por tanto, considerando el objetivo de este artículo, parece que los productos del tabaco tienen características que de alguna manera condicionan la relación entre sus ventas y la evolución socioeconómica. De esta forma, nos encontramos con una proliferación del tabaquismo, generado por productos distintos al cigarrillo, que no se revierte, contrariamente a lo que cabría esperar de la literatura.

Una de las formas generales que puede motivar este comportamiento es el esfuerzo que debe realizar el consumidor para comprar cigarrillos. En el periodo 2002-2021, el esfuerzo que deben realizar los consumidores para comprar cigarrillos ha crecido significativamente, tal y como establece la literatura previa (Cadahia et al., 2022). Otra vía general podría ser las connotaciones “menos nocivas” para la salud o “de moda” que tienen productos como el tabaco de liar sin aditivos y el tabaco de pipa de agua, respectivamente.

Parece que los resultados están en línea con artículos de la literatura para España que indican que la reacción de las ventas de cigarrillos en las recesiones, medida por el desempleo, es muy pronunciada. Esta reacción puede estar motivada precisamente por el uso de otros “productos de refugio”. En esta línea, los resultados confirman que en España se ha alcanzado un nivel de desarrollo socioeconómico en el que más PIB se asocia a una mayor conciencia sobre los efectos nocivos del consumo de tabaco. Además, los resultados también respaldan que, en las recesiones, medidas por el desempleo, los consumidores de cigarrillos buscan alternativas más baratas para seguir consumiendo tabaco. Por lo tanto, nuestros hallazgos podrían ser útiles para los formuladores de políticas en el diseño y la implementación oportuna de políticas antitabaco que beneficiarían tanto a los consumidores potenciales de drogas como al resto de la sociedad. Sin embargo, hay que considerar que el estudio tiene limitaciones,

CAPÍTULO 4: EVOLUCIÓN socioeconómica y ventas de productos del tabaco en España: un análisis a largo plazo de la Curva de Kuznets del Tabaco (TKC)

ya que nuestros resultados consideran datos afectados por una recesión económica particularmente fuerte.

Las investigaciones futuras podrían centrarse en analizar la influencia de estos resultados en los cambios regulatorios. Los gobiernos deberían considerar estos resultados heterogéneos en el desempeño de diferentes productos de tabaco al desarrollar políticas de control del tabaco porque las políticas homogéneas pueden conducir a resultados sesgados.

Capítulo 5

Conclusiones Integradoras

El análisis del mercado del tabaco en España, en el contexto del impacto de las políticas de control del tabaquismo, el desarrollo económico y la pandemia de COVID-19, arroja luz sobre una variedad de temas interrelacionados. A lo largo de esta investigación, se han explorado factores económicos y sociales, así como la efectividad de las herramientas tecnológicas avanzadas como el aprendizaje automático (ML). A continuación, se presentan las conclusiones integradoras en respuesta a los objetivos propuestos.

Uno de los principales objetivos del estudio fue analizar cómo el consumo de tabaco afecta tanto a la salud pública como a la economía de los países, específicamente en España. La investigación ha mostrado que el consumo de tabaco sigue siendo un factor clave de morbilidad y mortalidad, lo que genera enormes costes en los sistemas de salud debido al tratamiento de enfermedades relacionadas con el tabaco, como el cáncer, las enfermedades respiratorias y cardiovasculares. Estos costes están ampliamente documentados en la literatura, y se ha estimado que representan hasta el 11,5% del PIB en términos de gasto sanitario en algunos países (Papanicolas et al., 2018).

Por otro lado, los ingresos derivados de los impuestos al tabaco siguen siendo una fuente significativa de ingresos fiscales para el gobierno español, lo que convierte al tabaco en un producto de doble filo: altamente gravoso en términos de salud, pero necesario para mantener ciertos niveles de recaudación fiscal. Sin embargo, los gobiernos se enfrentan al reto de equilibrar la necesidad de ingresos fiscales con las

CAPÍTULO 5: CONCLUSIONES Integradoras

medidas para reducir el consumo de tabaco y, por ende, mejorar los indicadores de salud pública.

Otro de los objetivos fundamentales fue evaluar la efectividad de las políticas restrictivas aplicadas en España, particularmente las leyes de 2005 y 2010 que prohibieron la publicidad, promoción y patrocinio del tabaco. Estas políticas han tenido un impacto medible en la reducción de las ventas de cigarrillos y en la prevalencia de fumadores. Sin embargo, el análisis de las ventas de cigarrillos por marcas muestra que, si bien las políticas han limitado la capacidad de los fabricantes para diferenciar sus productos en el mercado, no han logrado una convergencia absoluta en términos de homogenización del mercado.

