

Facultad de Ciencias Económicas  
Universidad Nacional del Litoral



Tesis de Maestría en Economía Aplicada

El potencial predictivo de las búsquedas en Internet. Una aplicación  
sobre la demanda de la afluencia turística en destinos de Argentina

Autor: JORGE REYNOLDS, Ramiro Emmanuel

Directora: VICENTÍN MASARO, Jimena

Co-Director: GARCÍA ARANCIBIA, Rodrigo

**Diciembre 2024**

# Índice general

<b>1</b>	<b>Introducción</b>	<b>7</b>
1.1	Problema de investigación . . . . .	7
1.2	Objetivos . . . . .	9
1.2.1	Objetivo general . . . . .	9
1.2.2	Objetivos específicos . . . . .	9
1.3	Hipótesis . . . . .	9
1.4	Estructura de la tesis . . . . .	9
<b>2</b>	<b>Marco teórico y antecedentes</b>	<b>11</b>
2.1	El proceso de toma de decisiones . . . . .	11
2.1.1	Tipos de bienes y consumo turístico . . . . .	12
2.1.2	Tipos de turismo y características . . . . .	12
2.1.3	El proceso decisorio en la selección de consumo de servicios turísticos . . . . .	13
2.2	El sector turístico en Argentina . . . . .	13
2.3	La información imperfecta y sus efectos sobre la toma de decisiones . . . . .	14
2.4	Internet y los motores de búsqueda como mecanismos para sortear los problemas de información . . . . .	15
2.5	Antecedentes . . . . .	17
2.5.1	La utilización de Google Trends para ganar oportunidad en las decisiones . . . . .	17
2.5.2	El uso de información disponible en la web para conocer la demanda turística . . . . .	20
<b>3</b>	<b>Métodos y materiales</b>	<b>22</b>
3.1	Series de tiempo, estacionariedad y raíces unitarias . . . . .	22
3.1.1	Importancia del análisis de estacionariedad y consecuencias de la existencia de raíces unitarias . . . . .	25
3.2	Metodología para realizar predicciones . . . . .	27
3.2.1	Modelo univariado . . . . .	27
3.2.2	Modelo multivariado . . . . .	28
3.3	Variables y fuentes . . . . .	29
3.3.1	Variable de interés predictivo: demanda turística en Argentina . . . . .	29
3.3.2	Variables con potencial explicativo: búsquedas realizadas por usuarios de Google . . . . .	33
3.3.3	Modelos predictivos . . . . .	35
	Modelo univariado . . . . .	35
	Modelo VAR . . . . .	35
3.4	Proceso de implementación del análisis . . . . .	36
<b>4</b>	<b>Resultados: prediciendo la demanda turística en Argentina</b>	<b>38</b>
4.1	Análisis de las condiciones de partida para la modelación . . . . .	38

4.2	Los modelos univariados para la prección de la demanda turística . . . . .	39
4.2.1	Predicciones fuera de la muestra con los modelos univariados . . . . .	41
4.3	Analizando la eficacia de las búsquedas en Google como variables predictoras . . . . .	44
4.3.1	Predicciones fuera de la muestra con modelos multivariados . . . . .	47
4.4	Comparando estrategias de predicción: ¿son las búsquedas de Google útiles para predecir la demanda turística en Argentina? . . . . .	50
<b>5</b>	<b>Discusión, Conclusiones y Recomendaciones</b>	<b>52</b>
	<b>Bibliografía</b>	<b>56</b>

# Índice de figuras

Figura 1	Serie original ilustrativa y componentes de tendencia, estacional, ciclo e irregular . . .	24
Figura 2	Pernoctaciones totales por destino turístico . . . . .	30
Figura 3	Evolución de las búsquedas en Google de cada uno de los términos seleccionados . . .	34
Figura 4	Valores observados y estimados dentro de la muestra (enero 2004 a diciembre 2022) mediante modelos ARIMA . . . . .	39
Figura 5	Valores observados, estimación fuera de la muestra (enero a diciembre 2023) e intervalos al 95 % de confianza obtenidos mediante modelos ARIMA . . . . .	41
Figura 6	Valores observados y estimados dentro de la muestra mediante modelos VAR . . . . .	45
Figura 7	Valores observados, estimación fuera de la muestra e intervalos al 95 % de confianza obtenidos mediante modelos VAR . . . . .	47
Figura A1	Interface de la página de Google Trends . . . . .	59
Figura B1	Residuos de modelo univariado sobre las pernoctaciones en Bariloche . . . . .	61
Figura B2	Residuos de modelo univariado sobre las pernoctaciones en CABA . . . . .	61
Figura B3	Residuos de modelo univariado sobre las pernoctaciones en Ciudad de Jujuy . . . . .	62
Figura B4	Residuos de modelo univariado sobre las pernoctaciones en Ciudad de Mendoza . . . . .	62
Figura B5	Residuos de modelo univariado sobre las pernoctaciones en Pinamar . . . . .	63
Figura B6	Residuos de modelo univariado sobre las pernoctaciones en Ciudad de Santa Fe . . . . .	63
Figura B7	Residuos de modelo univariado sobre las pernoctaciones en Villa Carlos Paz . . . . .	64
Figura B8	Residuos de modelo multivariado sobre las pernoctaciones en Bariloche . . . . .	65
Figura B9	Residuos de modelo multivariado sobre las pernoctaciones en CABA . . . . .	66
Figura B10	Residuos de modelo multivariado sobre las pernoctaciones en Jujuy . . . . .	67
Figura B11	Residuos de modelo multivariado sobre las pernoctaciones en Mendoza . . . . .	68
Figura B12	Residuos de modelo multivariado sobre las pernoctaciones en Pinamar . . . . .	69
Figura B13	Residuos de modelo multivariado sobre las pernoctaciones en Santa Fe . . . . .	70
Figura B14	Residuos de modelo multivariado sobre las pernoctaciones en Villa Carlos Paz . . . . .	71

# Índice de cuadros

Cuadro 1	Error de estimación para cada mes de 2023 y promedio anual de los modelos ARIMA	44
Cuadro 2	Selección de parámetros $p$ y variables endógenas para modelos multivariados . . . . .	45
Cuadro 3	Error de estimación para cada mes de 2023 y promedio anual de los modelos VAR . .	50
Cuadro 4	Comparación de errores de estimación entre modelos ARIMA y modelos VAR . . . . .	51
Cuadro A1	Resultados de pruebas de raíces unitarias de las pernoctaciones totales . . . . .	60
Cuadro A2	Resultados de las pruebas de raíces unitarias ADF sobre las búsquedas . . . . .	60
Cuadro A3	Selección de parámetros $p$ , $d$ y $q$ para modelos univariados . . . . .	60

## Agradecimientos

A mi madre, a Micaela y a Isabel, por su amor incondicional, sostén fundamental de mi vida. Sin su apoyo emocional y motivación, este logro habría sido inalcanzable.

A la memoria de mi padre, cuya sabiduría y valores me han guiado en cada paso de mi vida.

A las personas que han hecho posible la realización de esta tesis. A Jimena, mi directora, por su invaluable orientación, apoyo constante y por compartir conmigo su vasto conocimiento. A Rodrigo, por su gran colaboración en este proceso.

A mis amigos de la infancia y a los que conocí a lo largo de mi formación académica.

A mis compañeros de cursado, en especial a Julieta, con quien compartimos extensas y gratas jornadas de estudio.

A la Bolsa de Comercio de Santa Fe, por financiar la realización de esta carrera de posgrado.

A todos y cada uno de los docentes y no docentes que contribuyeron a mi formación en cada etapa del proceso educativo. Gracias a la Facultad de Ciencias Económicas y a la Universidad Nacional del Litoral. Fundamentalmente, gracias a la universidad pública, pilar fundamental, hoy y siempre, de la movilidad social ascendente en nuestro país.

## Resumen

La incertidumbre económica, característica del contexto actual, hace que resulte necesario disponer de información precisa y oportuna sobre la evolución de las variables económicas, para mejorar el proceso de toma de decisiones. Aun así, la recopilación de datos y la evaluación de diferentes escenarios implican una inversión de tiempo y recursos, lo que convierte a la información en un bien costoso. En este sentido, el avance de la tecnología y particularmente de los motores de búsqueda en Internet, representa una importante ventaja, ya que proporciona un gran caudal de información de alta frecuencia y brinda la posibilidad de conocer patrones a partir de las estadísticas que surgen de la utilización de dichos buscadores.

Es en este sentido que la presente tesis se enfoca en el análisis del sector turístico en Argentina. Particularmente, se propone utilizar los motores de búsqueda como una herramienta para predecir la evolución de la mencionada actividad económica —medida a través de la afluencia de turistas a siete destinos específicos—, aprovechando para ésto los datos disponibles en el motor de búsquedas de Google, la cual se encuentra sistematizada a través de la aplicación complementaria Google Trends.

Así, se pretende conocer si es posible disminuir los tiempos de espera en la publicación de datos oficiales, modelando la evolución de la afluencia de turistas a partir de dos métodos: uno univariado (modelo ARIMA) y otro multivariado (modelo VAR), este último con la incorporación de información proveniente del motor de búsqueda. A partir de la aplicación de los mencionados métodos, se obtienen resultados que muestran mejoras significativas en la predicción de la demanda turística a partir de la inclusión de información proveniente de Google. El modelo multivariado reduce los errores de predicción en cuatro de los siete casos bajo estudio, y en un promedio de 4,6 puntos porcentuales para el total de los destinos analizados.

Sin embargo, es importante destacar que, aun cuando en la mayoría de los casos se logra una reducción de los errores de predicción al incorporar datos provenientes de Google, esto implica necesariamente una complejización de los modelos predictivos por la inclusión de un mayor número de parámetros.

Aun con la desventaja mencionada y considerando los fines prácticos, los resultados obtenidos con los modelos multivariados son superiores a los de los univariados, y permiten a los agentes económicos tomar mejores y más oportunas decisiones respecto a la evolución de la demanda turística, quedando pendientes para futuras investigaciones el desarrollo de mejoras metodológicas en aras de una mayor parsimonia de los modelos predictivos.

Palabras claves: información imperfecta, herramientas de Google, pronóstico económico inmediato, turismo en Argentina

## Abstract

Economic uncertainty, characteristic of the current context, makes it necessary to have precise and timely information about the evolution of economic variables to improve the decision-making process. Nevertheless, the collection of data and the evaluation of different scenarios involve an investment of time and resources, making information a costly commodity. In this sense, advancements in technology, particularly in Internet search engines, represent a significant advantage as they provide a vast amount of high-frequency information and allow for the identification of patterns based on statistics derived from the use of these search engines.

It is in this context that this thesis focuses on the analysis of the tourism sector in Argentina. Specifically, it proposes using search engines as a tool to predict the evolution of this economic activity—measured by tourist influx to seven specific destinations—by leveraging data available from Google’s search engine, which is systematized through the complementary application Google Trends.

Thus, the aim is to determine whether it is possible to reduce the waiting times for the publication of official data by modeling the evolution of tourist influx using two methods: one univariate (ARIMA model) and one multivariate (VAR model), with the latter incorporating information from search engines. The application of these methods yields results that show significant improvements in predicting tourism demand when including data from Google. The multivariate model reduces prediction errors in four out of the seven cases studied, with an average reduction of 4.6 percentage points across all analyzed destinations.

However, it is important to highlight that, although in most cases a reduction in prediction errors is achieved by incorporating data from Google, this necessarily complicates predictive models due to the inclusion of a greater number of parameters.

Despite this mentioned disadvantage and considering practical purposes, the results obtained with multivariate models outperform those from univariate models, allowing economic agents to make better and more timely decisions regarding tourism demand evolution. Future research should focus on developing methodological improvements aimed at achieving greater parsimony in predictive models.

Keywords: imperfect information, Google tools, economic nowcast, tourism in Argentina

# Capítulo 1

## Introducción

### 1.1. Problema de investigación

En el actual contexto económico caracterizado por la incertidumbre y la volatilidad de los mercados, donde los tiempos son cada vez más ajustados y el caudal de datos se incrementa de manera exponencial, es esencial contar con las herramientas adecuadas para transformar dichos datos en información fiable y actualizada para la toma de decisiones tanto en el ámbito público como privado. Como señala Varian (2010), existen muchos mercados —y muy complejos— en el mundo real, por lo que resulta muy costoso —y a veces imposible— obtener información precisa y oportuna de todos a la vez, o incluso, información detallada de uno de ellos en particular. A pesar de dichas dificultades, la recopilación de datos y la evaluación de diferentes escenarios es fundamental para evitar, o hacer menos probable, que los agentes económicos tomen decisiones equivocadas, dado que les permite anticiparse a los distintos escenarios posibles.

Afortunadamente, el progresivo desarrollo de la tecnología ha permitido una disminución de los costos de acceso a la información y ha facilitado su registro, acumulación y sistematización. En este sentido, Internet ha jugado un papel fundamental como herramienta para obtener, almacenar, sistematizar y analizar grandes volúmenes de datos<sup>1</sup>. Vale la pena destacar que, en lo referido a su uso, la pandemia de COVID-19 representó un hito realmente disruptivo. Antes de la misma, los canales de comunicación y el proceso de búsqueda de información a través de Internet ya estaban consolidados en la mayoría de las nuevas generaciones, abarcando diversos ámbitos de la vida cotidiana, como contratación de servicios, educación y comercio, por mencionar algunos. Sin embargo, durante la pandemia, esta tendencia se aceleró aún más, “*internetizando*” prácticamente todos los aspectos de la vida cotidiana y generando una mayor penetración del uso de esta red global en los diferentes estratos etarios.

Dentro de las numerosas herramientas y aplicaciones que ofrece Internet, una que se destaca por su alcance y utilidad son los motores de búsqueda. Estos ofrecen a los usuarios un gran caudal de información sobre tendencias de búsqueda y son, en la gran mayoría de los casos, el “portal de ingreso” de los usuarios de Internet a las páginas web disponibles de donde se derivarán distintas acciones, principalmente de consulta, acceso a material audiovisual, compra de bienes y servicios, entre otras.

Actualmente, existen a nivel mundial diversos motores de búsqueda, pero sin lugar a dudas el más utilizado

---

<sup>1</sup>Resulta redundante aclarar que Internet no solo ha permitido un mayor acceso a la información y ha facilitado su almacenamiento y organización, sino que además ha transformado de manera significativa todos los ámbitos de la vida social y económica, mejorando las comunicaciones, el comercio, la prestación de servicios y la interacción social, entre muchos otros aspectos.

es Google, habiendo alcanzado una participación de mercado de 92,38 % durante 2023<sup>2</sup>. Además de su servicio de búsqueda, esta firma ofrece a sus usuarios una serie de herramientas, dentro de las que se encuentra Google Trends, que permite visualizar el volumen de búsquedas de diferentes palabras y términos realizadas en Google a lo largo del tiempo. Esto posibilita, entre otras cosas, identificar tendencias, estudiar la distribución geográfica de las consultas y realizar pronósticos acerca de las mismas.

Los buscadores en general, y el de Google en particular, son plataformas de gran interés desde el punto de vista económico, dado que permiten realizar un análisis comportamental de los consumidores, quienes a partir de sus búsquedas y compras, generan información sobre sus patrones de preferencias y gustos. Dicho de otra manera, las búsquedas realizadas a través de estos portales permiten captar información de los consumidores y representan un valioso insumo para el diseño de estrategias de marketing.

En el caso puntual del sector turístico, contar con información actualizada es fundamental para tomar decisiones estratégicas efectivas. Dado que este entorno es altamente dinámico y competitivo, los proveedores deben adaptarse rápidamente a las tendencias del mercado y a las preferencias de los consumidores. Sin embargo, en Argentina, los datos oficiales sobre el turismo se publican con un rezago de casi dos meses, lo que dificulta que los agentes del sector actúen de manera oportuna.

Ante esta situación, la información obtenida de búsquedas en Internet, especialmente a través de motores de búsqueda como Google, puede resultar de gran valor. Estos datos ofrecen una visión clara sobre el comportamiento de los viajeros y permiten identificar patrones de interés, evaluar la demanda en tiempo real y anticipar cambios en las preferencias de los turistas, por lo que incorporar esta información podría mejorar la oportunidad de respuesta en el sector y mitigar los efectos del rezago en la publicación de estadísticas oficiales.

El análisis de estas búsquedas no solo ayuda a comprender qué destinos son más populares en un momento dado, sino que también permite a las empresas ajustar sus ofertas y estrategias de ventas para satisfacer mejor las necesidades del consumidor. Al aprovechar la información obtenida de estas plataformas, los actores del sector turístico pueden optimizar su planificación, mejorar la asignación de recursos y diseñar campañas de marketing efectivas. En definitiva, contar con datos precisos y actualizados es fundamental para tomar decisiones informadas que impulsen el crecimiento del turismo.

Por lo mencionado, y a sabiendas de que la información proveniente de Internet, y especialmente de los motores de búsqueda, tiene múltiples usos potenciales, este documento de investigación se focaliza en el análisis del sector turístico en Argentina como caso puntual de aplicación. De manera más específica, se propone utilizar los motores de búsqueda como insumo para realizar predicciones sobre la afluencia de turistas —medida a través de la cantidad de pernотaciones— a una serie de destinos dentro del país y evaluar la fiabilidad de dichas predicciones. Es decir, el propósito de este documento es evaluar la capacidad de predecir comportamientos de los consumidores, a partir de la información recopilada a través de Google Trends, para determinar si la misma contribuye a la realización de proyecciones confiables y oportunas sobre la demanda turística en Argentina, tomando como referencia temporal el periodo 2004-2023.

---

<sup>2</sup>Según el portal alemán Statista GmbH, que realiza estudios de mercado y de opinión, elabora indicadores económicos y releva estadísticas oficiales sobre el uso de Internet y sus diversas herramientas.

## 1.2. Objetivos

### 1.2.1. Objetivo general

Realizar predicciones de la afluencia de turistas a diferentes destinos de la Argentina, que permitan adelantar la disponibilidad respecto a las estadísticas oficiales.