La ley de 2010, más restrictiva que la de 2005, ha generado mayores niveles de homogeneización en el mercado, con menos diferenciación entre las marcas de cigarrillos. Este hallazgo es consistente con la teoría económica, que sugiere que las regulaciones más estrictas tienden a limitar la capacidad de los fabricantes para explotar los vacíos legales y promocionar sus productos de manera indirecta. A pesar de ello, los resultados también indican que las leyes no han eliminado por completo la capacidad de las marcas mejor posicionadas para mantenerse competitivas en el mercado, lo que sugiere que aún existe espacio para mejorar en la aplicación de estas regulaciones.

Una parte importante del estudio ha sido la evaluación de la utilidad del aprendizaje automático (ML) como herramienta para analizar el mercado del tabaco. El uso de modelos predictivos, como ARIMA y SARIMA, ha permitido predecir el impacto de la COVID-19 en las ventas de tabaco en España. Los modelos estadísticos muestran que el cierre de fronteras y las restricciones de movilidad durante la pandemia provocaron una caída significativa en las ventas de cigarrillos en provincias turísticas y fronterizas, como Alicante y Baleares, mientras que, en regiones como Cádiz y Sevilla, el impacto fue mucho menor.

Este hallazgo subraya la importancia de utilizar herramientas tecnológicas avanzadas como el ML para realizar análisis más precisos y en tiempo real de las fluctuaciones en el mercado del tabaco. Además, destaca cómo el contexto geográfico y socioeconómico influye en el comportamiento de los consumidores, lo que sugiere que las políticas de control del tabaco deben adaptarse a las particularidades de cada región para ser verdaderamente efectivas.

El impacto de la pandemia de COVID-19 en el mercado del tabaco ha sido otro tema central de este análisis. Los resultados muestran que la pandemia ha tenido efectos mixtos en las ventas de tabaco en España. Mientras que algunas provincias experimentaron una disminución significativa en las ventas debido al cierre de fronteras y la falta de turismo, otras apenas se vieron afectadas. Esto pone de manifiesto que el impacto de la pandemia no ha sido uniforme en todo el país, lo que sugiere que las políticas de control del tabaco deben tener en cuenta las diferencias regionales y demográficas.

Además, los resultados también señalan que la pandemia pudo haber ofrecido una oportunidad única para impulsar la disminución del tabaquismo. Sin embargo, algunos estudios sugieren que, en ciertos grupos de la población, el consumo de tabaco aumentó como respuesta al estrés y la ansiedad provocados por la crisis sanitaria. Esto refleja la necesidad de seguir investigando cómo las crisis globales pueden influir en el comportamiento del consumidor y qué intervenciones podrían mitigar los efectos negativos sobre el control del tabaquismo.

El estudio también ha aplicado el marco teórico de la Curva de Kuznets para examinar la relación entre el desarrollo económico y el consumo de tabaco. Los resultados apoyan la hipótesis de que, en las primeras etapas del desarrollo económico, el aumento de los ingresos disponibles lleva a un incremento en el consumo de tabaco. Sin embargo, a medida que las economías se desarrollan y aumentan los niveles de educación y conciencia sobre la salud, el consumo de tabaco tiende a disminuir.

Este patrón ha sido particularmente evidente en España, donde las ventas de cigarrillos han mostrado una tendencia decreciente en las últimas décadas, correlacionada con un aumento en el nivel socioeconómico. No obstante, la investigación también ha revelado que otros productos de tabaco, como los cigarrillos para liar (RYO) y el tabaco de pipa, han actuado como "productos refugio" en tiempos de recesión económica, lo que sugiere que los consumidores cambian a productos alternativos más baratos cuando los ingresos disminuyen.

El análisis también revela que los ciclos económicos tienen un impacto significativo en el consumo de tabaco en España. Durante las expansiones económicas, cuando los ingresos aumentan, las ventas de cigarrillos también tienden a crecer. Sin embargo, durante las recesiones, el desempleo actúa como un factor determinante en la reducción del consumo de cigarrillos (Martín-Álvarez et al., 2020b). Este hallazgo es consistente con estudios previos que sugieren que las recesiones económicas pueden llevar a una disminución en el consumo de tabaco debido a la falta de ingresos disponibles.

A pesar de esto, los datos muestran que, en tiempos de recesión, los consumidores de tabaco tienden a cambiar a productos más baratos como los RYO o el tabaco de pipa, que se perciben como alternativas menos nocivas o más accesibles económicamente. Este fenómeno plantea un desafío adicional para las políticas de control del tabaco, ya que los esfuerzos por reducir el consumo de cigarrillos pueden verse neutralizados por el aumento en el uso de otros productos de tabaco.

Los resultados de este estudio ofrecen una serie de recomendaciones para la formulación de políticas públicas y futuras investigaciones. En primer lugar, es evidente que las políticas de control del tabaco deben ser más adaptables a las diferencias regionales y demográficas. El uso de herramientas tecnológicas como el aprendizaje automático puede ser crucial para monitorear y analizar en tiempo real los patrones de consumo y las respuestas a las políticas antitabaco.