### 1.2.2. Objetivos específicos

- Realizar un análisis descriptivo de la evolución de la afluencia de turistas a diferentes destinos de Argentina durante el período comprendido entre enero de 2004 y diciembre de 2023. Identificar cambios estructurales dentro de dicha evolución.
- Utilizar una estrategia de predicción univariada para las pernoctaciones en cada uno de los destinos turísticos seleccionados, y analizar sus características en cuanto a oportunidad y confiabilidad, medida a partir de los errores de predicción.
- Utilizar una estrategia de predicción multivariada para las pernoctaciones en cada uno de los destinos turísticos seleccionados, y analizar sus respectivas definiciones relacionadas con la oportunidad y confiabilidad.
- Comparar las estimaciones obtenidas mediante ambos modelos y concluir respecto al grado de precisión de los mismos en la predicción de las pernoctaciones en los destinos seleccionados. Identificar las ventajas, desventajas y limitaciones de cada uno de ellos para los propósitos de predicción.

## 1.3. Hipótesis

Partiendo del problema de investigación y los objetivos propuestos y, en base a la revisión bibliográfica realizada, la hipótesis inicial de este documento es que la información proveniente de Google Trends contribuye a la elaboración de predicciones confiables y oportunas sobre la demanda turística de los distintos destinos de Argentina, contribuyendo a reducir la incertidumbre de los agentes en el proceso de toma de decisiones. Por lo mencionado, se espera que el modelo multivariado que incorpora como variables explicativas las búsquedas en Google, permita realizar predicciones con menor error de estimación que las del modelo univariado.

## 1.4. Estructura de la tesis

El presente primer capítulo resulta de carácter introductorio y expone el problema de investigación, los objetivos y la hipótesis de trabajo. El segundo capítulo, destinado al marco teórico y antecedentes, sintetiza algunos aspectos relevantes referidos al proceso de toma de decisiones, los problemas de información imperfecta, y la importancia de Internet y los motores de búsqueda como herramientas para mitigarlos. El mismo comprende además la exposición de trabajos previos y antecedentes que resultan relevantes a los fines del desarrollo de esta investigación. El tercer capítulo se centra en los métodos y materiales, y sintetiza las principales características tanto de los modelos a implementar (univariado y multivariado) como de las series bajo análisis (pernoctaciones y búsquedas de palabras claves realizadas en el motor de búsquedas de Google). El cuarto apartado contiene la exposición de los resultados de los modelos y su comparación, como

así también sus correspondientes pruebas de especificación y estabilidad. En el quinto y último capítulo se expone una breve discusión, conclusiones y recomendaciones de cara a futuras líneas de trabajo.

## Capítulo 2

# Marco teórico y antecedentes

En esta sección se realiza una breve descripción del marco teórico y estado del arte sobre el que se fundamenta el presente documento. Se describen las características del proceso de toma de decisiones, tanto en general como en el caso particular del consumo de bienes y servicios, y específicamente de servicios turísticos. Se aborda además una descripción de la problemática de la información imperfecta, como así también la evolución de la utilización de Internet, y puntualmente del motor de búsquedas de Google, como fuente de datos que podría contribuir a mitigar los problemas derivados de ésta. Adicionalmente, este capítulo contiene una revisión de antecedentes referidos a la utilización de herramientas del mencionado motor de búsquedas y aplicaciones relacionadas, como instrumentos para la realización de estimaciones económicas.

### 2.1. El proceso de toma de decisiones

Dentro del campo de la teoría económica, el proceso de toma de decisiones ocupa un lugar central, dado que constituye uno de los ejes fundamentales a partir de los cuales se desarrollan los modelos y análisis que buscan comprender y explicar el comportamiento de los agentes económicos. En este sentido, la teoría de la decisión, como ámbito específico de estudio del comportamiento de dichos agentes, encuentra sus inicios en los aportes realizados por John von Neumann y Oskar Morgenstern (1944), quienes sentaron las bases de la teoría de juegos y la toma de decisiones bajo incertidumbre. En aportes más recientes, y partiendo de los documentos pioneros de los mencionados autores, Varian (2010) concibe a la toma de decisiones como un proceso racional, a través del cual los individuos o las organizaciones evalúan diferentes alternativas y eligen aquella que les reporta la mayor utilidad o beneficio, dentro de las opciones asequibles.

En el ámbito de la microeconomía, el proceso de toma de decisiones de los consumidores y productores de manera individual es el eje fundamental para entender los agregados y sus variables, como la oferta, la demanda, los precios de equilibrio y la asignación óptima de recursos. De manera más específica, y partiendo del supuesto de racionalidad, se asume que los agentes económicos buscan maximizar su utilidad o beneficios, sujetos a restricciones presupuestarias o tecnológicas, como lo plantean Pindyck y Rubinfeld (2018).

Bajo las mencionadas premisas, la teoría del consumidor postula que los individuos toman decisiones de consumo a partir de comparar las diferentes canastas de bienes y servicios disponibles, eligiendo aquella que les proporciona la mayor satisfacción o utilidad. Este proceso depende no solo de las preferencias individuales, sino también de los precios de los bienes y servicios, que influyen en la accesibilidad de las canastas de consumo. Los consumidores deben considerar su restricción presupuestaria, es decir, su disponibilidad monetaria, lo

que determina, a partir de los precios de los bienes, las combinaciones que efectivamente puede adquirir. De manera análoga, la teoría de la empresa postula como los productores toman decisiones de producción y asignación de recursos con el objetivo de maximizar sus beneficios, sujetos a sus restricciones tecnológicas (Snyder, Stewart & Nicholson, 2015). En este contexto, los precios de los insumos y del producto final son cruciales, ya que afectan la rentabilidad y la viabilidad de las decisiones de producción y consumo.

En resumen, el proceso de toma de decisiones de los agentes económicos, ya sean consumidores, productores, hogares o gobiernos, constituye un elemento central en la teoría económica. La forma en que los agentes toman sus decisiones, en un contexto de información, recursos y restricciones, es fundamental para comprender y predecir el comportamiento de variables económicas clave.

### 2.1.1. Tipos de bienes y consumo turístico

En el análisis del proceso decisorio de compra de bienes y servicios resulta relevante incorporar una mención sobre los tipos de bienes en general, y el consumo turístico en particular. Los bienes se pueden clasificar en varias categorías, incluyendo bienes de consumo, bienes de capital, bienes duraderos y bienes no duraderos (Mankiw, 2012). El consumo turístico se ubica principalmente dentro de la categoría de bienes de consumo, específicamente como un servicio. Los servicios turísticos, como alojamiento, transporte y actividades recreativas, son intangibles y se consumen en el momento de su prestación. Además, el consumo turístico puede ser considerado un bien experiencial, ya que se centra en la creación de experiencias y recuerdos para el consumidor.

### 2.1.2. Tipos de turismo y características

Según la Organización Mundial del Turismo (OMT), organismo especializado de la Organización de las Naciones Unidas (ONU), existen diversos tipos de turismo, de los que se pueden mencionar, a modo de ejemplo, el turismo de placer, el turismo de trabajo, el turismo cultural y el turismo de salud, entre otros. De manera general, cada uno de estos puede definirse de la siguiente manera:

- **Turismo de placer:** se centra en la recreación, el ocio y la búsqueda de experiencias agradables. Los viajeros participan en actividades como excursiones, visitas a atracciones turísticas y actividades al aire libre. La motivación principal es disfrutar y relajarse, lo que influye en la elección de destinos, alojamiento y actividades.
- **Turismo de trabajo:** también conocido como turismo de negocios, comprende viajes realizados con fines laborales, como asistir a conferencias, reuniones o ferias comerciales. Los viajeros de negocios a menudo buscan eficiencia y conveniencia en su alojamiento y transporte, priorizando la proximidad a centros de negocios y la disponibilidad de servicios como salas de reuniones. La motivación principal es cumplir con objetivos profesionales, lo que puede influir en la selección de destinos y la duración de la estancia.
- **Turismo cultural:** se enfoca en el conocimiento y la experiencia de diferentes culturas, historia y patrimonio. Los viajeros participan en actividades como visitas a museos, monumentos, festivales y eventos culturales. La motivación principal es aprender y enriquecer su conocimiento sobre otros modos de vida y tradiciones. Dentro de esta categoría podría incluirse también el turismo educativo, que comprende a quienes por motivos de formación académica viajan de forma frecuente a determinados centros urbanos sin residir de manera definitiva en estos.

- **Turismo de salud:** implica viajar con el propósito de mejorar la salud y el bienestar. Puede incluir tratamientos médicos, terapias alternativas, spas y retiros. Los viajeros buscan destinos que ofrezcan servicios de salud de calidad y un entorno relajante.

Cada una de las modalidades mencionadas presenta características distintivas que influyen en el tipo de información que los viajeros buscan antes de tomar una decisión de consumo. Por ejemplo, quienes planean un viaje de turismo cultural probablemente realicen búsquedas relacionadas con museos, monumentos históricos, festividades tradicionales y actividades educativas. En cambio, los interesados en turismo de salud se enfocarán en encontrar información sobre spas, tratamientos médicos, centros de bienestar y destinos conocidos por sus propiedades curativas. De manera similar, los turistas de placer tenderán a buscar playas, parques temáticos, vida nocturna y actividades recreativas, mientras que los viajeros de negocios se concentrarán en ubicación, conectividad o instalaciones para conferencias.

En resumen, el tipo de turismo define las expectativas y necesidades de información de los viajeros, lo que podría verse reflejado en el comportamiento de búsqueda por parte de los consumidores durante el proceso decisorio, y de manera previa a la realización efectiva del consumo.

### **2.1.3. El proceso decisorio en la selección de consumo de servicios turísticos**

En el caso específico del consumo de servicios turísticos, las decisiones de los consumidores son influenciadas por una variedad de factores, que incluyen, sin carácter exhaustivo, las preferencias personales (que definen el tipo de turismo que se desea consumir), las recomendaciones por parte de otros consumidores, el presupuesto disponible y los precios de los servicios turísticos. Esta última variable juega un papel crucial. Los consumidores deben conocer y considerar los precios de transporte, alojamiento y actividades al momento de planificar un viaje, ya que estos determinan la cantidad y calidad de los productos que pueden adquirir.

En general, los individuos evalúan diferentes opciones basándose en su utilidad percibida, que puede incluir la calidad del servicio, la experiencia esperada y la relación calidad-precio. Sin embargo, no todas las decisiones de consumo de servicios turísticos se basan en una planificación previa. Existe también lo que se conoce como “turismo espontáneo”, que consiste en aquellos casos en que individuos que no planifican con antelación, de repente encuentran la oportunidad de visitar un determinado destino y la aprovechan.

Aun cuando existen casos de turismo espontáneo, la teoría de la decisión sugiere que los consumidores buscan maximizar su satisfacción al comparar las alternativas disponibles, considerando sus limitaciones presupuestarias y los datos obtenidos de diversas fuentes. En este sentido, el uso de herramientas digitales (reseñas en línea, redes sociales y motores de búsqueda) ha facilitado el acceso a información en tiempo real sobre precios, disponibilidad y características de los servicios turísticos, lo que permite a los consumidores tomar decisiones más informadas en el proceso de compra de los diferentes tipos de bienes y servicios disponibles en el mercado.

## **2.2. El sector turístico en Argentina**

En el caso puntual del sector turístico argentino, el mismo se caracteriza por su diversidad geográfica y climática. Argentina es un país con una amplia variedad de destinos que comprenden actividades y experiencias distintivas para quienes los visitan. En este sentido, cuenta con regiones montañosas, que presentan paisajes únicos y donde se pueden practicar deportes de invierno; extensas llanuras, ideales para el turismo rural; regiones de clima cálido y seco, con paisajes característicos; regiones de clima frío, con glaciares y lagos, por

mencionar solo algunos ejemplos. Esta diversidad geográfica se traduce en una amplia oferta que abarca desde el turismo de aventura y actividades al aire libre hasta el turismo cultural y gastronómico.

Adicionalmente, y dado que el clima en Argentina varía significativamente de una región a otra, los principales focos turísticos tienen una marcada estacionalidad. Las playas de la Costa Atlántica son populares en verano, mientras que las estaciones de esquí en la Cordillera de los Andes atraen a turistas en invierno. Esta variabilidad climática permite que el país sea un destino atractivo durante todo el año, adaptándose a las preferencias de diferentes tipos de turistas.

Sin embargo, el turismo de placer no es el único que reviste relevancia, sino que además se pueden destacar ejemplos de turismo de trabajo, cultural y de salud a lo largo de todo el país. De manera más específica, la concentración de gran parte de la actividad económica en centros urbanos como la Ciudad Autónoma de Buenos Aires (CABA), la Ciudad de Rosario o Córdoba, entre otras, genera que un importante número de personas se movilicen hacia dichas ubicaciones tanto por motivos laborales como por la necesidad de llevar adelante su formación profesional. También tienen lugar movimientos de personas hacia centros urbanos por motivos de salud, dado que la mayoría de las instituciones de atención para patologías complejas se encuentran en determinadas localizaciones.

Por lo anterior, se puede afirmar que en Argentina se observa una gran diversidad de turismo, como así también un importante caudal de personas que por diferentes motivos se movilizan fuera de su lugar de residencia y asumen el rol de turistas.

En lo que respecta a la relevancia económica del turismo en Argentina, es importante destacar que, si bien el rubro en el que puede incluirse el sector dentro de las cuentas nacionales (hoteles y restaurantes) representó menos del 2,0% entre el año 2004 y la pandemia de COVID-19, la participación de la mencionada categoría dentro del Valor Agregado Bruto<sup>1</sup> evidenció una tendencia creciente, pasando de 1,68% en el primer año del período a 1,98% en 2019. Luego de caer a un mínimo histórico por el evento mencionado, registró una rápida recuperación a partir de 2021 para retornar a niveles similares a los previos a la pandemia. De acuerdo con los datos oficiales publicados por el Instituto Nacional de Estadística y Censos (INDEC), el rubro de hoteles y restaurantes, contribuyó a la economía del país en unos USD 13.332,2 millones durante el año 2023.

### **2.3. La información imperfecta y sus efectos sobre la toma de decisiones**

Partiendo de la importancia que posee la información en el proceso de toma de decisiones, la teoría de la decisión ha pasado a ocupar un lugar de gran relevancia no solo en la economía, sino también en otras ciencias sociales. Como se mencionó en el apartado 2.1, esta teoría se centra en explicar cómo los individuos y organizaciones eligen entre diferentes opciones bajo condiciones de incertidumbre. Según Becker (1976), los agentes toman decisiones de compra basadas en sus preferencias, ingresos y precios de los bienes o servicios disponibles, buscando maximizar su bienestar al elegir entre diferentes alternativas.

Sin embargo, la toma de decisiones es aún más compleja que simplemente conocer el vector de todos los precios. La dinámica decisoria se enmarca en un sistema cambiante en el que interactúan múltiples factores, alterando las condiciones de mercado. En este contexto, la disponibilidad de información se vuelve crucial,

---

<sup>1</sup>Según el Instituto Nacional de Estadística y Censos (INDEC), el Valor Agregado Bruto se encuentra compuesto por el Producto Bruto Interno (PBI) menos impuestos (a los ingresos brutos, específicos, a los débitos y créditos bancarios, a las exportaciones) netos de subsidios, Impuesto al Valor Agregado (IVA) e impuesto a los productos importados (derechos de importación).

dado que, de hecho, la información de la que disponen los agentes a la hora de tomar decisiones suele ser imperfecta.

La información imperfecta representa un obstáculo crítico para la optimización de las decisiones de los consumidores. Cuando los individuos carecen de datos precisos sobre las alternativas disponibles —incluyendo sus características y precios—, corren mayores riesgos de tomar decisiones subóptimas que no maximizan su utilidad y, como contrapartida, da lugar a que otros agentes pueden aprovecharse de dicha limitación de manera especulativa. Por ejemplo, si un consumidor desconoce el precio real de un bien o servicio, puede terminar pagando más de lo necesario, o consumiendo menores cantidades de las que efectivamente podría, lo que reduce su excedente y, en definitiva, su bienestar; mientras que también existen casos en los que oferentes o vendedores fijan precios superiores a los de equilibrio, aprovechando la imposibilidad de conocimiento de dicha situación por parte de los demandantes. Otro ejemplo opera cuando se ignoran atributos relevantes de los diferentes productos disponibles, por lo que el consumidor podría terminar adquiriendo uno que no se ajusta plenamente a sus preferencias. En este caso, la contraparte que ofrece dicho producto o servicio puede aprovechar la demanda residual que está disponible por efecto de la falla de mercado. En ambos casos ejemplificados, la información imperfecta impide que el consumidor alcance el nivel de utilidad máximo posible dados sus ingresos y gustos y, a su vez, da lugar a que exista un mercado para oferentes o productos que de otra manera no existiría.

Partiendo de las situaciones mencionadas, en el ámbito del estudio de la información imperfecta, es fundamental definir el concepto de información asimétrica. Según Stiglitz (2000), la información asimétrica ocurre cuando una de las partes en una transacción posee más información que la otra. Esto puede dar lugar a dos problemas principales: la selección adversa y el riesgo moral. La selección adversa se presenta cuando productos de baja calidad desplazan a los de alta calidad, lo que no solo reduce la eficiencia del mercado, sino que también disminuye el bienestar social al generar equilibrios no eficientes ante las limitaciones de opciones disponibles para los consumidores. Por otro lado, el riesgo moral, como lo destaca Holmström (1979), ocurre cuando, dado que una de las partes tiene más información sobre sus acciones o intenciones, esta puede actuar de manera tal que sus comportamientos no maximicen el bienestar común.

Teniendo en cuenta que en los mercados reales la información imperfecta es la norma más que la excepción, resulta necesario mejorar la calidad y disponibilidad de información con la que cuentan tanto los consumidores como los oferentes para contribuir a decisiones más informadas y consecuentemente a una asignación más eficiente de recursos. Para reducir la incertidumbre, los agentes pueden recurrir a herramientas como la planificación, la predicción y la búsqueda de información de indicadores que les permitan tomar decisiones más informadas.