CAPÍTULO 5: CONCLUSIONES Integradoras

En segundo lugar, las políticas deben enfocarse no solo en los cigarrillos tradicionales, sino también en los productos sustitutos como los RYO y el tabaco de pipa. Estos productos, percibidos como menos nocivos, actúan como impulsores del consumo de tabaco en tiempos de recesión, lo que sugiere que las estrategias de control deben ser integrales y abarcar todos los tipos de productos de tabaco.

Finalmente, se recomienda que los responsables de las políticas sigan fortaleciendo las restricciones sobre la publicidad, promoción y patrocinio del tabaco, dado que las regulaciones más estrictas han demostrado ser más eficaces para homogeneizar el mercado y reducir la diferenciación entre marcas (Almeida et al., 2021). Sin embargo, también es importante vigilar de cerca los vacíos legales que permiten a las empresas tabacaleras seguir promocionando indirectamente sus productos.

Referencias

Act 28/2005 (2005) Ley de medidas sanitarias frente al tabaquismo y reguladora de la venta, el suministro, el consumo y la publicidad de los productos del tabaco. BOE No. 209, 42241–42250, <https://www.boe.es/boe/dias/2005/12/27/pdfs/A42241-42250.pdf> [01 November 2022].

Act 42/2010 (2010) Ley por la que se modifica la Ley 28/2005, de 26 de diciembre, de medidas sanitarias frente al tabaquismo y reguladora de la venta, el suministro, el consumo y la publicidad de los productos del tabaco. BOE No. 318, 1–7, <https://www.boe.es/buscar/pdf/2010/BOE-A-2010-20138-consolidado.pdf> [01 November 2022].

Ahluwalia, I. B., Myers, M., & Cohen, J. E. (2020). COVID-19 pandemic: an opportunity for tobacco use cessation. *The Lancet. Public Health*, 5(11), e577. [http://dx.doi.org/10.1016/S2468-2667\(20\)30236-X](http://dx.doi.org/10.1016/S2468-2667(20)30236-X)

Almeida, A., Galiano, A., Golpe, A. A., & Martín-Álvarez, J. M. (2021). The usefulness of marketing strategies in a regulated market: evidence from the Spanish tobacco market. *E+M. Economics and Management*, 24, 171–188. <https://doi.org/10.15240/tul/001/2021-2-011>

Álvarez, R., Vicente, N., Polo, L., Ríos, P., Ferrández, P., AM, F. M. & Galán, G. C. (2021). Tobacco use in Spain during COVID-19 lockdown: an evaluation through social media. *Revista Española de Salud Pública*, 95. PMID: 33724261.

Asare, S., Majmundar, A., Islami, F., Bandi, P., Fedewa, S., Westmaas, L. J., & Nargis, N. (2022). Changes in cigarette sales in the United States during the COVID-19 pandemic. *Annals of Internal Medicine*, 175(1), 141-143. <https://doi.org/10.7326/M21-3350>

Bader, P., Boisclair, D., Ferrence, R.G. (2011). Effects of Tobacco Taxation and Pricing on Smoking Behavior in High-Risk Populations: A Knowledge Synthesis. *International Journal of Environmental Research and Public Health*, 8, 4118 - 4139. <https://doi.org/10.3390/ijerph8114118>

Beam, A. L., & Kohane, I. S. (2018). Big data and machine learning in health care. *Jama*, 319(13), 1317-1318. <https://doi.org/10.1001/jama.2017.18391>

Ben-David, D. (1998). Convergence clubs and subsistence economies. *Journal of Development Economics*, 55(1), 155-171. [https://doi.org/10.1016/S0304-3878\(98\)00037-2](https://doi.org/10.1016/S0304-3878(98)00037-2)

Box George, E. P., Jenkins Gwilym, M., Reinsel Gregory, C., & Ljung Greta, M. (1976). *Time series analysis: forecasting and control*. San Francisco: Holden Bay.

Burki, T. K. (2021). Tobacco industry capitalises on the COVID-19 pandemic. *The Lancet Respiratory Medicine*, 9(10), 1097-1098. [https://doi.org/10.1016/S2213-2600\(21\)00361-1](https://doi.org/10.1016/S2213-2600(21)00361-1)

Cadahia, P., Golpe, A., Martín-Álvarez, J. M., & Asensio, E. (2021). Measuring anomalies in cigarette sales using official data from Spanish provinces: Are the anomalies detected by the Empty Pack Surveys (EPSs) used by Transnational Tobacco Companies (TTCs) the only anomalies?. *Tobacco Induced Diseases*, 19. <https://doi.org/10.18332/tid/143321>

Cadahia, P., Golpe, A. A., Martín-Álvarez, J. M., & Asensio, E. (2022). The Importance of Price, Income, and Affordability in the Demand for Cigarettes in Spain. *Addicta: The Turkish Journal on Addictions*, 9(3). <https://doi.org/10.5152/ADDICTA.2022.22054>