## **2.4. Internet y los motores de búsqueda como mecanismos para sortear los problemas de información**

A lo largo de la historia, diversos mecanismos han sido implementados para mitigar los efectos negativos de la información imperfecta. La introducción de regulaciones y estándares de calidad ha sido una estrategia clave para aumentar la transparencia y reducir la asimetría de información en los mercados. Según Stiglitz (2000), las políticas públicas pueden jugar un papel crucial en la mejora del acceso a la información y en la creación de condiciones que favorezcan la competencia y la calidad de los productos ofrecidos. En este contexto, la creación de instituciones que faciliten el intercambio de información, como las agencias de calificación de productos y servicios, ha permitido a los consumidores tomar decisiones en contextos más informados. Akerlof

(1970), en su análisis sobre el mercado de “limones”, destaca cómo la información asimétrica puede llevar a la desconfianza del consumidor y a la reducción de la calidad de los productos en el mercado. Por lo tanto, establecer mecanismos que aseguren la calidad y la transparencia se convierte en una necesidad para mejorar el bienestar social.

Sin embargo, y de manera independiente a la creación de instituciones abocadas a mitigar los mencionados fallos del mercado, la llegada de Internet y los motores de búsqueda ha dado lugar a un salto significativo en la disponibilidad de información. La capacidad de acceder a datos en tiempo real y de manera gratuita ha permitido a los consumidores y productores superar muchas de las barreras impuestas por la información imperfecta. Xiang y Fesenmaier (2017) destacan que el uso de motores de búsqueda ha facilitado la obtención de información sobre precios, disponibilidad, servicios y experiencias previas, permitiendo a los agentes económicos tomar decisiones más informadas y eficientes.

El avance tecnológico de las últimas décadas ha revolucionado la economía de la información, posibilitando el acceso a un universo de datos que antes habría sido impensable. Como destaca Varian (2010), el nivel de conectividad y la reducción de los costos de búsqueda han sido tan significativos que se les considera parte de una “revolución de la información”. Esta revolución ha permitido que los agentes económicos valoren la información como un bien útil, mejorando su situación al reducir los costos de acceso y aumentar la disponibilidad de datos.

A modo de contextualización histórica, se debe destacar que los inicios de la utilización de Internet datan de finales de la década de 1960, mientras que su consolidación a nivel global se dio a partir de la década de 1980, generando un desarrollo tecnológico sin precedentes. En el caso particular de Argentina, la disponibilidad comercial del servicio residencial se materializó por primera vez en el año 1995. Cinco años después, en el 2000, solo el 7,0 % de la población argentina accedía a Internet, según la Unión Internacional de Telecomunicaciones (UIT)<sup>2</sup>. Una década después, en 2010, esta cifra se ubicaba en 45,0 %, mientras que a finales de 2023 el 89,2 % de la población nacional utilizaba Internet en sus computadoras o teléfonos celulares, según datos del Instituto Nacional de Estadística y Censos (INDEC)<sup>3</sup>.

Dentro del mencionado contexto histórico, la expansión de Internet trajo aparejado el desarrollo de diversas herramientas, programas y productos, de los que se destacan los motores de búsqueda cuyo algoritmo de funcionamiento brinda a los usuarios la posibilidad de arribar a sitios por medio de palabras claves. Su principal función es encontrar archivos almacenados y direcciones útiles que posibiliten a los usuarios acceder a los sitios de interés vinculados a su consulta. Actualmente, algunos de los motores de búsqueda disponibles son Aol, Ask, Bing, Dogpile, DuckDuckGo, Ecosia, Flickr, Google, StartPage, Yahoo Search, entre otros; aunque el más importante en los países occidentales es Google, con participaciones en el mercado de 92,07 % durante 2022 y 92,38 % durante 2023, según Statista GmbH<sup>4</sup>.

Además de su función como motor de búsqueda, Google ofrece diversas aplicaciones que complementan su servicio, entre las que se encuentran:

- Google Ads: permite crear anuncios;
- Google AdSense: posibilita la publicación de anuncios creados a través de Google Ads;
- Google Analytics: ofrece información del tráfico de los sitios web según la audiencia, la adquisición, el

---

<sup>2</sup>Según lo informado por el Banco Mundial en base a datos de la Unión Internacional de Telecomunicaciones. Consultar: <https://datos.bancomundial.org/indicador/IT.NET.USER.ZS?locations=AR>.

<sup>3</sup>Según lo informado en el Módulo de Acceso y Uso de Tecnologías de la Información y la Comunicación (MAUTIC), en el marco de la Encuesta Permanente de Hogares (EPH). Consultar: <https://www.indec.gob.ar/indec/web/Nivel3-Tema-4-26>.

<sup>4</sup>Consultar: <https://es.statista.com/estadisticas/664510/desglose-porcentual-de-las-busquedas-online-mundiales-por-buscador>.

comportamiento y las conversiones que se llevan a cabo en los mismos;

- Google Blogger: permite publicar un blog o bitácora en la red;
- Google Maps: permite realizar búsqueda de ubicaciones, geolocalización de un punto concreto, información sobre rutas y lugares de interés cercanos;
- Google Places: contribuye a dar mayor visibilidad a los negocios y brinda información acerca de la ubicación de los mismos a través de Google Maps;
- Google Trends: expone los términos de búsqueda más populares y sus tendencias a lo largo del tiempo, permitiendo la aplicación de diferentes filtros temporales y geográficos.

Cada una de estas herramientas se nutre de la interacción constante con los usuarios, creando un ecosistema que permite a la empresa monetizar la información recopilada. La retroalimentación, que consiste en recabar datos para mejorar productos y servicios, transforma a Google en una red que atrapa información, generando servicios útiles tanto para consumidores como para empresas. Así, Google no solo ofrece herramientas valiosas, sino que también monetiza la información a través de publicidad y análisis de flujos de búsqueda.

Dentro de la enorme cantidad de usos y aplicaciones de Internet y los motores de búsqueda, resulta relevante destacar que, en el caso particular del análisis de la demanda turística, la disponibilidad de información es un factor fundamental. En el proceso de toma de decisiones de consumo, los turistas eligen los destinos a visitar, definen la cantidad de tiempo que esperan permanecer en cada lugar, qué tipo de alojamiento contratar y qué actividades realizar durante su estadía. En este proceso, los agentes hacen uso de todas las posibles fuentes disponibles, incluyendo las búsquedas a través de Internet.

Tanto es así que la mencionada expansión de Internet y el creciente uso de los motores de búsqueda en las últimas décadas consolidan a esta fuente de información como la más relevante en el ámbito de la demanda turística. Actualmente, como lo destacan Xiang y Fesenmaier (2017) y Sigala (2020), la gran mayoría de la información que utilizan los agentes para materializar su demanda de servicios turísticos se obtiene de búsquedas en Internet, donde pueden acceder a datos sobre precios y disponibilidad de alojamientos, servicios incluidos, clima durante los días de estadía, destinos y actividades, medios de transporte recomendados, fotos y videos del alojamiento y zonas cercanas, entre otros.

Habiendo realizado una descripción de la evolución de Internet, los motores de búsqueda y su relevancia como fuente de información en general, y dentro del sector turístico en particular, es importante mencionar que existen numerosos antecedentes sobre la utilización de datos de alta frecuencia provenientes de Internet, así como estimaciones sobre la demanda turística en diferentes países. A continuación, se exponen algunos aportes respecto al uso de Internet en general y Google Trends en particular, como fuentes de datos para la realización de pronósticos sobre diferentes variables económicas, y se mencionan trabajos de investigación en los que la herramienta de Google se utilizó para predecir el comportamiento de variables del sector turístico.

## **2.5. Antecedentes**

### **2.5.1. La utilización de Google Trends para ganar oportunidad en las decisiones**

Los estudios que se presentan a continuación demuestran cómo la información extraída de diversas plataformas de Internet puede ser utilizada para realizar análisis y predicciones económicas, contribuyendo a una mejor comprensión de las dinámicas del mercado y del comportamiento del consumidor.

Un primer documento a destacar es el elaborado por Bollen, Mao y Zeng (2011), que tiene como objetivo analizar cómo el sentimiento expresado en las redes sociales, específicamente en plataformas como Twitter, puede influir en los rendimientos del mercado de valores. Los autores buscan establecer una relación entre las fluctuaciones en el sentimiento social y las variaciones en los precios de las acciones. Para esto, recopilaron datos de publicaciones en Twitter relacionadas con acciones específicas y utilizaron técnicas de análisis de sentimiento para cuantificar las emociones expresadas en esos tweets. Luego, aplicaron modelos econométricos para correlacionar el sentimiento con los rendimientos de las acciones en un marco temporal definido. La investigación encontró que un sentimiento positivo en las redes sociales tiende a preceder aumentos en los precios de las acciones, sugiriendo que las plataformas sociales pueden servir como indicadores anticipados del comportamiento del mercado.

Por otra parte, Boegershausen, Borah y Stephen (2020) analizan el uso de técnicas de web scraping para extraer datos de sitios de comercio electrónico y evaluar el comportamiento del consumidor. Los autores realizan una recopilación de estudios sobre la aplicación de dicha herramienta como medio para la obtención de información que permite analizar los precios, las características de productos y las calificaciones de los consumidores de los bienes y servicios demandados. Los resultados de los casos bajo análisis indican que las reseñas positivas y las estrategias de precios dinámicos son determinantes claves en el comportamiento de compra de los consumidores.

En lo que refiere específicamente al uso de herramientas de la firma Google, existen numerosos antecedentes relacionados con aplicaciones en diferentes ramas de la actividad económica. Por ejemplo, el trabajo de investigación realizado por Choi y Varian (2012) subraya la importancia de contar con información oportuna sobre las condiciones económicas, especialmente en un contexto donde los datos se publican con cierto rezago. Los autores destacan que la información proveniente de Google Trends puede ser un recurso valioso para la predicción de la actividad económica en diversas industrias, como la afluencia de turistas a destinos específicos, las ventas minoristas y de automóviles, y el mercado inmobiliario. Al comparar modelos de predicción, Choi y Varian concluyen que aquellos que incorporan datos de Google Trends como variables explicativas logran estimaciones más precisas que los modelos que no lo hacen. Esto sugiere que la capacidad de anticipar cambios en la actividad económica puede mejorarse significativamente mediante el uso de datos en tiempo real provenientes de Internet, lo que representa una importante ventaja en la toma de decisiones económicas. Adicionalmente, los autores enfatizan que, aunque los modelos autorregresivos simples son efectivos para series con tendencias suaves, pueden fallar ante cambios abruptos de tendencia, lo que refuerza la relevancia de Google Trends como herramienta complementaria en la evaluación económica.

En otro análisis sobre el sector automotriz, Blanco (2014) sostiene que las publicaciones de periodicidad mensual o trimestral continúan siendo las principales fuentes de información macroeconómica y que las mismas se encuentran disponibles con importantes rezagos. Destaca además que fueron surgiendo diversas técnicas que permiten disponer de datos de alta frecuencia, y que Google Trends representa una herramienta que podría complementar los indicadores comúnmente utilizados, permitiendo una evaluación inmediata de las condiciones económicas. Por lo anterior, se propone por objetivos pronosticar las ventas de automóviles y construir un indicador mensual de consumo utilizando datos provenientes de Google Trends. El autor menciona dos considerables desventajas de la inclusión de datos de la fuente mencionada. En primer lugar, la necesidad de establecer un criterio objetivo para seleccionar las palabras claves más adecuadas dentro de aquellas con mayor potencial predictivo y, por otra parte, el hecho de que, al trabajar con un índice estandarizado entre 0 y 100, los resultados no pueden interpretarse de manera tan directa como en el caso de un indicador del volumen total de búsquedas. Aun así, los resultados de su trabajo permiten afirmar que la inclusión de información proveniente de Google Trends podría mejorar la capacidad predictiva de los modelos simples de series de tiempo.

En relación con la importancia de la oportunidad de la información, y lo conveniente de disponer de datos en tiempo real, resulta relevante el aporte de Carrière-Swallow y Labbe (2011), quienes utilizan Google Trends para elaborar un índice de las compras de automóviles en Chile, denominado Google Trends Automotive Index (GTAI). Los autores destacan la necesidad de acortar los tiempos de espera propios de las estadísticas oficiales, que afectan negativamente a la oportuna toma de decisiones del sector privado y del público. Para esto proponen el mencionado indicador, el cual se construye a partir de datos de Internet y permite observar tendencias. Mencionan además la novedad de los datos de Google como fuente de información para la predicción de la demanda agregada en modelos de corto plazo en economías emergentes, ya que se derivan del accionar de micro usuarios y tienen una buena representatividad, se publican con alta frecuencia y en intervalos regulares.

Otro ámbito de aplicación de modelos predictivos con información proveniente de Google es el mercado laboral. En este rubro, se puede destacar el trabajo de Askitas y Zimmermann (2009), autores estos que hacen uso de datos provenientes de Google Trends para modelar y predecir las tasas de desempleo en Alemania. Los resultados del trabajo permiten sostener que existe una fuerte correlación entre las búsquedas de palabras clave y las tasas de desempleo de ese país.

En la misma línea, Naccarato et al. (2018) analizan el mercado laboral en Italia. En este caso, y con el objetivo de predecir la tasa de desempleo juvenil, los autores utilizan información proveniente de Google Trends en modelos univariados y multivariados. Los resultados confirman dos hechos: i) existe una estrecha conexión entre el desempleo juvenil y la búsqueda de empleo en Internet, y ii) el uso de datos de Google Trends modelados a través de vectores autorregresivos (VAR) produce mejores resultados en términos de errores de pronóstico que el autorregresivo integrado de medias móviles (ARIMA), es decir, el modelo multivariado, que incluye información sobre las búsquedas, parece estimar mejor que el univariado que no incorpora dichos datos.

Al igual que Blanco (2014), Naccarato et al. (2018) identifican como aspecto relevante al propio proceso de selección de las palabras claves a utilizar, y destacan que la selección inadecuada puede llevar a errores significativos, como correlaciones espurias, que afectan la validez de los análisis y pronósticos. Para mitigar estos riesgos, sugieren realizar una selección de las palabras clave basada en un conocimiento profundo del fenómeno en estudio.

También se pueden mencionar antecedentes de análisis de demanda de bienes de consumo. En este rubro se destaca la investigación de Bortoli y Combes (2015), quienes hacen uso de información de búsquedas en Google para predecir la evolución del gasto en este tipo de bienes en los hogares de Francia. Los autores concluyen que la utilización de esta herramienta arroja resultados satisfactorios solamente en el caso de los bienes durables. Como elemento favorable, destacan la oportunidad y accesibilidad de los datos provenientes de las búsquedas en Internet.

En cuanto al estudio de agregados macroeconómicos mediante información obtenida a través de Google, destaca entre otras la investigación de Austin et al. (2021), en la cual se examina cómo los datos extraídos de Google Places y Google Trends pueden ser utilizados para desarrollar indicadores de alta frecuencia alineados con los conceptos estadísticos, clasificaciones y definiciones utilizados en la producción de medidas oficiales. Los autores analizan la trayectoria del Producto Bruto Interno (PBI) de Australia, Brasil, Canadá, Filipinas, Francia y Sudáfrica durante la etapa inicial de la pandemia de COVID-19. Para este fin, desarrollan un conjunto de herramientas metodológicas para los compiladores nacionales, interesados en utilizar los datos de Google para mejorar la oportunidad y frecuencia de los indicadores económicos.

Análogamente, Bhattacharjee y Kohns (2022) evalúan los beneficios de la utilización de datos provenientes de Google Trends para pronosticar el crecimiento real del PBI de Estados Unidos a través de modelos

Bayesianos Estructurales de Series de Tiempo (BSTS por sus siglas en inglés). Como resultado de su análisis, los autores encuentran que Google Trends mejora los pronósticos puntuales inmediatos y de densidad en tiempo real dentro de la muestra bajo investigación, dándose las mejores predicciones antes de la publicación de la información oficial. Para probar la efectividad de las estimaciones en contextos de gran volatilidad, aplican el modelo al período de pandemia de COVID-19 y concluyen que la información de Google Trends ayuda a caracterizar la incertidumbre que tuvo lugar durante dicha recesión, como así también durante el período de recuperación posterior.

Otro documento relevante sobre esta temática es el elaborado por Bantis et al. (2022) que utilizan datos de búsqueda de Google Trends para predecir y pronosticar la evolución del PBI para una economía desarrollada (Estados Unidos) y una economía emergente (Brasil). Los autores se centran en evaluar la contribución marginal de estos datos sobre los predictores tradicionales. Como resultado de su investigación, encuentran que los modelos que internalizan información proveniente de Google proporcionan mejores estimaciones en comparación con los modelos que excluyen esta información. Adicionalmente, concluyen que las mejoras en las estimaciones generadas por la utilización de datos de Google Trends resultan muy similares para Brasil y Estados Unidos.

Si bien, a lo largo de esta sección se han mencionado diversos documentos de gran relevancia, en toda revisión bibliográfica sobre la utilización de información proveniente de Google Trends resulta inevitable mencionar el relevamiento realizado por Jun et al. (2018), quienes describen el estado del arte analizando 657 trabajos científicos que, con diferentes objetivos, utilizan Google Trends en las áreas de comunicaciones, medicina, salud, negocios y economía, entre los años 2006 y 2016. Los autores destacan que el foco de atención de estos trabajos pasa de la mera descripción y diagnóstico de tendencias a la realización de pronósticos y estimaciones. Resumen, además, las principales ventajas de Google Trends que se mencionan en los distintos documentos de investigación, dentro de las cuales se incluyen la gran disponibilidad de datos, y la oportunidad y gratuidad de los mismos; en cuanto a las limitaciones, destacan las generalizaciones excesivas basadas en las predicciones con datos de esta fuente que no siempre coinciden con los fenómenos del mundo real.

Con este recuento bibliográfico como marco general, la presente tesis representa la continuación de un trabajo que viene realizando el autor en cuanto a la utilización de diferentes metodologías para mejorar las predicciones económicas utilizando las mencionadas fuentes de datos. En particular, mencionar Jorge (2020) y Camusso y Jorge (2021). El primer documento tuvo como objetivo predecir de manera oportuna el valor del Índice Compuesto Coincidente de Actividad Económica de la Provincia de Santa Fe (ICASF<sub>e</sub>), indicador que se publica con dos meses de rezago, mientras que en el segundo se utilizaron las herramientas Google Trends y Google Correlate<sup>5</sup> para construir un indicador de alta frecuencia de las ventas de supermercados de Santa Fe. En ambos casos, los resultados obtenidos evidenciaron una significativa utilidad de las herramientas de Google en la obtención de estimaciones precisas y oportunas, mejorando los resultados a los que se arribaron mediante modelos que no internalizan información proveniente de las búsquedas.

### **2.5.2. El uso de información disponible en la web para conocer la demanda turística**

Además del mencionado trabajo de Choi y Varian (2012), que se centra en el estudio tanto de la afluencia de turistas a diferentes destinos como en otras variables económicas (ventas minoristas, de automóviles y de

---

<sup>5</sup>Aunque no se encuentra disponible desde finales del año 2019, Google Correlate permitía, dado un patrón de interés temporal o espacial, determinar qué consultas imitaban mejor dichos datos. Al ingresar una serie temporal, Google Correlate brindaba un listado con las 100 palabras claves cuyas búsquedas presentaban un comportamiento similar a la serie ingresada. Esto posibilitaba contar con una herramienta objetiva para la selección de los términos más pertinentes a analizar posteriormente mediante Google Trends.

inmuebles), existen numerosos antecedentes respecto a la predicción de la demanda del sector turístico a nivel mundial. Un ejemplo de esto es el estudio de Artola y Galan (2012), en el cual se utilizan herramientas de Google para analizar los flujos de turistas británicos hacia destinos dentro de España. Los autores destacan las ventajas de la información de Internet, debido a su volumen, disponibilidad en tiempo real y gratuidad.

De manera análoga, Park et al. (2017) proponen realizar pronósticos de la afluencia de turistas japoneses a Corea del Sur mediante la utilización de datos de Google Trends, contrastando modelos univariados y multivariados. Dicho trabajo tiene la particularidad de que incorpora una variable *dummy* que permite modelar la ocurrencia de eventos que potencialmente podrían afectar el número de turistas que ingresan al país, como pueden ser eventos deportivos, políticas de incentivos en ese sector, ocurrencia de crisis económicas que afecten negativamente el caudal de visitantes, entre otros. Tanto en el caso de Artola y Galán como en el de Park et al. se concluye que los modelos que incorporan información proveniente de Internet tienen un mejor rendimiento que aquellos que solo internalizan datos sobre la variable dependiente, como es el caso de los modelos autorregresivos.

Un antecedente más reciente es el trabajo de Dergiades et al. (2018), quienes utilizan información de búsquedas en Internet para predecir el consumo de productos turísticos en Chipre entre los años 2004 y 2015, considerando que los motores de búsquedas constituyen una herramienta líder en la planificación de vacaciones. Los autores destacan que el uso de esta información permite desarrollar estrategias de comercialización a mediano plazo, facilita la formulación de políticas de precios, da lugar a una programación adecuada de las inversiones, y posibilita la efectiva asignación de recursos limitados.

Finalmente, el estudio de Correa (2021) examina si los criterios de búsqueda de Google Trends permiten predecir el arribo de turistas a Colombia, comparando un modelo base que utiliza como predictor solamente los rezagos propios de la llegada de turistas (modelo univariado) con dos especificaciones alternativas, a saber: un modelo que incluye como predictor los datos mensuales de Google Trends y un modelo que incluye datos semanales de esa misma fuente<sup>6</sup>. Los resultados presentan evidencia estadísticamente significativa de que la información proveniente de Google Trends aporta beneficios a la evaluación y predicción de la llegada de turistas y que el modelo que incluye datos de alta frecuencia (semanales) agrega un mayor valor predictivo que el modelo que incluye datos mensuales sobre las búsquedas.

---

<sup>6</sup>Para compatibilizar la diferencia en la unidad temporal de medida de las variables explicativas (búsquedas) y la llegada de turistas, el autor utiliza el método *Mixed Data Sampling* (MIDAS), que permite usar datos de alta frecuencia para explicar la evolución de variables de baja frecuencia.

# Capítulo 3

## Métodos y materiales

En el presente capítulo se exponen los fundamentos teóricos y metodológicos que permiten el análisis de series de tiempo, tanto en general como en el caso específico de la predicción de pernoctaciones en diferentes destinos turísticos de Argentina. En primer lugar, se define qué es una serie de tiempo, destacando la relevancia del análisis de su estacionariedad y las consecuencias de la existencia de raíces unitarias en la modelación.

Posteriormente, se presenta una síntesis sobre los tipos de modelos econométricos aplicables, diferenciando entre modelos univariados y multivariados. Como caso representativo del primer grupo, se utiliza un modelo ARIMA (*Autoregressive Integrated Moving Average*), mientras que para el análisis multivariado, se implementa un modelo VAR (*Vector Autoregression*), que permite explorar las interrelaciones entre las pernoctaciones y las búsquedas realizadas por los usuarios en Google y publicadas por esta empresa a través de Google Trends. Es importante destacar que la variable de interés predictivo en ambos casos está representada por las pernoctaciones, siendo esta la propia variable explicativa en el modelo univariado y adicionándose las búsquedas para el caso multivariado.

Finalmente, y a través de un enfoque metodológico específico, se detalla el procedimiento a utilizar para la aplicación de ambos modelos, exponiendo la representación analítica de cada uno para la predicción de la demanda turística en los destinos seleccionados.

### 3.1. Series de tiempo, estacionariedad y raíces unitarias

Tal como lo plantean Wooldridge (2009), y Gujarati y Porter (2010), una serie de tiempo es un conjunto de observaciones secuenciales de una variable, indexadas cronológicamente, que se recopilan en intervalos regulares (con periodicidad diaria, semanal, mensual, trimestral o anual). De manera simbólica, se puede expresar de la siguiente manera:

$$y_t = t_t + s_t + c_t + i_t, \quad (3.1)$$

donde  $y_t$  representa el valor de la variable  $y$  en el momento  $t$ ,  $t_t$  representa al componente de tendencia,  $s_t$  al estacional,  $c_t$  al cíclico e  $i_t$  al irregular.

En este tipo de series, el orden temporal de las observaciones proporciona información valiosa, ya que permite comprender cómo los eventos pasados pueden influir en los futuros a través de patrones, tendencias y

comportamientos de los datos.

Para el caso puntual del análisis económico, algunas de las series de tiempo más comúnmente estudiadas son las correspondientes al precio de los valores negociables (diaria), los activos y pasivos monetarios de los bancos centrales (semanal), la cantidad de personas empleadas, el nivel de actividad económica, el índice de precios al consumidor (mensuales), el producto bruto, la pobreza (trimestrales) y la ejecución presupuestaria de las partidas del sector público (anual), entre otras.

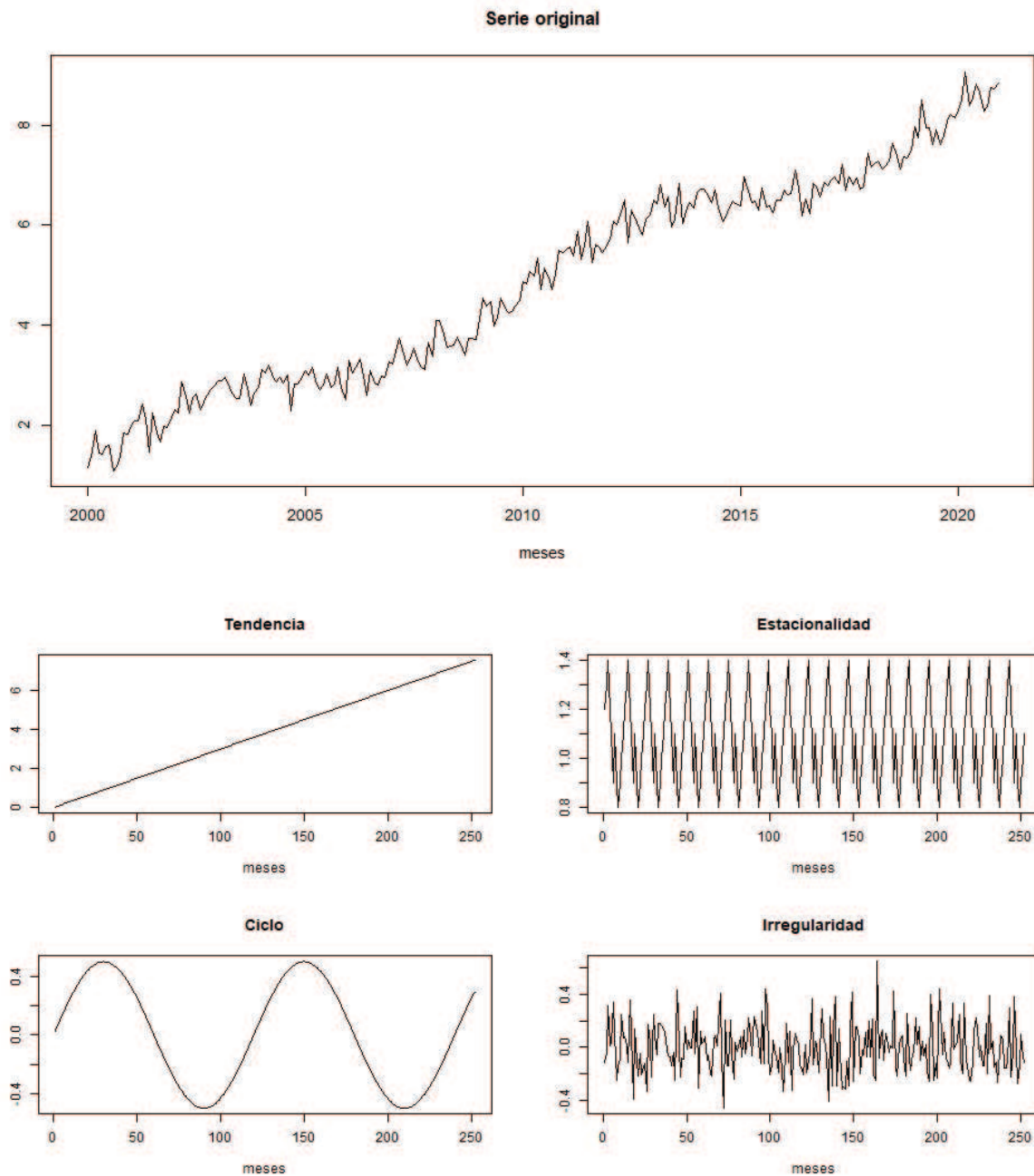
Como se expresa en la ecuación 3.1 las series de tiempo pueden componerse por cuatro elementos principales. Tal y como destacan Chatfield y Xing (2019), entre otros numerosos autores, los mencionados componentes tienen la siguiente interpretación:

- **Tendencia** ( $t_t$ ): representa el movimiento a largo plazo de la serie, e indica si la variable bajo análisis está aumentando, disminuyendo o permaneciendo constante a lo largo del tiempo. Una dificultad inherente a esta definición es decidir qué se entiende por “largo plazo”, por lo que, al hablar de tendencia, se debe tener en cuenta el número de observaciones disponibles y hacer una evaluación subjetiva de dicho concepto.
- **Estacionalidad** ( $s_t$ ): refiere a patrones recurrentes y relativamente predecibles que ocurren en intervalos regulares y dentro de un período de un año. Existen numerosas variables económicas que poseen un patrón estacional, dentro de las que se pueden destacar las ventas de determinados productos (mayores ventas de helados en verano que en invierno, mayores ventas de abrigos en los meses invernales) y servicios (mayor consumo de gas durante el invierno, mayor cantidad de turistas visitando playas durante los meses de verano).
- **Ciclos** ( $c_t$ ): son fluctuaciones en la serie que ocurren a lo largo de intervalos más largos que los correspondientes a la estacionalidad, comprendiendo un período mayor al año (desde 3 o 4 años hasta más de 10 años, dependiendo de la variable medida). Para el caso del análisis económico, las oscilaciones cíclicas a menudo están relacionadas con factores económicos más amplios que la mera estacionalidad, tales como las fases económicas de expansión y recesión.
- **Componente irregular** ( $i_t$ ): es la variabilidad aleatoria en la serie que no puede explicarse por los componentes anteriores, a menudo es considerada como un error de medición o influencias externas no controladas y poco frecuentes. Un ejemplo representativo de este componente podría ser el efecto que tuvo en sus primeros meses la pandemia de COVID-19 sobre el nivel general de actividad económica.

En la Figura 1, y a modo de ejemplo, se expone la evolución de una serie meramente ilustrativa y la correspondiente desagregación de cada uno de los elementos que la componen.

Considerando la relevancia de las series de tiempo y entendiendo sus componentes, queda claro que el análisis de éstas es crucial en economía porque permite comprender la dinámica de las variables y realizar predicciones sobre su comportamiento futuro. Por ejemplo, en el caso del crecimiento económico, las series de tiempo del PBI permiten identificar si la economía se encuentra en una fase de expansión o recesión, y proyectar su evolución a corto y mediano plazo. De manera similar, el estudio de las series de inflación es fundamental para la formulación de políticas monetarias y la toma de decisiones de inversión.

Figura 1: Serie original ilustrativa y componentes de tendencia, estacional, ciclo e irregular



Fuente: elaboración propia.

En el rubro del turismo, el análisis temporal de variables como la cantidad de pernoctaciones, el gasto turístico o la ocupación hotelera es esencial para la planificación y gestión del sector. Las mencionadas series permiten detectar tendencias, estacionalidades y ciclos en la demanda turística, facilitando la toma de decisiones estratégicas por parte de los agentes públicos y privados. Además, la modelización de dichas series puede utilizarse para predecir la evolución futura de la actividad sectorial, lo que contribuye a la asignación eficiente de recursos y la adaptación a los cambios en las preferencias de los turistas.

En resumen, el análisis de series de tiempo es una herramienta poderosa para comprender y predecir el comportamiento de variables económicas. Sin embargo, para que la aplicación de modelos predictivos y la realización de pronósticos tenga validez, las series de tiempo bajo análisis deben cumplir con ciertos requisitos, dentro de los que se destaca la estacionariedad o no existencia de raíces unitarias.

### 3.1.1. Importancia del análisis de estacionariedad y consecuencias de la existencia de raíces unitarias

Antes de continuar con la evaluación de los diferentes modelos predictivos, es importante comprender el concepto de estacionariedad de las series de tiempo. De manera general, se puede afirmar que un proceso estocástico o una serie de tiempo es estacionario si se cumplen tres condiciones: i) su media es constante en el tiempo, ii) su varianza es constante en el tiempo, y iii) el valor de la covarianza entre dos periodos depende únicamente de la distancia entre estos, independientemente del momento en el que se calcule dicha covarianza (Gujarati & Porter, 2010).

En contraposición a las series estacionarias, se encuentran las no estacionarias, cuyo ejemplo clásico son los modelos de caminata aleatoria (MCA). Estos poseen características tales que, la mejor predicción para el valor futuro de una variable aleatoria temporal se constituye como el valor actual de la misma más un “choque o shock” puramente aleatorio. De modo tal que:

$$y_t = \rho y_{t-1} + u_t, \quad -1 \leq \rho \leq 1, \quad (3.2)$$

donde el coeficiente de autocorrelación  $\rho$  mide el efecto de  $y_{t-1}$  sobre  $y_t$ . En el caso en que  $\rho$  es igual a 1, tenemos lo que se conoce como problema de raíz unitaria; lo que necesariamente implica que enfrentamos una situación de no estacionariedad.

El cumplimiento de la condición de estacionariedad es importante porque en los casos en los que el valor futuro de  $y$  ( $y_{t+1}$ ) depende exclusivamente del valor actual ( $y_t$ ) más un choque aleatorio ( $u_t$ ), cualquier pronóstico que intente realizarse sobre la evolución de la serie temporal representaría un ejercicio inútil. Dicho de otra manera, si una serie de tiempo es no estacionaria, solo podemos estudiar su comportamiento durante el periodo bajo análisis, no resultando posible generalizar las conclusiones para periodos futuros. Esto implica que, para propósitos de pronóstico, los modelos predictivos aplicados a las series de tiempo no estacionarias tienen poco valor práctico (Gujarati & Porter, 2010).

Para corroborar la existencia de raíces unitarias y la consecuente no estacionariedad de una serie de tiempo, existen diferentes pruebas. Una alternativa consiste en evaluar la significancia estadística de los coeficientes de autocorrelación, denotados por  $\rho_k$ , donde el subíndice  $k$  representa la cantidad de rezagos. En esta línea, Bartlett (1946) demuestra que una serie de tiempo es puramente aleatoria cuando la distribución muestral de  $\rho_k$  es normal, con media 0 y varianza  $1/n$ , es decir  $\hat{\rho}_k \sim N(0, 1/n)$ . A partir de esto, se construye un intervalo de confianza para  $\rho_k$ . Si el valor cero está incluido dentro del intervalo, no se rechaza la hipótesis de que el verdadero coeficiente de correlación ( $\rho_k$ ) es cero, entonces existe evidencia a favor de la estacionariedad de la serie, al menos respecto al rezago  $k$ .

Para los casos en los que se pretende probar la hipótesis conjunta de que todos los  $\rho_k$  son simultáneamente cero, se puede utilizar el método desarrollado por Box y Pierce (1970), denominado estadístico  $Q$ , o una de sus variantes, el estadístico Ljung-Box (LB). Ambos tienen una distribución chi-cuadrado ( $\chi^2$ ) para muestras grandes. A mayores valores del estadístico  $Q$  (o del estadístico LB), mayor es la evidencia de no estacionariedad de la serie.