Cantarero-Prieto, D., Lanza-Leon, P., & Pascual-Saez, M. (2023). Does a Kuznets curve exist for tobacco consumption? Evidence from OECD countries. *Applied Economics Letters*, 1-4. <https://doi.org/10.1080/13504851.2023.2211331>

Cantarero-Prieto, D., Pascual-Saez, M., & Gonzalez Diego, M. (2019). Examining an Alcohol consumption Kuznets Curve for developed countries. *Applied Economics Letters*, 26(17), 1463-1466. <https://doi.org/10.1080/13504851.2019.1581901>

Case, A., Lubotsky, D., & Paxson, C. (2002). Economic status and health in childhood: The origins of the gradient. *American Economic Review*, 92(5), 1308-1334. <https://doi.org/10.1257/000282802762024520>

Castello-Climent, A., Domenech, R. (2014). Human Capital Inequality, Life Expectancy and Economic Growth. *The Economic Journal*, 124(581), 170-201. <https://doi.org/10.1111/eoj.12099>

CAPÍTULO 5: REFERENCIAS

- Chaloupka, F.J. (1999) Macro-social influences: The effects of prices and tobacco-control policies on the demand for tobacco products, *Nicotine and Tobacco Research*, 1(1):S105–S109, <https://doi.org/10.1080/14622299050011681>
- Clancy, L., Gallus, S., Leung, J., & Egbe, C. O. (2020). Tobacco and COVID-19: Understanding the science and policy implications. *Tobacco Induced Diseases*, 18. <https://doi.org/10.18332/tid/131035>
- Congregado, E., Feria-Gallardo, J., Golpe, A. A., & Iglesias, J. (2016). The environmental Kuznets curve and CO 2 emissions in the USA: Is the relationship between GDP and CO 2 emissions time varying? Evidence across economic sectors. *Environmental Science and Pollution Research*, 23, 18407-18420. <https://doi.org/10.1007/s11356-016-6982-9>
- Contreary, K. A., Chattopadhyay, S. K., Hopkins, D. P., Chaloupka, F. J., Forster, J. L., Grimshaw, V. & Community Preventive Services Task Force. (2015). Economic impact of tobacco price increases through taxation: a community guide systematic review. *American Journal of Preventive Medicine*, 49(5), 800-808. <https://doi.org/10.1016/j.amepre.2015.04.026>
- Costa-Font, J., Hernandez-Quevedo, C., & Sato, A. (2018). A health ‘Kuznets’ curve? Cross-sectional and longitudinal evidence on concentration indices. *Social Indicators Research*, 136, 439-452. <https://doi.org/10.1007/s11205-017-1558-8>
- Coughlin, L. N., Tegge, A. N., Sheffer, C. E., & Bickel, W. K. (2020). A machine-learning approach to predicting smoking cessation treatment outcomes. *Nicotine and Tobacco Research*, 22(3), 415-422. <https://doi.org/10.1093/ntr/nty259>
- Croissant Y, Millo G (2008). “Panel Data Econometrics in R: The plm Package.” *Journal of Statistical Software*, 27(2), 1–43. <https://doi.org/10.18637/jss.v027.i02>
- Davagdorj, K., Lee, J. S., Park, K. H., & Ryu, K. H. (2019). A machine-learning approach for predicting success in smoking cessation intervention. In 2019 *IEEE 10th International Conference on Awareness Science and Technology (iCAST)* (pp. 1-6). IEEE. <https://doi.org/10.1109/ICAwST.2019.8923252>
- Deaton, A. (2003). Health, inequality, and economic development. *Journal of Economic Literature*, 41(1), 113-158. <https://doi.org/10.1257/jel.41.1.113>
- Del Arco-Osuna, M. Á., Blasco, J., Almeida, A., & Martín-Álvarez, J. M. (2023). Impact of the Spanish smoke-free laws on cigarette sales by brands, 2000–2021: Evidence from a club convergence approach. *Tobacco Induced Diseases*, 21. <https://doi.org/10.18332/tid/174407>
- Driezen, P., Kasza, K. A., Gravely, S., Thompson, M. E., Fong, G. T., Cummings, K. M., & Hyland, A. (2022). Was COVID-19 associated with increased cigarette purchasing, consumption, and smoking at home among US smokers in early 2020? Findings from the US arm of the International Tobacco Control (ITC) Four Country Smoking and Vaping Survey. *Addictive Behaviors*, 129, 107276. <https://doi.org/10.1016/j.addbeh.2022.107276>