Finalmente, otra de las pruebas paramétricas de mayor renombre es la desarrollada por Dickey y Fuller (1979), quienes utilizando simulaciones de Monte Carlo, y estudiando la distribución de las mismas —y de los estadísticos— calculan los valores críticos de la prueba. Dickey-Fuller (DF) (citado en Gujarati y Porter, 2010) indican que una serie de tiempo  $y$  puede seguir diferentes modelos:

1- La serie  $y_t$  es una caminata aleatoria, sin constante ni tendencia, entonces puede expresarse de la siguiente manera:

$$\Delta y_t = \delta y_{t-1} + u_t \quad (3.3)$$

2- La serie  $y_t$  es una caminata aleatoria con constante, entonces:

$$\Delta y_t = \beta_1 + \delta y_{t-1} + u_t \quad (3.4)$$

3- La serie  $y_t$  es una caminata aleatoria con constante y tendencia determinista, la cual puede escribirse como:

$$\Delta y_t = \beta_1 + \beta_2 t + \delta y_{t-1} + u_t \quad (3.5)$$

Por lo tanto, la prueba consiste en analizar la significancia del parámetro  $\delta$ .

Una versión más compleja de este test es la prueba Dickey-Fuller Aumentada (ADF) que adiciona a las tres ecuaciones anteriores una serie de valores rezagados de la variable dependiente  $\Delta y_t$ . La estructura general de la prueba ADF es la siguiente:

$$\Delta y_t = \beta_1 + \beta_2 t + \delta y_{t-1} + \sum_{i=1}^p \alpha_i \Delta y_{t-i} + \varepsilon_t, \quad (3.6)$$

donde  $\varepsilon_t$  es error puro de ruido blanco en el momento  $t$ , y la prueba ADF sigue consistiendo en evaluar la significancia del parámetro  $\delta$ .

Además de la falta de capacidad predictiva de los diferentes modelos en los casos de series no estacionarias, se pueden mencionar al menos dos inconvenientes adicionales. El primero es que cualquier predicción que se realice partiendo de series de tiempo no estacionarias, da lugar a errores de estimación no estacionarios y que, por lo tanto, tienen problemas de autocorrelación. Si bien, la autocorrelación de los errores puede ocurrir aún cuando las series utilizadas en la estimación son estacionarias, la no estacionariedad de la serie es condición necesaria para la no estacionariedad del error. Otro inconveniente relacionado con la existencia de raíces unitarias es la aparición de regresiones espurias.

Una alternativa válida y muy utilizada para salvar los problemas de no estacionariedad es la modelación de las primeras diferencias de la serie bajo análisis. Sin embargo, la utilidad de esta solución está sujeta al orden de la no estacionariedad y a los objetivos propios de las estimaciones y aplicaciones de cada caso.

Habiendo destacado la relevancia del análisis de raíces unitarias, se procede a desarrollar las características generales de los modelos univariados y multivariados.

## 3.2. Metodología para realizar predicciones

### 3.2.1. Modelo univariado

Como explican Gujarati y Porter (2010), los modelos univariados se utilizan para realizar predicciones de una variable utilizando información pasada de la misma, i.e. se asume la existencia de una relación entre su valor pasado y presente. En series de tiempo, los modelos univariados más comunes para realizar predicciones son los denominados ARIMA. Esta es una sigla en la que AR denota que la serie a explicar sigue un proceso autorregresivo, es decir, el valor de la variable  $y$  en el momento  $t$  es simplemente una proporción de su valor en el momento  $t - p$ , más una perturbación aleatoria en el momento  $t$ . Así, por ejemplo, un modelo autorregresivo de orden 1, denotado como AR(1), puede ser expresado de la siguiente manera:

$$y_t = \alpha_1 y_{t-1} + u_t, \quad (3.7)$$

siendo  $y_t$  el valor de  $y$  en el momento  $t$ ,  $\alpha_1$  la proporción en la que  $y_t$  se compone de  $y_{t-1}$ , y  $u_t$  la perturbación aleatoria de la variable  $y$  en el momento  $t$ . Generalizando, un modelo AR( $p$ ) puede ser expresado como:

$$y_t = \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \dots + \alpha_p y_{t-p} + u_t \quad (3.8)$$

Por otro lado, una serie en particular también puede estar compuesta por un proceso de medias móviles MA de orden  $q$ , que se denota como MA( $q$ ), y puede definirse como:

$$y_t = \mu + \beta_0 u_t + \beta_1 u_{t-1} + \beta_2 u_{t-2} + \dots + \beta_q u_{t-q} \quad (3.9)$$

donde  $\mu$  es una constante, los  $\beta_i$  son las proporciones constantes en las que  $u_{t-i}$  afectan a  $y_t$ , y  $u$ , al igual que antes, es el término de error estocástico, por lo que el valor de  $y$  en el periodo  $t$  es igual a una constante más un promedio móvil de los términos de error presente y pasados.

Estos procesos se pueden dar de manera conjunta en una serie de tiempo. Ello es lo que se denomina ARMA( $p,q$ ), donde  $p$  es el orden del proceso AR y  $q$  el del MA. El proceso ARMA( $p,q$ ) puede ser expresado como sigue:

$$y_t = \theta + \alpha_1 y_{t-1} + \alpha_2 y_{t-2} + \dots + \alpha_p y_{t-p} + \beta_0 u_t + \beta_1 u_{t-1} + \beta_2 u_{t-2} + \dots + \beta_q u_{t-q} \quad (3.10)$$

donde  $\theta$  es un valor constante y cada uno de los demás componentes representa los argumentos descritos precedentemente.

Finalmente, es importante destacar que los procesos mencionados pueden dar lugar a series que no son estacionarias (la media o la varianza no son constantes, o la covarianza no es invariante en el tiempo), de hecho, esto es lo más común para series de tiempo económicas.

En estos casos, se debe establecer el orden de integración  $d$ , de manera tal que al diferenciar la serie original dicha cantidad de veces se logra la estacionariedad de la misma. El mencionado orden de integración representa al componente I de los modelos ARIMA.

Si una serie de tiempo es integrada de orden 1 ( $d = 1$  o lo que es igual, I(1)), sus primeras diferencias son I(0) (estacionarias), mientras que si una serie de tiempo es I(2), sus segundas diferencias son estacionarias.

Esto puede generalizarse, por lo que si una serie de tiempo es  $I(d)$ , después de diferenciarla  $d$  veces se obtiene una serie estacionaria.

De lo expuesto precedentemente se deduce que si debemos diferenciar una serie de tiempo  $d$  veces para hacerla estacionaria y luego aplicarle el modelo  $ARMA(p,q)$ , entonces la serie original se presenta a través de un  $ARIMA(p,d,q)$ . Esto significa que la serie de tiempo es autorregresiva integrada de promedios móviles, donde  $p$  denota el número de términos autorregresivos,  $d$  el número de veces que la serie debe diferenciarse para hacerse estacionaria y  $q$  el número de términos de promedios móviles. Formalmente, dicho proceso estocástico —o serie de tiempo— puede ser escrito de la siguiente manera:

$$y_t = -(\Delta^d y_t - y_t) + \theta_0 + \sum_{i=1}^p \alpha_i \Delta^d y_{t-i} + \sum_{j=1}^q \beta_j u_{t-j} + \varepsilon_t \quad (3.11)$$

donde  $y_t$  representa el valor de la variable de interés predictivo en el momento  $t$ ,  $\theta_0$  es una constante,  $\alpha_i$  mide la incidencia de los valores de  $y$  en  $t - i$  sobre su valor actual, y  $\beta_j$  mide el efecto de los errores estocásticos ( $u$ ) en el momento  $t - j$  sobre el valor actual de  $y$ .

### 3.2.2. Modelo multivariado

Así como el modelo  $ARIMA$  utiliza información del pasado de una variable para la realización de predicciones, el modelo de Vectores Autorregresivos ( $VAR$ ) es una herramienta estadística que permite analizar la relación dinámica entre múltiples variables endógenas. La multiplicidad de variables no es lo único que caracteriza a los modelos  $VAR$ , sino que además, en estos se asume que dichas variables son explicadas de manera endógena (i.e. dentro del modelo) y que no presentan relaciones de causalidad unilateral. Esto implica que los valores contemporáneos de éstas no se incluyen como factores explicativos en ninguna de las ecuaciones del modelo, sino que se establece un sistema de ecuaciones simultáneas.

A modo de ejemplo, partiendo del caso de un modelo  $VAR$  de orden 1 con dos variables endógenas, la representación formal del mismo viene dada por:

$$\begin{aligned} y_{1t} &= \alpha_{10} + \alpha_{11}y_{2t} + \alpha_{12}y_{1t-1} + \alpha_{13}y_{2t-1} + \varepsilon_{1t} \\ y_{2t} &= \alpha_{20} + \alpha_{21}y_{1t} + \alpha_{22}y_{1t-1} + \alpha_{23}y_{2t-1} + \varepsilon_{2t} \end{aligned} \quad (3.12)$$

Dado que los términos  $y_{1t}$  y  $y_{2t}$  se encuentran en ambas ecuaciones del sistema 3.12, mediante operaciones algebraicas simples se puede obtener la siguiente forma reducida del modelo  $VAR(1)$ :

$$\begin{aligned} y_{1t} &= \beta_{10} + \beta_{11}y_{1t-1} + \beta_{12}y_{2t-1} + u_{1t} \\ y_{2t} &= \beta_{20} + \beta_{21}y_{1t-1} + \beta_{22}y_{2t-1} + u_{2t} \end{aligned} \quad (3.13)$$

En forma matricial, el sistema de ecuaciones 3.13 puede representarse de la siguiente manera:

$$\begin{pmatrix} y_{1t} \\ y_{2t} \\ \vdots \\ y_{kt} \end{pmatrix} = \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_k \end{pmatrix} + \begin{pmatrix} a_{11} & a_{12} & \cdots & a_{1k} \\ a_{21} & a_{22} & \cdots & a_{2k} \\ \vdots & \vdots & \ddots & \vdots \\ a_{k1} & a_{k2} & \cdots & a_{kk} \end{pmatrix} \begin{pmatrix} y_{1t-1} \\ y_{2t-1} \\ \vdots \\ y_{kt-1} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \\ \vdots \\ \varepsilon_{kt} \end{pmatrix}$$

Siguiendo la lógica anterior, la formulación básica de un modelo VAR de orden  $p$  implica una colección de  $p$  ecuaciones simultáneas, en la cual cada una corresponde a una variable de interés. Si consideramos  $k$  variables endógenas denotadas como  $y_1, y_2, \dots, y_k$ , el modelo VAR( $p$ ) se puede exponer de la siguiente manera:

$$y_t = c + \sum_{i=1}^p A_i y_{t-i} + \varepsilon_t \quad (3.14)$$

donde  $y_t$  es un vector de dimensiones  $k \times 1$  que contiene las variables endógenas en el momento  $t$ ,  $c$  es un vector constante de dimensiones  $k \times 1$ ,  $A_i$  son matrices de coeficientes de dimensiones  $k \times k$  que cuantifican la fuerza y dirección de las relaciones de autorregresión entre las variables en los *lags*  $i$  y  $\varepsilon_t$  es el vector de errores en el momento  $t$ , que se asume es un proceso estocástico multivariado con media cero y covarianza constante.

### 3.3. Variables y fuentes

Para responder a los objetivos del presente trabajo (realizar predicciones de la demanda turística en diferentes destinos de Argentina evaluando la utilidad de la información de las búsquedas en Google como una aproximación sobre las preferencias), se analizan las variables y utilizan las fuentes que se detallan a continuación.

#### 3.3.1. Variable de interés predictivo: demanda turística en Argentina

La variable de interés predictivo en esta investigación es la demanda turística en Argentina, reflejada a través de la cantidad de pernотaciones (variable que será denotada en este documento como *pernoct*) en diferentes destinos del país. La información sobre dicha variable proviene de la Encuesta de Ocupación Hotelera (EOH) publicada por el Instituto Nacional de Estadística y Censos (INDEC)<sup>1</sup> con periodicidad

<sup>1</sup>La Encuesta de Ocupación Hotelera (EOH) releva información sobre cantidad de establecimientos hoteleros y para-hoteleros, cantidad de habitaciones o unidades disponibles y ocupadas, cantidad de plazas disponibles y ocupadas, porcentaje de ocupación de habitaciones o unidades y de plazas, cantidad de viajeros, pernотaciones y duración promedio de la estadía. Según el informe de aspectos metodológicos de la EOH actualizado en mayo de 2020, se entiende por habitación/unidad disponible a todo cuarto o espacio amoblado ofrecido por un hotel, hostería, hospedaje o residencial, por un cierto precio por noche; por unidad disponible (para el caso de apart hotel, cabaña, bungalow) a cada departamento, cabaña o bungalow equipado que un establecimiento ofrece, y que otorga, además algunos de los servicios de hotelería; y por plazas disponibles al número total de camas fijas y supletorias. Las camas matrimoniales se contabilizan como dos plazas. El relevamiento comprende diferentes destinos de cada una de las siete regiones en las que dicha encuesta divide a Argentina, a saber, Región Buenos Aires, Región Ciudad de Buenos Aires, Región Córdoba, Región Cuyo, Región Litoral, Región Norte y Región Patagonia.

En lo que refiere a la demanda de servicios turísticos, esta fuente provee información, desde enero del año 2004 hasta el presente, respecto a las ya mencionadas **pernотaciones** (cantidad total de plazas ocupadas en establecimientos hoteleros de 1 a 5 estrellas, apart-hoteles y para-hoteleros tales como hoteles sindicales, albergues, cabañas, bungalows, hospedajes, bed and breakfast, hosterías y residenciales), como así también sobre **la cantidad de viajeros** (toda persona que se ha trasladado de su lugar de residencia habitual por razones de diversa índole, tales como el ocio, los negocios, la visita a familiares o amigos, que realiza una o más pernотaciones seguidas en el mismo establecimiento hotelero o para-hoteleros y que abona por tal servicio), y **la duración promedio de las estadías**. La primera de las mencionadas variables representa, en este documento, la evolución

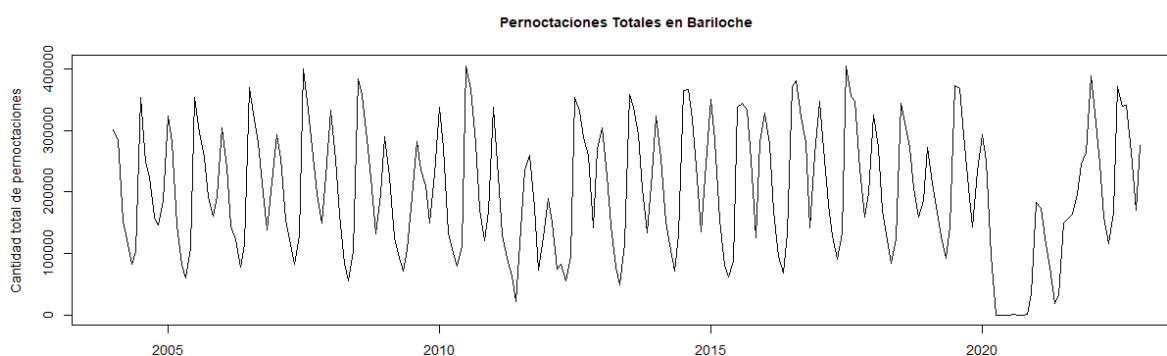
mensual y un rezago de casi dos meses.

Partiendo de la información publicada por INDEC a través de la EOH, y considerando la división geográfica dispuesta por dicha encuesta, se recopilieron datos sobre las pernoctaciones en una ubicación geográfica de cada región<sup>2</sup>. A continuación se detallan los destinos seleccionados.

- **Región Buenos Aires:** Pinamar.
- **Región Ciudad de Buenos Aires:** Ciudad Autónoma de Buenos Aires (CABA).
- **Región Córdoba:** Villa Carlos Paz.
- **Región Cuyo:** Ciudad de Mendoza.
- **Región Litoral:** Ciudad de Santa Fe.
- **Región Norte:** Ciudad de Jujuy.
- **Región Patagonia:** San Carlos de Bariloche<sup>3</sup>.

En la Figura 2, se expone la evolución de la cantidad de pernoctaciones en cada uno de los destinos turísticos mencionados precedentemente.

Figura 2: Pernoctaciones totales por destino turístico

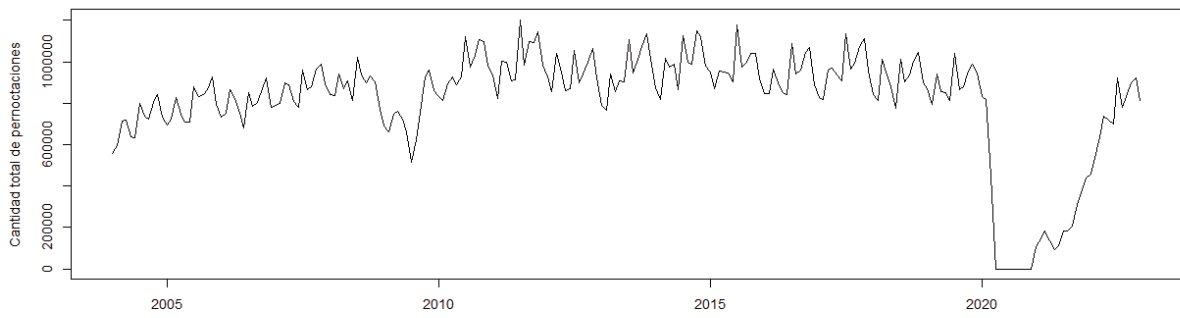


de la demanda turística.

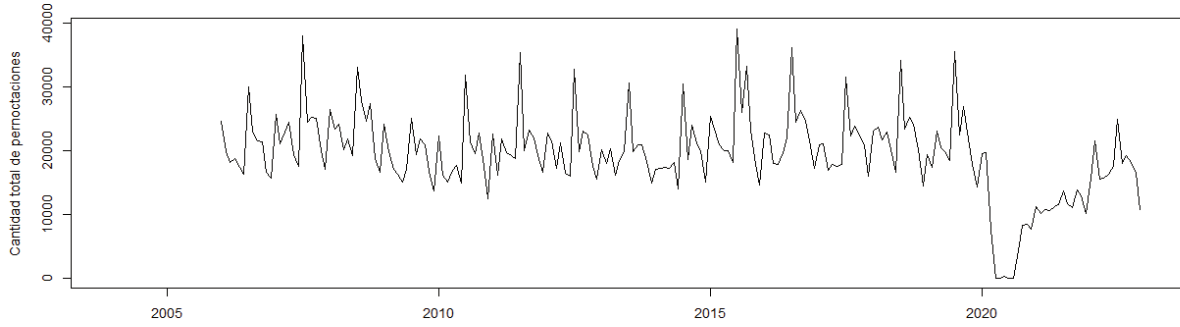
<sup>2</sup>Es importante destacar que las pernoctaciones, la cantidad de viajeros y la duración promedio de las estadías se dividen a su vez según se trate de turistas residentes o no residentes de nuestro país, y de alojamientos hoteleros o para-hoteleros. Teniendo en cuenta esto, y mediante una primera instancia de análisis exploratorio de los datos recopilados, se observa que las dos variables que evidencian mayor cantidad de observaciones relevadas son las pernoctaciones totales y la duración promedio de las estadías. En los casos de la cantidad de viajeros y las desagregaciones de las pernoctaciones y estadía promedio dentro de sus diferentes categorías, existen períodos en los que se discontinuó la publicación de los datos.

<sup>3</sup>De las series seleccionadas por ser las más completas (pernoctaciones y estadía promedio), se cuenta con información desde enero de 2004 para Pinamar, Villa Carlos Paz, Mendoza, Bariloche y CABA, desde enero de 2006 para Jujuy y desde enero de 2007 para Santa Fe. Por considerarla más pertinente, se opta por analizar la evolución de la **cantidad de pernoctaciones** como indicador de la demanda turística, debido a que arroja mayor información sobre la evolución de la misma que la aportada por la duración promedio de la estadía de los viajeros.

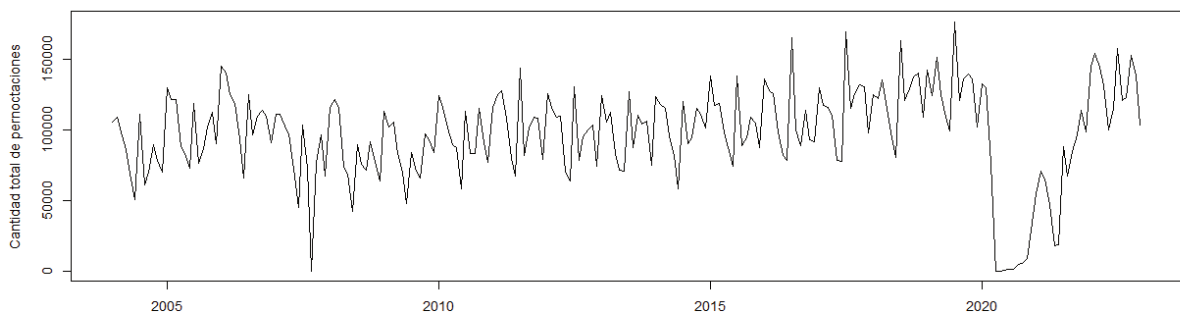
**Pernoctaciones Totales en Ciudad Autónoma de Buenos Aires**



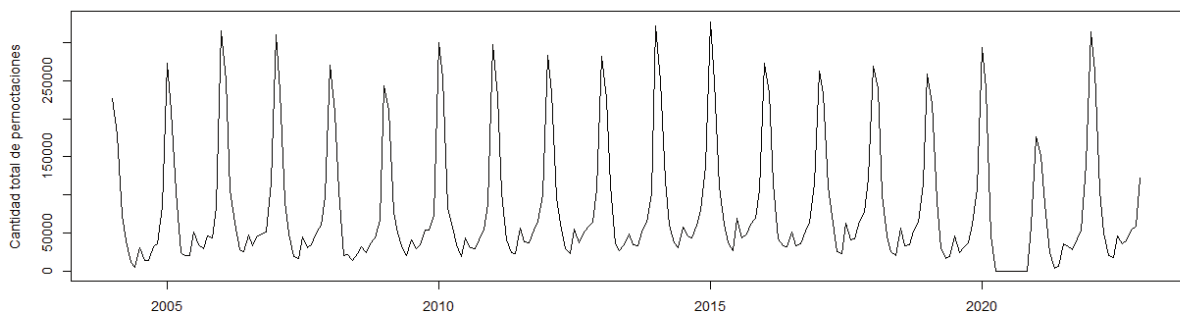
**Pernoctaciones Totales en Ciudad de Jujuy**

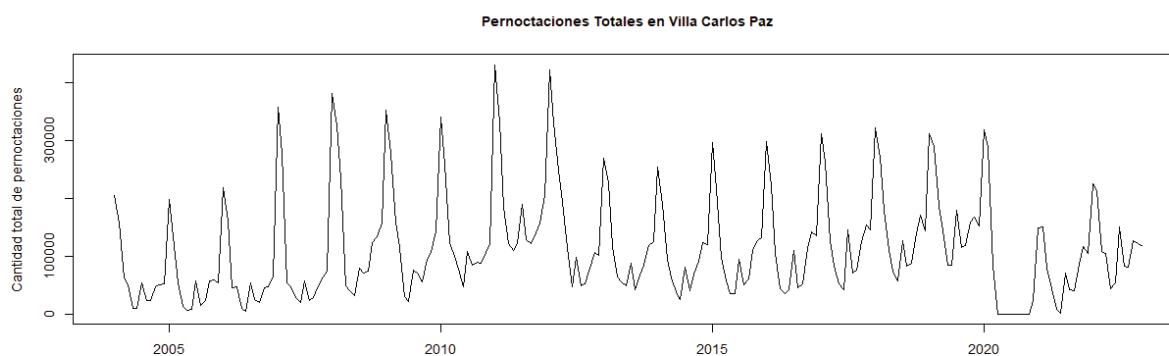
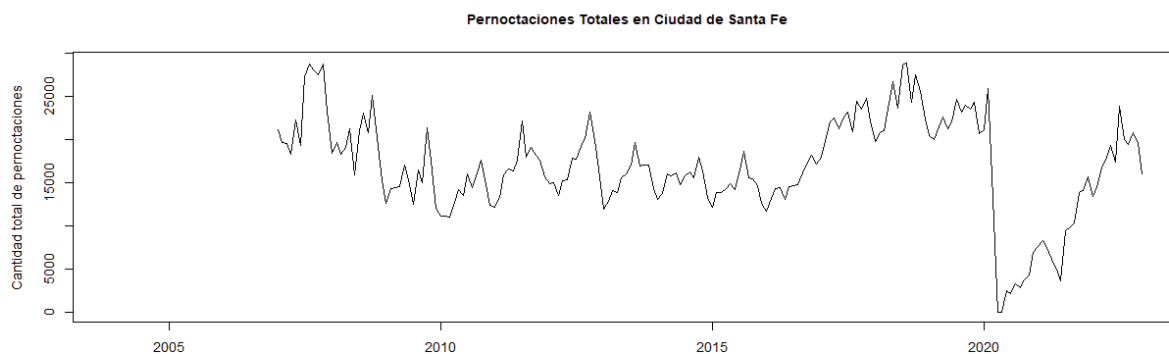


**Pernoctaciones Totales en Ciudad de Mendoza**



**Pernoctaciones Totales en Pinamar**





**Fuente:** elaboración propia en base a datos del Instituto Nacional de Estadística y Censos (INDEC).

Se observan al menos dos características económicas relevantes de las series correspondientes a las pernoctaciones.

En primer lugar, el patrón estacional, en algunos destinos más marcado que en otros, arroja información sobre la diferencia entre la cantidad de turistas que visitan cada destino durante los diferentes meses del año. Así, por ejemplo, las ciudades de Bariloche, Pinamar y Villa Carlos Paz poseen una marcada estacionalidad. En el caso de la primera, con dos picos importantes, uno en los meses invernales, propio de la temporada alta de nevadas, y otro en los meses finales del año, vinculado esto a la ocurrencia de los viajes de egresados que implican para la ciudad una considerable fuente de visitas. Para Pinamar y Villa Carlos Paz, la mayor cantidad de visitas se produce durante los meses de verano, por las características geográficas y climáticas propias de estos destinos.

Por otra parte, en todas las series bajo análisis se observa claramente el impacto de la pandemia de COVID-19 sobre la afluencia de turistas. Las restricciones a la circulación de personas, como medida preventiva para reducir la propagación del virus generaron que en cada uno de los casos analizados la cantidad de pernoctaciones cayera a cero durante al menos un par de meses. Con posterioridad a la pandemia, la cantidad de pernoctaciones registró un crecimiento que ubicó los valores de las series en sus niveles previos al comienzo de la misma.

### 3.3.2. Variables con potencial explicativo: búsquedas realizadas por usuarios de Google

Como variables que potencialmente explicarían la evolución de la demanda turística y que podrían ser utilizadas en los modelos multivariados, se consideran las búsquedas de palabras claves realizadas por los usuarios de Google en el motor de búsquedas de dicha firma (esta variable se denota en este documento como  $pc_k$ , donde  $k$  representa la cantidad de términos buscados a ser incluidos). Para ésto, se recurre a Google Trends<sup>4</sup>, un complemento de libre acceso y gratuito que permite conocer la evolución de las búsquedas de un determinado término o palabra clave (*keyword*) a lo largo de un período de tiempo específico, lo que posibilita identificar las variaciones en las búsquedas durante dicho período<sup>5</sup>. Los datos se encuentran disponibles desde enero del año 2004 hasta la actualidad.

Es importante destacar que, aun cuando su utilización resulta intuitiva y la información brindada puede ser de gran utilidad, esta herramienta posee como principal desventaja que los datos obtenidos respecto a la evolución de las búsquedas de las distintas palabras claves no se muestran en términos absolutos, sino mediante un indicador, con escala de 0 a 100, donde 100 representa el punto más alto en niveles de búsquedas de cada consulta en particular<sup>6</sup>.

Sin embargo, la anteriormente mencionada no es la única desventaja de Google Trends. Existe un limitante adicional que no tiene que ver estrictamente con el funcionamiento de la herramienta, y es que, al discontinuarse la publicación de datos en Google Correlate, actualmente no se cuenta con un mecanismo objetivo de selección de potenciales palabras claves cuyas búsquedas se encuentren vinculadas con el fenómeno que se desea estudiar<sup>7</sup>. Es por esto que, de manera subjetiva, se decidió seleccionar un conjunto de términos que *a priori*, y por consideración del autor, podrían tener algún grado de correspondencia con la cantidad de turistas que visitan los diferentes destinos de Argentina. Los términos seleccionados fueron: “Clima en ...”, “Hoteles en ...” y “Pasajes a ...” para cada uno de los siete destinos turísticos, y adicionalmente se consideró uno en particular para cada caso<sup>8</sup>. La Figura 3 expone la evolución de las búsquedas de cada palabra clave.

---

<sup>4</sup>Ver sitio: <https://trends.google.es/home>.

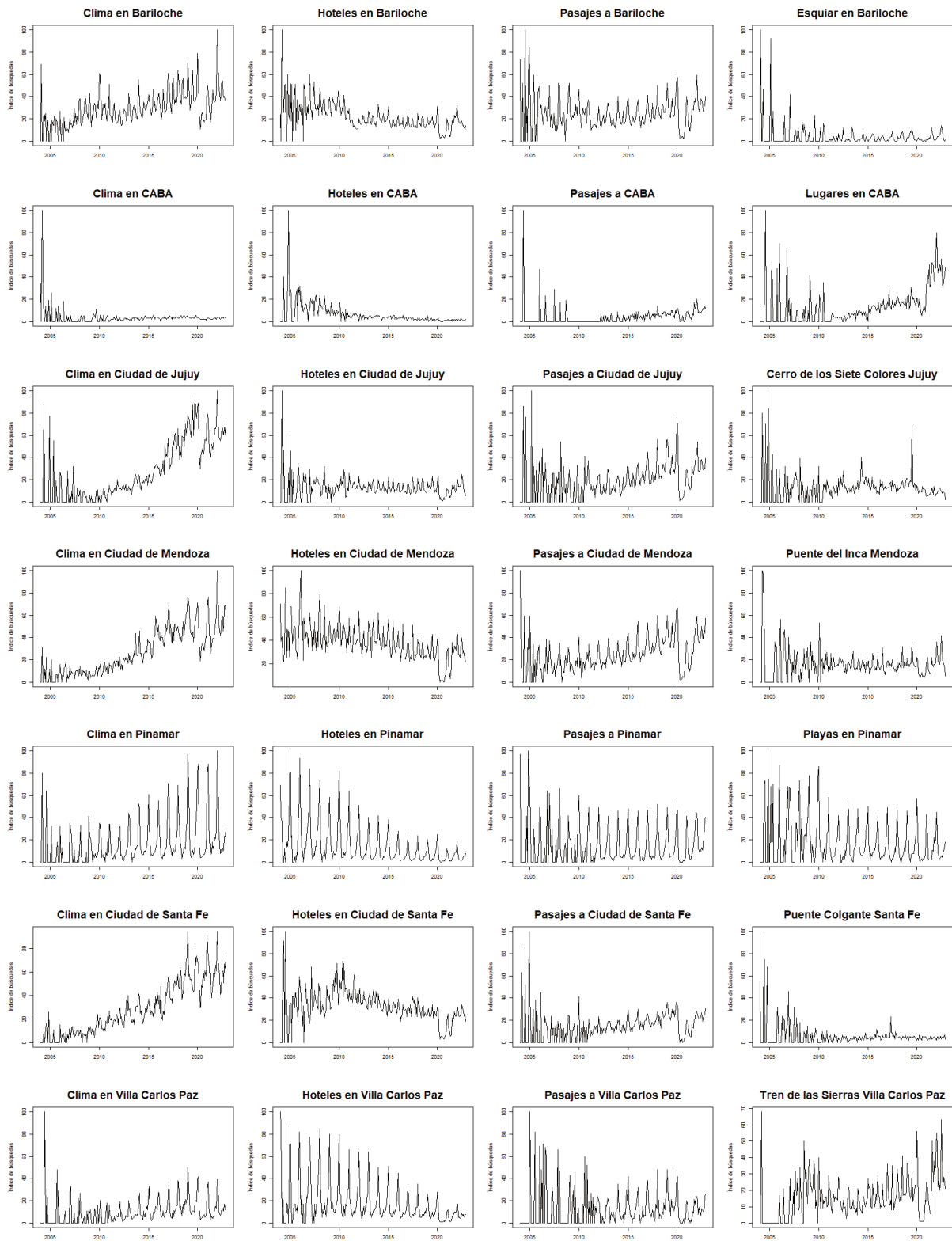
<sup>5</sup>La herramienta brinda la posibilidad de obtener información de las búsquedas con diferentes periodicidades, i.e. anual, mensual, semanal, e incluso por hora. El mecanismo de funcionamiento resulta intuitivo y de uso sencillo para el usuario. Este ingresa una palabra clave en el buscador, teniendo la posibilidad de aplicar filtros de búsqueda, entre los cuales se puede mencionar el área geográfica (cuya dimensión espacial mínima es la provincia), el período de tiempo, la categoría (arte, ciencias, compras, deportes, etc.) y el tipo de búsqueda (en la web, en noticias, imágenes). En el Anexo A se brindan más detalles acerca de la interface y la modalidad de uso de Google Trends.

<sup>6</sup>Este aspecto negativo fue oportunamente destacado por Blanco (2014).

<sup>7</sup>Este aspecto negativo fue destacado por Blanco (2014) y Naccarato et al. (2018).

<sup>8</sup>Los términos particulares seleccionados para cada destino son: Esquiar en Bariloche, Lugares en Ciudad Autónoma de Buenos Aires, Cerro de los Siete Colores Jujuy, Puente del Inca Mendoza, Puente Colgante Santa Fe, Playas en Pinamar y Tren de las Sierras Villa Carlos Paz.

Figura 3: Evolución de las búsquedas en Google de cada uno de los términos seleccionados



Fuente: elaboración propia en base a datos de Google.

Con el objetivo de predecir la evolución de la demanda turística, representada mediante la cantidad de pernотaciones en cada uno de los destinos seleccionados, se realizan a continuación las aclaraciones pertinentes acerca de la implementación del análisis. El mismo comprende tanto al modelo univariado como al multivariado, que internaliza información proveniente de Google para predecir el desempeño futuro de las variables de interés.

### 3.3.3. Modelos predictivos

#### Modelo univariado

Dado que la variable de interés predictivo son las pernотaciones en los diferentes destinos turísticos, el modelo ARIMA expresa el valor presente de dicha variable como función de valores pasados de la misma y los errores estocásticos ( $u$ ). De manera analítica:

$$pernoct_t^k = -(\Delta^d pernoct_t^k - pernoct_t^k) + \theta_0 + \sum_{i=1}^p \alpha_i \Delta^d pernoct_{t-i}^k + \sum_{j=1}^q \beta_j u_{t-j} + \varepsilon_t \quad (3.15)$$

donde  $pernoct_t^k$  representa a las pernотaciones en el destino  $k$  (con  $k =$  Bariloche, CABA, Jujuy, Mendoza, Pinamar, Santa Fe y Villa Carlos Paz), durante el momento  $t$ ,  $\theta_0$  es una constante,  $\alpha_i$  mide la incidencia de los valores de las pernотaciones en  $t - i$  sobre su valor actual, y  $\beta_j$  mide el efecto de los errores estocásticos ( $u$ ) en el momento  $t - j$  sobre el valor actual de las pernотaciones.