- Duflo, Esther. (2012). Women Empowerment and Economic Development. *Journal of Economic Literature*, 50(4):1051-1079. <https://doi.org/10.1257/jel.50.4.1051>
- Dumortier, A., Beckjord, E., Shiffman, S., & Sejdić, E. (2016). Classifying smoking urges via machine learning. *Computer methods and programs in biomedicine*, 137, 203-213. <https://doi.org/10.1016/j.cmpb.2016.09.016>
- Escario, J. J., & Molina, J. A. (2004). Modeling the optimal fiscal policy on tobacco consumption. *Journal of Policy Modeling*, 26(1), 81-93. [https://doi.org/10.1016/S0161-8938\(03\)00008-0](https://doi.org/10.1016/S0161-8938(03)00008-0)
- Esplá, E. P., Faus, C. C., Baldó, A. J., Enrique, I. B., & Vives, E. C. (2021). COVID-19 and smoking: an opportunity to quit. *Archivos de bronconeumologia*, 57(12), 784. <https://doi.org/10.1016/j.arbr.2021.10.009>
- Estévez-Danta, A., Bijlsma, L., Capela, R., Cela, R., Celma, A., Hernández, F., & Quintana, J. B. (2022). Use of illicit drugs, alcohol and tobacco in Spain and Portugal during the COVID-19 crisis in 2020 as measured by wastewater-based epidemiology. *Science of the Total Environment*, 836, 155697. <https://doi.org/10.1016/j.scitotenv.2022.155697>
- Fajnzylber, P., Lederman, D., Loayza, N. (2002). Inequality and Violent Crime. *Journal of Law and Economics*, 45(1), 1-39. <https://doi.org/10.1086/338347>
- Fernández, M. F., Artacho-Cordón, F., Freire, C., Pérez-Lobato, R., Calvente, I., Ramos, R., & Olea, N. (2015). Trends in childrens exposure to second-hand smoke in the INMA-Granada cohort: An evaluation of the Spanish anti-smoking law. *Environmental Research*, 138, 461–468. <https://doi.org/10.1016/j.envres.2015.03.013>
- Fichtenberg, C. M., Glantz, S. A. (2002). Effect of smoke-free workplaces on smoking behaviour: Systematic review. *British Medical Journal*, 325, 188–194. <https://doi.org/10.1136/bmj.325.7357.188>
- Fotourehchi, Z., & Çalışkan, Z. (2018). Is it possible to describe a Kuznets curve for health outcomes? An empirical investigation. *Panoeconomicus*, 65(2), 227-238. <https://doi.org/10.2298/PAN15081411F>
- Fu, R., Kundu, A., Mitsakakis, N., Elton-Marshall, T., Wang, W., Hill, S. & Chaiton, M. O. (2021). Machine learning applications in tobacco research: a scoping review. *Tobacco Control*. <http://dx.doi.org/10.1136/tobaccocontrol-2020-056438>
- Galán, I., Mata, N., Estrada, C., Díez-Gañán, L., Velázquez, L., Zorrilla, B., ... & Ortiz, H. (2007). Impact of the " Tobacco control law" on exposure to environmental tobacco smoke in Spain. *BMC Public Health*, 7(1), 1-7. <https://doi.org/10.1186/1471-2458-7-224>
- Glied, S., & Smith, P. C. (Eds.). (2013). *The Oxford Handbook of Health Economics*. Oxford University Press.
- Golpe, A. A., Martín-Álvarez, J. M., Galiano, A., & Asensio, E. (2022). Effect of IQOS introduction on Philip Morris International cigarette sales in Spain: a Logarithmic Mean

CAPÍTULO 5: REFERENCIAS

Divisa Index decomposition approach. *Gaceta Sanitaria*, 36, 293-300. <https://doi.org/10.1016/j.gaceta.2021.12.007>

Gomajee, R., Torregrossa, H., Bolze, C., Melchior, M., & Lesueur, F. E. K. (2021). Decrease in cross-border tobacco purchases despite intensification of antitobacco policies in France. *Tobacco Control*, 30(4), 428-433. <http://dx.doi.org/10.1136/tobaccocontrol-2019-055540>

Greco, A. M., & Rotthoff, K. W. (2015). Economic growth and obesity: Findings of an Obesity Kuznets curve. *Applied Economics Letters*, 22(7), 539-543. <https://doi.org/10.1080/13504851.2014.955251>

Grossman, G. M., & Krueger, A. B. (1995). Economic growth and the environment. *The Quarterly Journal of Economics*, 110(2), 353-377. <https://doi.org/10.2307/2118443>

Gupta, A., Ghanshala, K., & Joshi, R. C. (2021). Machine Learning Classifier Approach with Gaussian Process, Ensemble boosted Trees, SVM, and Linear Regression for 5G Signal Coverage Mapping. *International Journal of Interactive Multimedia and Artificial Intelligence*, Vol. 6, No 6. <https://doi.org/10.9781/ijimai.2021.03.004>

Hastie, T., Tibshirani, R., Friedman, J. H., & Friedman, J. H. (2009). The elements of statistical learning: data mining, inference, and prediction (Vol. 2, pp. 1-758). New York: Springer. <https://doi.org/10.1007/978-0-387-21606-5>