#### Modelo VAR

El modelo VAR comprende como variables endógenas a las pernотaciones como así también a las búsquedas realizadas por los usuarios en Google. En este caso, un modelo de orden  $p$ , se expresa de la siguiente forma analítica:

$$\begin{aligned} \begin{pmatrix} pernoct_t \\ pc_{1t} \\ \vdots \\ pc_{kt} \end{pmatrix} &= \begin{pmatrix} c_1 \\ c_2 \\ \vdots \\ c_k \end{pmatrix} + \begin{pmatrix} a_{11}^1 & a_{12}^1 & \cdots & a_{1k}^1 \\ a_{21}^1 & a_{22}^1 & \cdots & a_{2k}^1 \\ \vdots & \vdots & \ddots & \vdots \\ a_{k1}^1 & a_{k2}^1 & \cdots & a_{kk}^1 \end{pmatrix} \begin{pmatrix} pernoct_{t-1} \\ pc_{1t-1} \\ \vdots \\ pc_{kt-1} \end{pmatrix} \\ &+ \begin{pmatrix} a_{11}^2 & a_{12}^2 & \cdots & a_{1k}^2 \\ a_{21}^2 & a_{22}^2 & \cdots & a_{2k}^2 \\ \vdots & \vdots & \ddots & \vdots \\ a_{k1}^2 & a_{k2}^2 & \cdots & a_{kk}^2 \end{pmatrix} \begin{pmatrix} pernoct_{t-2} \\ pc_{1t-2} \\ \vdots \\ pc_{kt-2} \end{pmatrix} + \cdots \\ &+ \begin{pmatrix} a_{11}^p & a_{12}^p & \cdots & a_{1k}^p \\ a_{21}^p & a_{22}^p & \cdots & a_{2k}^p \\ \vdots & \vdots & \ddots & \vdots \\ a_{k1}^p & a_{k2}^p & \cdots & a_{kk}^p \end{pmatrix} \begin{pmatrix} pernoct_{t-p} \\ pc_{1t-p} \\ \vdots \\ pc_{kt-p} \end{pmatrix} + \begin{pmatrix} \varepsilon_{1t} \\ \varepsilon_{2t} \\ \vdots \\ \varepsilon_{kt} \end{pmatrix} \end{aligned}$$

donde  $pernoct_{t-i}$  representa a las pernотaciones en el destino específico en el momento  $t - i$ ,  $pc_{kt-i}$  denota a las búsquedas de la palabra clave  $k$  realizada por usuarios de Google en el momento  $t - i$ ,  $c$  es un vector

constante de dimensiones  $k \times 1$ ,  $a_i$  son matrices de coeficientes de dimensiones  $k \times k$  que cuantifican la fuerza y dirección de las relaciones de autorregresión entre las variables en los *lags*  $i$  y  $\varepsilon_t$  es el vector de errores en el momento  $t$ , que se asume es un proceso estocástico multivariado con media cero y covarianza constante.

### 3.4. Proceso de implementación del análisis

Siguiendo los trabajos de Naccarato et al. (2018), y Choi y Varian (2012)<sup>9</sup>, se utilizan modelos ARIMA y VAR para la realización de las estimaciones sobre la evolución de la demanda turística (representada en este caso por la cantidad de pernoctaciones) en los diferentes destinos seleccionados.

Específicamente, el procedimiento se realiza de la siguiente manera. En primer lugar, se analiza la estacionariedad de las series de tiempo con interés predictivo, con el objetivo de definir la estructura de los modelos a aplicar. Para ello se utilizan las pruebas de raíces unitarias de Dickey-Fuller sin intercepto ni tendencia, con intercepto, con tendencia, y la versión aumentada.

En segundo lugar, y en base a los resultados obtenidos a través de las pruebas de estacionariedad, se procede a determinar el orden de cada uno de los componentes en ambos modelos, es decir, el orden del componente autorregresivo (AR), de integración (I) y de suavización a través de medias móviles (MA) para el modelo ARIMA<sup>10</sup> y el orden de autorregresión para el modelo VAR<sup>11</sup>. La selección de dichos valores responde a los criterios comúnmente utilizados<sup>12</sup>. Este procedimiento representa la etapa de selección de la especificación de los modelos y la estimación de los parámetros. Una vez seleccionados los órdenes de los parámetros para cada caso se aplican pruebas de autocorrelación serial, homocedasticidad y normalidad de los residuos para verificar la estabilidad del modelo.

Finalmente, se realizan las predicciones correspondientes con el ARIMA y el VAR para cada uno de los destinos turísticos. Si bien se cuenta con datos mensuales tanto de las pernoctaciones como de las búsquedas en Google desde enero de 2004 hasta diciembre de 2023, se realizan las estimaciones de los modelos (ARIMA y VAR para cada uno de los destinos turísticos) utilizando los datos comprendidos entre enero 2004 y diciembre 2022, período denominado “para modelación”, sobre el cual se estiman los valores de los coeficientes correspondientes. Luego, se realizan predicciones sobre el período que va desde enero a diciembre de 2023 (por “fuera de la muestra”). Esta estrategia permite comparar las predicciones con valores reales, es decir, se predice sobre un período para el cual se tiene las observaciones con el objetivo de comparar la fiabilidad de las predicciones. De esta manera, con la realización de estimaciones fuera de la muestra se puede medir la eficacia de cada alternativa a partir de los errores observados<sup>13</sup> y, determinar así, si la inclusión de información sobre las búsquedas de los usuarios de Google contribuye a mejorar las predicciones.

Es importante destacar que para la implementación del análisis se utiliza R, el lenguaje de programación y entorno de software libre para análisis estadístico y gráfico, por la gran cantidad de herramientas de las que dispone, su uso generalizado, su versatilidad y gratuidad. Adicionalmente, se recurre a diferentes librerías de dicho software que tienen como propósito tanto el análisis econométrico en sí (dentro de las que se pueden

---

<sup>9</sup>En ambos documentos se comparan los resultados obtenidos a través de modelos univariados y multivariados (que incluyen información proveniente de Google) para la predicción de variables económicas.

<sup>10</sup>El proceso de selección de los parámetros  $p$ ,  $d$  y  $q$  se llevó adelante mediante la implementación del programa AutoARIMA en R. Para los casos de variables no estacionarias en niveles, se obtuvieron valores de  $d$  iguales a 1 o valores de  $q$  mayores a 0, lo que permitió corregir la no estacionariedad de las series correspondientes.

<sup>11</sup>El proceso implicó la utilización del comando VARSelect, que permite estimar modelos VAR con las variables que resultan significativas y la cantidad óptima de rezagos en base a los criterios de selección más utilizados.

<sup>12</sup>Los criterios más difundidos son el de selección de Akaike (AIC), de selección de Bayes (BIC) y el criterio de Hannan-Quinn (HQ).

<sup>13</sup>Como indicador para la comparación de la eficacia de los modelos se utiliza la raíz del error cuadrático medio (RMSE, por sus siglas en inglés).

mencionar a las siguientes: `forecast`, `vars`, `dynlm`, `lmtest`, `urca`, `tseries` y `mFilter`) como mejorar la interface de exposición de gráficos y cuadros (`ggplot2`, `kableExtra`, `knitr`, `scales`, `PerformanceAnalytics`, `tidyverse` y `fpp2`).

## Capítulo 4

# Resultados: prediciendo la demanda turística en Argentina

El presente capítulo expone los resultados de la aplicación del modelo univariado (ARIMA) y del multivariado (VAR) para cada uno de los destinos turísticos analizados. Adicionalmente, se comparan los resultados obtenidos mediante ambas variantes con el objetivo de determinar si la inclusión de información proveniente de Google mejora las estimaciones realizadas a través del modelo univariado.

### 4.1. Análisis de las condiciones de partida para la modelación

Como se mencionó en el Capítulo 3, tanto la modelación univariada como la multivariada requieren que las series de tiempo sean estacionarias, ya que de lo contrario, el valor futuro de la variable que se desea explicar dependerá exclusivamente de su valor actual más un choque aleatorio que es difícil predecir mediante la modelación y, por tanto, solo podremos estudiar su comportamiento durante el periodo bajo análisis, no resultando posible generalizar las conclusiones para periodos futuros.

A partir de la aplicación de la prueba ADF a las pernoctaciones en cada una de las ciudades consideradas, se concluye que son estacionarias en nivel las correspondientes a Bariloche, Mendoza, Pinamar y Villa Carlos Paz, mientras que las series de pernoctaciones de CABA, Jujuy y Santa Fe no resultan estacionarias 95 % de confianza, aunque sí sus primeras diferencias.

De manera análoga, se realizaron las pruebas ADF para cada una de las series de búsquedas en Google seleccionadas para la modelación multivariada. Todos los índices de búsqueda resultaron estacionarios al 90 % de significancia, excepto Clima en Jujuy, Hoteles en Santa Fe, Lugares en CABA y Cerro de los Siete Colores en Jujuy.

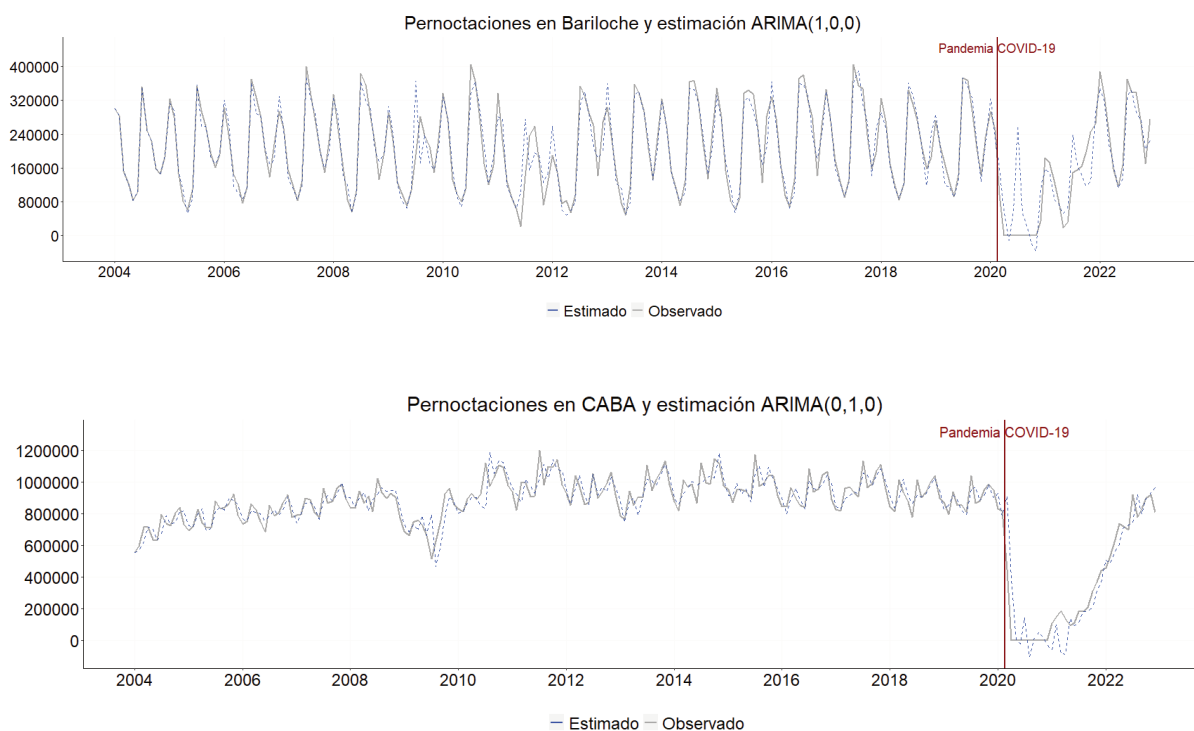
Un mayor nivel de detalle sobre los resultados específicos de cada una de las pruebas de raíces unitarias, se presenta en los cuadros A1 y A2 del Anexo A. Las mismas contienen los resultados de las pruebas de Dickey-Fuller sin intercepto ni tendencia, con intercepto, con tendencia y aumentada para las series originales y en primeras diferencias de las pernoctaciones y para las series originales de búsquedas.

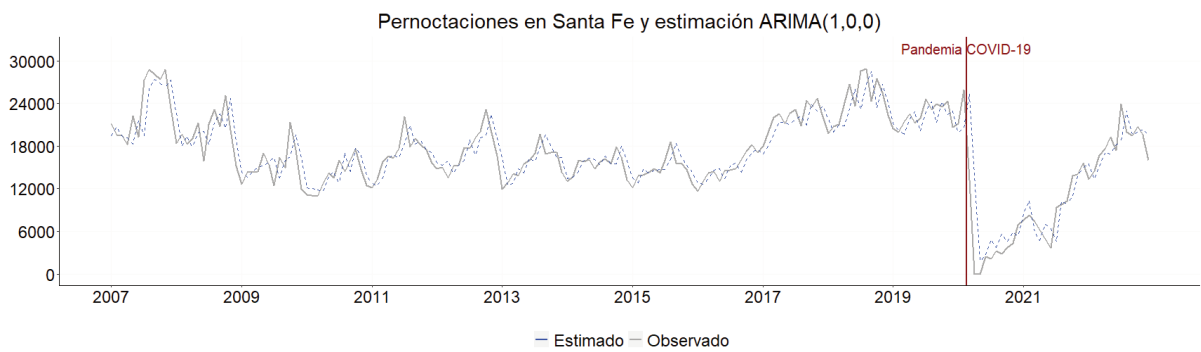
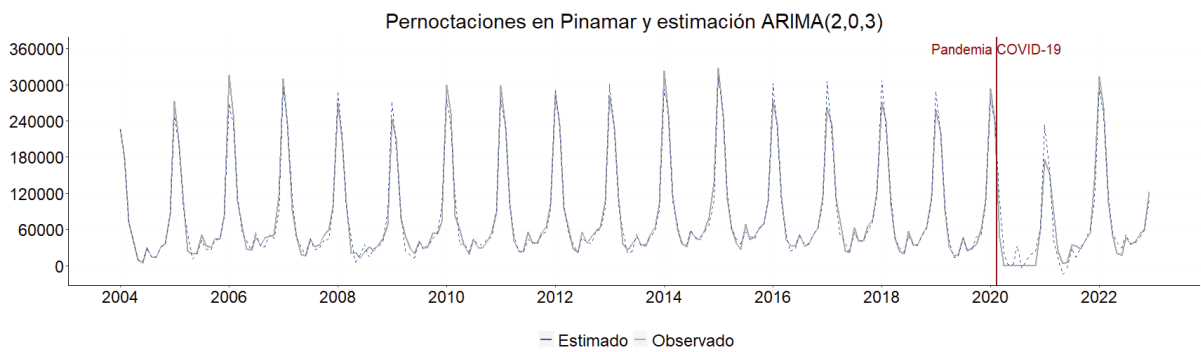
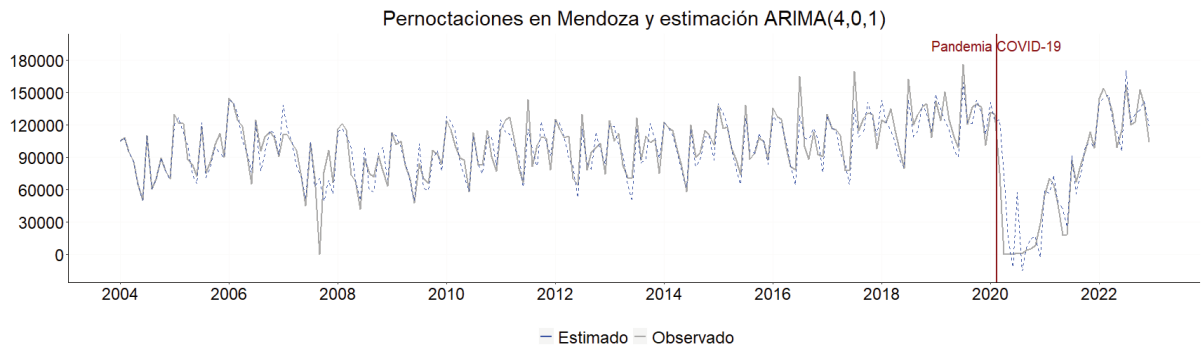
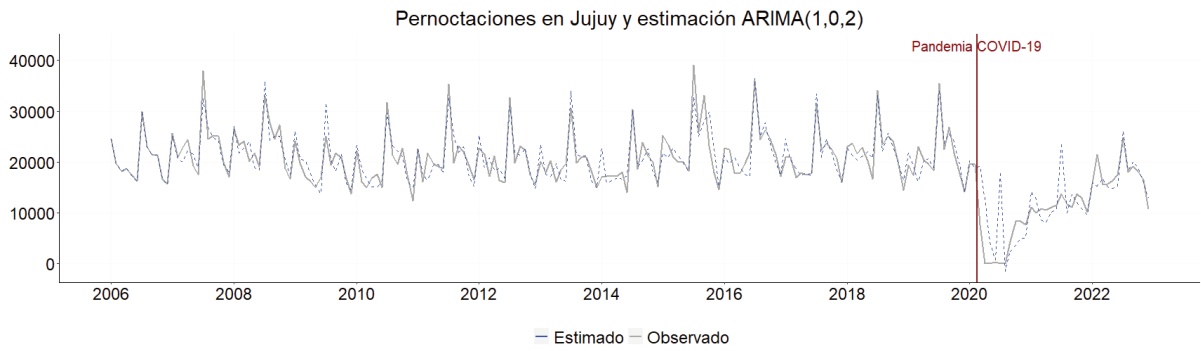
## 4.2. Los modelos univariados para la prección de la demanda turística

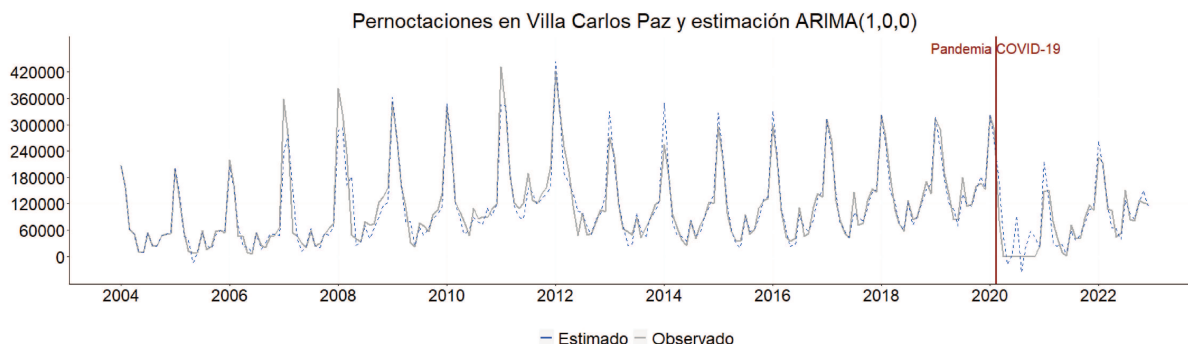
En relación al modelo univariado y, habiendo constatado la estacionariedad de todas las series relacionadas con las pernoctaciones de las ciudades, se realizó la selección de los parámetros  $p$ ,  $d$  y  $q$  de los modelos  $ARIMA(p, d, q)$  específicos de cada una de ellas siguiendo la metodología y procedimiento explicado en la sección 3.4. Los parámetros, propiamente dichos, se exponen en el Cuadro A3 del Anexo A, mientras que en las figuras que van desde B1 a B7 del Anexo B se pueden visualizar los residuos de cada uno de los modelos, mediante los cuales se puede constatar que los mismos parecen mostrar una estructura compatible con la distribución normal aunque registran algunos valores atípicos que pueden explicarse, principalmente, por los efectos de la pandemia de COVID-19. Por lo demás, los residuos oscilan alrededor de cero sin cambios sostenidos en su variabilidad a lo largo del tiempo.