Hausman, J. A. (1978). Specification tests in econometrics. *Econometrica: Journal of the econometric society*, 1251-1271. <https://doi.org/10.2307/1913827>

Hefler, M., & Gartner, C. E. (2020). The tobacco industry in the time of COVID-19: time to shut it down?. *Tobacco Control*, 29(3), 245-246. <http://dx.doi.org/10.1136/tobaccocontrol-2020-055807>

Hiscock R., Bauld L., Amos A., Fidler J.A., Munafò M. (2012) Socioeconomic status and smoking: a review. *Annals of the New York Academy of Sciences*, 1248(1):107–23. <https://doi.org/10.1111/j.1749-6632.2011.06202.x>

Huda, S., Yearwood, J., & Borland, R. (2010). Cluster based rule discovery model for enhancement of government's tobacco control strategy. In 2010 *Fourth International Conference on Network and System Security* (pp. 383-390). IEEE. <https://doi.org/10.1109/NSS.2010.14>

Hyndman RJ, Khandakar Y (2008). "Automatic time series forecasting: the forecast package for R." *Journal of Statistical Software*, 27(3), 1–22. <https://doi.org/10.18637/jss.v027.i03>

Hyndman R, Athanasopoulos G, Bergmeir C, Caceres G, Chhay L, O'Hara-Wild M, Petropoulos F, Razbash S, Wang E, Yasmien F (2022). forecast: Forecasting functions for time series and linear models. R package version 8.16, <https://pkg.robjhyndman.com/forecast/>

Iglesias, J., Carmona Arango, M., Golpe Moya, A., Martín-Álvarez, J.M. (2013). The Kuznets Curve and CO2 emission in Spain, 1850-2008. *Economía Industrial*, 389, 135-144. <http://hdl.handle.net/10272/10071>

Kasihmuddin, M. S. B. M., Mansor, M. A. B., Abdulhabib Alzaeemi, S., & Sathasivam, S. (2021). Satisfiability logic analysis via radial basis function neural network with artificial bee colony algorithm. *International Journal of Interactive Multimedia and Artificial Intelligence*, Vol. 6, No 6. <https://doi.org/10.9781/ijimai.2020.06.002>

Kasza, K. A., Hyland, A. J., Brown, A., Siahpush, M., Yong, H. H., McNeill, A. D., ... & Cummings, K. M. (2011). The effectiveness of tobacco marketing regulations on reducing smokers' exposure to advertising and promotion: findings from the International Tobacco Control (ITC) Four Country Survey. *International Journal of Environmental Research and Public Health*, 8(2), 321-340. <https://doi.org/10.3390/ijerph8020321>

Kim, J., & Lee, S. (2022). Impact of the COVID-19 pandemic on tobacco sales and national smoking cessation services in Korea. *International Journal of Environmental Research and Public Health*, 19(9), 5000. <https://doi.org/10.3390/ijerph19095000>

Kim, N., McCarthy, D. E., Loh, W. Y., Cook, J. W., Piper, M. E., Schlam, T. R., & Baker, T. B. (2019). Predictors of adherence to nicotine replacement therapy: Machine learning evidence that perceived need predicts medication use. *Drug and alcohol dependence*, 205, 107668. <https://doi.org/10.1016/j.drugalcdep.2019.107668>

Kuznets, S. (1955). Population, income and capital. In *Economic Progress: Proceedings of a Conference held by The International Economic Association at Santa Margherita Ligure, Italy* (pp. 3-20). London: Palgrave Macmillan UK. https://doi.org/10.1007/978-1-349-08440-1_1

Lee, B. P., Dodge, J. L., Leventhal, A., & Terrault, N. A. (2021). Retail alcohol and tobacco sales during COVID-19. *Annals of internal medicine*, 174(7), 1027-1029. <https://doi.org/10.7326/M20-7271>

Mamoshina, P., Kochetov, K., Cortese, F., Kovalchuk, A., Aliper, A., Putin, E. & Zhavoronkov, A. (2019). Blood biochemistry analysis to detect smoking status and quantify accelerated aging in smokers. *Scientific reports*, 9(1), 1-10. <https://doi.org/10.1038/s41598-018-35704-w>

Marmot, M. (1999). Epidemiology of socioeconomic status and health: are determinants within countries the same as between countries?. *Annals of the New York Academy of Sciences*, 896(1), 16-29. <https://doi.org/10.1111/nyas.1999.896.issue-1>

Martín-Álvarez, J. M., Golpe, A. A., Iglesias, J., & Ingelmo, R. (2020a). Price and income elasticities of demand for cigarette consumption: what is the association of price and economic activity with cigarette consumption in Spain from 1957 to 2016?. *Public Health*, 185, 275-282. <https://doi.org/10.1016/j.puhe.2020.05.059>

Martín-Álvarez, J.M., Almeida, A., Galiano, A., & Golpe, A. A. (2020b). Asymmetric behavior of tobacco consumption in Spain across the business cycle: a long-term

CAPÍTULO 5: REFERENCIAS

regional analysis. *International Journal of Health Economics and Management*, 20(4), 391-421. <https://doi.org/10.1007/s10754-020-09286-y>

Martín-Álvarez, J. M., Almeida, A., Golpe, A. A., & Asensio, E. (2023). Electronic device or regulated tobacco product? Learning from the diffusion of heated tobacco products in Spain. *Public Health*, 219, 61-66. <https://doi.org/10.1016/j.puhe.2023.03.017>

Martínez-Cao, C., de La Fuente-Tomas, L., Menéndez-Miranda, I., Velasco, Á., Zurrón-Madera, P., García-Álvarez, L. & Bobes, J. (2021). Factors associated with alcohol and tobacco consumption as a coping strategy to deal with the coronavirus disease (COVID-19) pandemic and lockdown in Spain. *Addictive Behaviors*, 121, 107003. <https://doi.org/10.1016/j.addbeh.2021.107003>

Moodie, C., Hoek, J., Hammond, D., Gallopel-Morvan, K., Sendoya, D., Rosen, L. & Van der Eijk, Y. (2022). Plain tobacco packaging: progress, challenges, learning and opportunities. *Tobacco Control*, 31(2), 263-271. <https://doi.org/10.1136/tobaccocontrol-2020-056138>

Murphy, K. P. (2012). *Machine learning: a probabilistic perspective*. MIT press.

Papanicolas, I., Woskie, L. R., & Jha, A. K. (2018). Health care spending in the united states and other high-income countries. *JAMA*, 319, 1024. <https://doi.org/10.1001/jama.2018.1150>

Patterson, A. C. (2023). Is economic growth good for population health? A critical review. *Canadian Studies in Population*, 50(1), 1. <https://doi.org/10.1007/s42650-023-00072-y>

Pedroni, P. (1999). Critical values for cointegration tests in heterogeneous panels with multiple regressors. *Oxford Bulletin of Economics and statistics*, 61(S1), 653-670. <https://doi.org/10.1111/1468-0084.0610s1653>

Pedroni, P. (2004). Panel cointegration: asymptotic and finite sample properties of pooled time series tests with an application to the PPP hypothesis. *Econometric theory*, 20(3), 597-625. <https://doi.org/10.1017/S0266466604203073>

Pesaran, M. H. (2007). A simple panel unit root test in the presence of cross-section dependence. *Journal of Applied Econometrics*, 22(2), 265-312. <https://doi.org/10.1002/jae.951>

Pfaff, B. (2008). *Analysis of integrated and cointegrated time series with R*. Springer Science & Business Media.

Phillips, P. C., & Sul, D. (2007). Transition modeling and econometric convergence tests. *Econometrica*, 75(6), <https://doi.org/10.1002/jae.1080>

Phillips, P. C., & Sul, D. (2009). Economic transition and growth. *Journal of Applied Econometrics*, 24(7), 1153-1185. <https://doi.org/10.1002/jae.1080>

Pinilla, J., & Abásolo, I. (2017). The effect of policies regulating tobacco consumption on smoking initiation and cessation in Spain: is it equal across socioeconomic groups?. *Tobacco Induced Diseases*, 15(1), 1-10. <https://doi.org/10.1186/s12971-016-0109-4>

Pinilla, J., López-Valcárcel, B. G., & Negrín, M. A. (2019). Impact of the Spanish smoke-free laws on cigarette sales, 2000–2015: Partial bans on smoking in public places failed and only a total tobacco ban worked. *Health Economics, Policy and Law*, 14(4), 536-552. <https://doi.org/10.1017/S1744133118000270>

Pons-Vigués, M., Rando-Matos, Y., Rodríguez-Blanco, T., Ballvé-Moreno, J. L., Ripoll, J., Llobera, J., Morán, J., López-Jiménez, T., Violán-Fors, C., & Bolibar, B. (2019). Effect of the comprehensive smoke-free law on time trends in smoking behaviour in primary healthcare patients in Spain: A longitudinal observational study. *BMJ Open*, 9(3), e020120. <https://doi.org/10.1136/bmjopen-2017-020120>

Preston, S. H. (1975). The changing relation between mortality and level of economic development. *Population Studies*, 29(2), 231-248. <https://doi.org/10.1080/00324728.1975.10410201>

R Core Team (2022). R: A language and environment for statistical computing. *R Foundation for Statistical Computing*, Vienna, Austria. <https://www.R-project.org/>

Reps, J. M., Rijnbeek, P. R., & Ryan, P. B. (2019). Supplementing claims data analysis using self-reported data to develop a probabilistic phenotype model for current smoking status. *Journal of Biomedical Informatics*, 97, 103264. <https://doi.org/10.1016/j.jbi.2019.103264>

Robalino-López, A., Mena-Nieto, Á., García-Ramos, J. E., & Golpe, A. A. (2015). Studying the relationship between economic growth, CO2 emissions, and the environmental Kuznets curve in Venezuela (1980–2025). *Renewable and Sustainable Energy Reviews*, 41, 602-614. <https://doi.org/10.1016/j.rser.2014.08.081>

Rodríguez-Andrés, A., & Amavilah, V. H. (2022). Convergence in suicide rates in FSU countries: a clustering approach. *Applied Economics Letters*, 1-6. <https://doi.org/10.1080/13504851.2022.2041167>

RStudio Team (2020). RStudio: Integrated Development for R. RStudio, PBC, Boston, MA URL <http://www.rstudio.com/>

Ruhm, C. J. (2000). Are recessions good for your health?. *The Quarterly Journal of Economics*, 115(2), 617-650. <https://doi.org/10.1162/003355300554872>

Ruhm, C. J. (2015). Recessions, healthy no more?. *Journal of Health Economics*, 42, 17-28. <https://doi.org/10.1016/j.jhealeco.2015.03.004>

Saloojee, Y., & Mathee, A. (2021). COVID-19 and a temporary ban on tobacco sales in South Africa: impact on smoking cessation. *Tobacco Control*. <http://dx.doi.org/10.1136/tobaccocontrol-2020-056293>

CAPÍTULO 5: REFERENCIAS

Sichera, R., & P. Pizzuto. 2019. ConvergenceClubs: A Package for Performing the Phillips and Sul's Club Convergence Clustering Procedure. *The R Journal* 11 (2): 143–151. <https://doi.org/10.32614/RJ-2019-048>

Singh, A., & Katyan, H. (2019). Classification of nicotine-dependent users in India: a decision-tree approach. *Journal of Public Health*, 27(4), 453-459. <https://doi.org/10.1007/s10389-018-0973-x>

Solow, R. M. (1956). A contribution to the theory of economic growth. *The quarterly Journal of Economics*, 70(1), 65-94. <https://doi.org/10.2307/1884513>

Suelves, J. M., Gomez-Zuniga, B., & Armayones, M. Changes in smoking behaviour due to the COVID-19 pandemic in Spain. *Tobacco Prevention & Cessation*, 7(Supplement). <https://doi.org/10.18332/tpc/143664>

Tirgil, A., Görüş, M. Ş., & Özgür, Ö. (2021). Club Convergence in Cigarette Consumption and Health Policies in Pre-Pandemic Period. *Duzce Medical Journal*, 23(Special Issue), 71-77. <https://doi.org/10.18678/dtfd.859119>

Trapletti, A., & Hornik, K. (2020). tseries: Time Series Analysis and Computational Finance. <https://CRAN.R-project.org/package=tseries>

Vázquez, J. C., & Redolar-Ripoll, D. (2020). COVID-19 outbreak impact in Spain: A role for tobacco smoking?. *Tobacco Induced Diseases*, 18. <https://doi.org/10.18332/tid/120005>

Vázquez, J. C., & Redolar-Ripoll, D. (2020). Epidemiological data from the COVID-19 outbreak in Spain for the promotion of tobacco smoking cessation policies. *Tobacco use insights*, 13, 1179173X20924028. <https://doi.org/10.1177/1179173X20924028>

Wickham H, Averick M, Bryan J, Chang W, McGowan LD, François R, Grolemond G, Hayes A, Henry L, Hester J, Kuhn M, Pedersen TL, Miller E, Bache SM, Müller K, Ooms J, Robinson D, Seidel DP, Spinu V, Takahashi K, Vaughan D, Wilke C, Woo K, Yutani H (2019). “Welcome to the tidyverse.” *Journal of Open Source Software*, 4(43), 1686. <https://doi.org/10.21105/joss.01686>

Wickham H, Bryan J (2023). readxl: Read Excel Files. <https://readxl.tidyverse.org>, <https://github.com/tidyverse/readxl>

World Bank, (1992). World Development Report 1992. New York: Oxford University Press.

World Health Organization (WHO) (2003) WHO Framework Convention on Tobacco Control. <https://apps.who.int/iris/rest/bitstreams/50793/retrieve> [01 November 2022].

World Health Organization (WHO) (2008) Report on the global tobacco epidemic, 2008: the MPOWER package, Geneva: WHO, https://apps.who.int/iris/bitstream/handle/10665/43818/9789241596282_eng.pdf?sequence=1&isAllowed=y [01 November 2022].

Yach, D. (2020). Tobacco use patterns in five countries during the COVID-19 lockdown. *Nicotine & Tobacco Research*. <https://doi.org/10.1093/ntr/ntaa097>

Yaffee, R. A., & McGee, M. (2000). *An introduction to time series analysis and forecasting: with applications of SAS® and SPSS®*. Elsevier.