Con los modelos  $ARIMA$  seleccionados en cada caso se realizaron las estimaciones de la evolución de las pernoctaciones dentro de la muestra (período comprendido entre enero de 2004 y diciembre de 2022), comparando los resultados obtenidos con los valores reales observados. La Figura 4 expone ambas series para cada destino bajo estudio.

Figura 4: Valores observados y estimados dentro de la muestra (enero 2004 a diciembre 2022) mediante modelos  $ARIMA$







**Fuente:** elaboración propia en base a datos del Instituto Nacional de Estadística y Censos (INDEC).

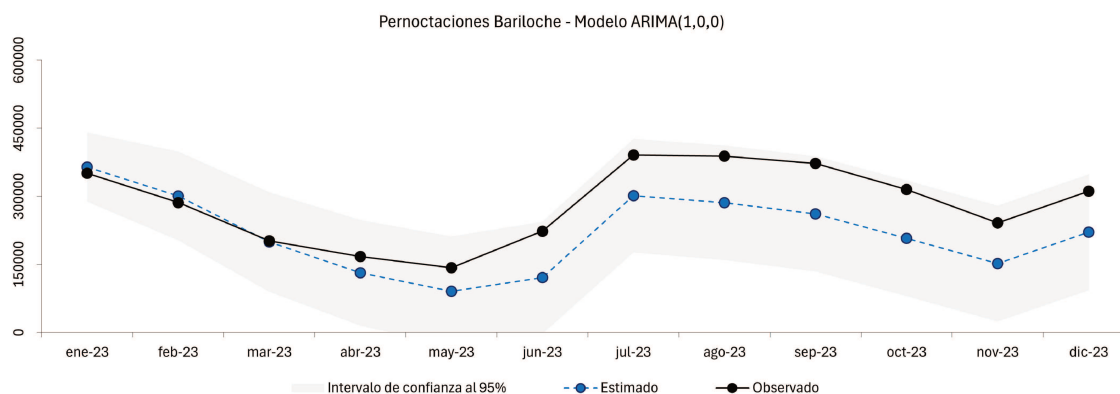
Como es de esperarse, el modelo ARIMA seleccionado para cada variable replica bastante acertadamente la evolución de los verdaderos valores de la misma, incluso en períodos en los que las pernoctaciones muestran comportamientos atípicos (como podría ser el mencionado período de la pandemia de COVID-19), y tanto para las series con un marcado patrón estacional —como los casos de Bariloche, Pinamar y Villa Carlos Paz—, como para aquellas que no lo tienen —CABA, Jujuy, Mendoza y Santa Fe—.

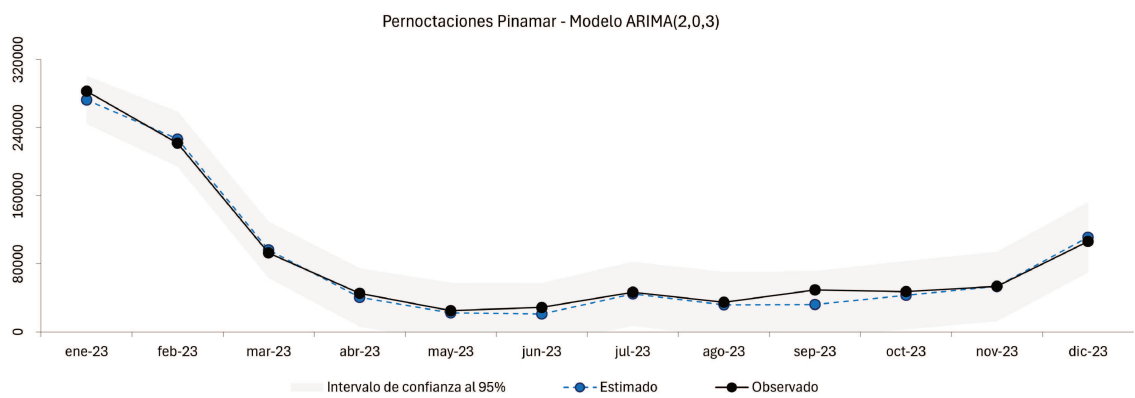
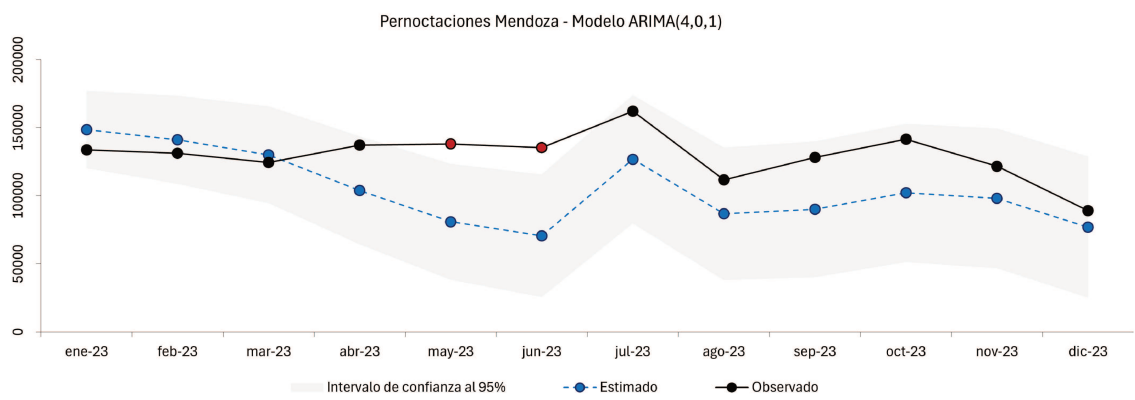
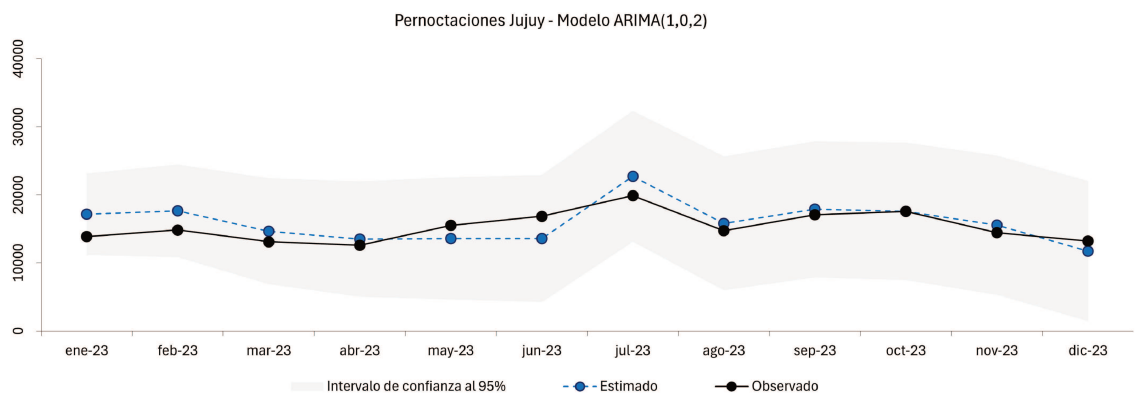
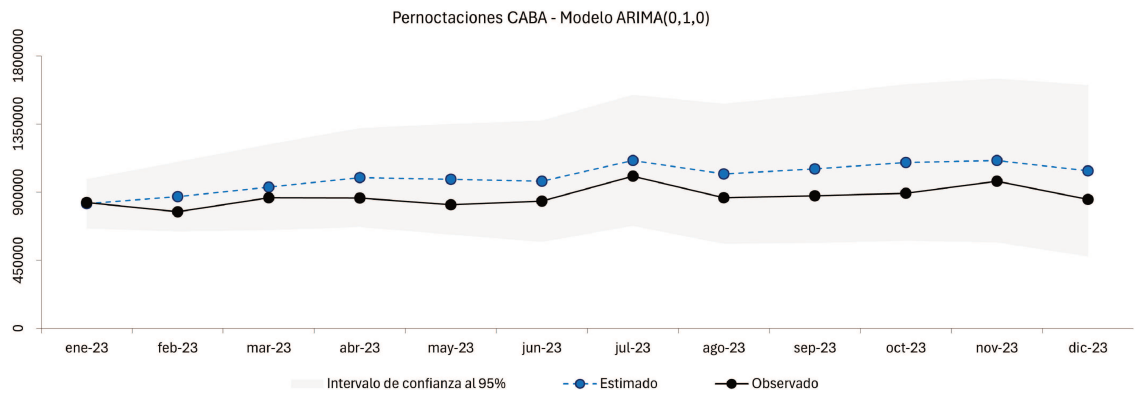
En lo que refiere a la relación entre precisión y parsimonia de los modelos obtenidos, es decir, considerando un equilibrio entre parametrización y ganancia de precisión, en líneas generales no se observa que los modelos con menor cantidad de parámetros sean menos precisos que aquellos con mayor grado de especificación. Por ejemplo, para el caso de la Ciudad de Santa Fe, cuyo modelo es un ARIMA(1,0,0), la raíz del error cuadrático medio (RMSE) es la menor de los siete destinos, mientras que también se encuentra dentro de los modelos con menor RMSE el correspondiente a la Ciudad de Mendoza, cuya estructura es un ARIMA(4,0,1).

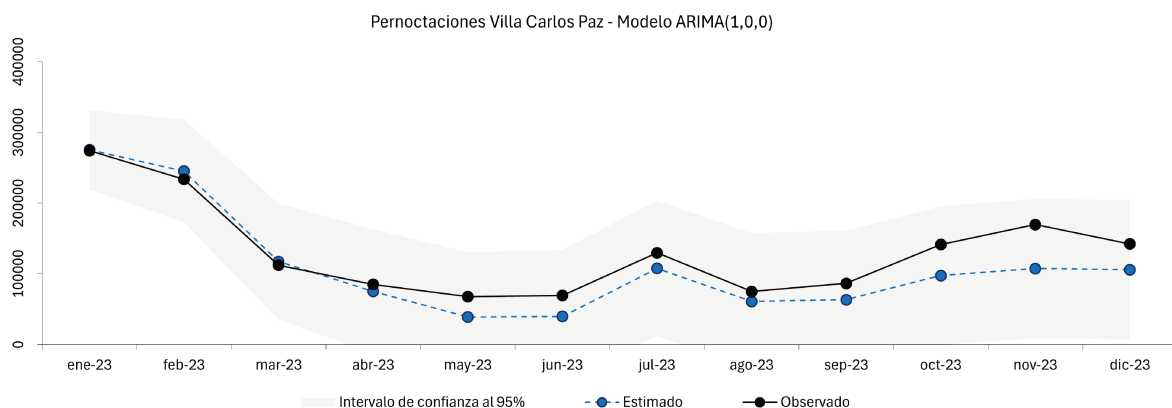
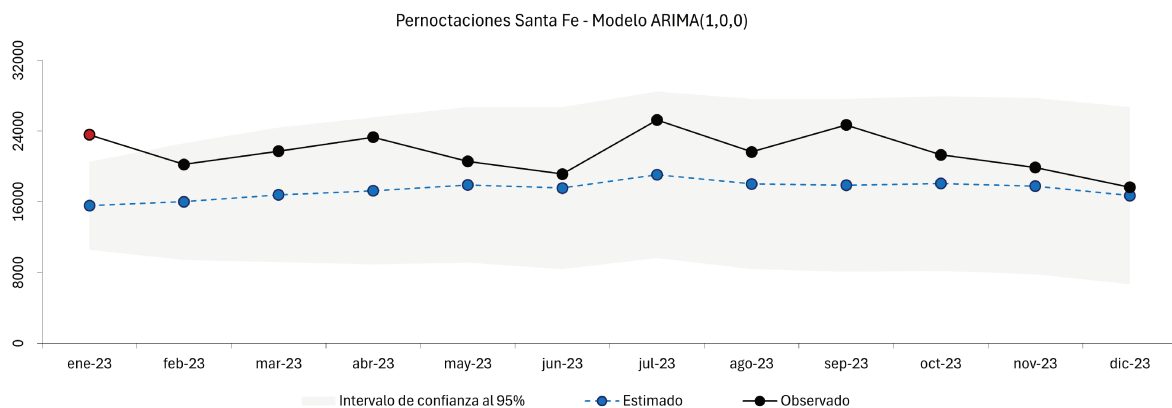
#### 4.2.1. Predicciones fuera de la muestra con los modelos univariados

Utilizando cada uno de los modelos seleccionados, se realizan estimaciones fuera de la muestra y se comparan dichas estimaciones con los verdaderos valores de la demanda turística. La Figura 5 muestra la evolución de las pernoctaciones y las estimaciones con intervalos de confianza al 95 % en cada caso.

Figura 5: Valores observados, estimación fuera de la muestra (enero a diciembre 2023) e intervalos al 95 % de confianza obtenidos mediante modelos ARIMA







**Fuente:** elaboración propia en base a datos del Instituto Nacional de Estadística y Censos (INDEC) y Google.

De las predicciones realizadas con los modelos seleccionados se pueden mencionar algunos aspectos relevantes. En primer lugar, se evidencia un grado de precisión heterogéneo entre los diferentes casos, observándose que para los tres destinos turísticos con patrones estacionales más marcados —Bariloche, Pinamar y Villa Carlos Paz—, los valores reales en cada uno de los meses se encuentran dentro del intervalo de confianza del modelo estimado. Esto también se cumple para CABA y Jujuy, que poseen un patrón estacional menos marcado que las tres antes mencionadas.

Por otra parte, se observa que solamente en las predicciones de Mendoza y Santa Fe, los valores reales se ubicaron fuera de los intervalos de confianza de las estimaciones, para Mendoza en los meses de mayo y junio de 2023, y para la Ciudad de Santa Fe en enero de 2023. Adicionalmente, es importante destacar que las mejores estimaciones corresponden a Pinamar, Jujuy y CABA, respectivamente, presentando los tres destinos grados de estacionalidad diferentes.

Por lo mencionado, se concluye que el grado de precisión de los modelos ARIMA es elevado y no guarda correlación con el nivel de estacionalidad de las series.

Complementando lo expuesto en la Figura 5, el Cuadro 1 muestra la evolución de los valores reales, las predicciones y sus intervalos de confianza, y los errores promedios de estimación para cada uno de los modelos ARIMA.

Cuadro 1: Error de estimación para cada mes de 2023 y promedio anual de los modelos ARIMA

	Bariloche	CABA	Jujuy	Mendoza	Pinamar	Santa Fe	VCP
	ARIMA	ARIMA	ARIMA	ARIMA	ARIMA	ARIMA	ARIMA
	(1,0,0)	(0,1,0)	(1,0,2)	(4,0,1)	(2,0,3)	(1,0,0)	(1,0,0)
ene-23	4,0%	1,0%	23,8%	11,1%	3,6%	34,0%	0,5%
feb-23	5,1%	13,0%	19,1%	7,5%	2,2%	20,9%	5,0%
mar-23	1,3%	8,0%	11,9%	4,5%	3,8%	22,7%	4,6%
abr-23	21,5%	15,8%	7,3%	24,3%	10,2%	26,1%	11,6%
may-23	36,2%	20,5%	12,4%	41,4%	10,9%	12,9%	42,4%
jun-23	45,9%	15,7%	19,3%	47,8%	27,0%	8,3%	42,7%
jul-23	22,9%	10,4%	14,2%	21,7%	4,1%	24,5%	16,9%
ago-23	26,5%	18,2%	7,3%	22,3%	8,5%	16,7%	18,6%
sep-23	29,9%	20,4%	4,8%	29,8%	34,7%	27,6%	26,6%
oct-23	34,1%	22,9%	0,0%	27,9%	8,9%	15,2%	31,0%
nov-23	37,2%	14,1%	7,7%	19,4%	0,2%	10,7%	36,7%
dic-23	29,0%	22,2%	11,0%	13,6%	4,6%	5,3%	25,6%
Promedio 2023	24,5%	15,2%	11,6%	22,6%	9,9%	18,7%	21,8%

Fuente: elaboración propia en base a datos del Instituto Nacional de Estadística y Censos (INDEC).

La tendencia de los errores evidencia que para Bariloche, CABA y Villa Carlos Paz, el grado de precisión del modelo univariado se reduce a medida que se predicen fechas más lejanas en el tiempo. Esto implica que los coeficientes estimados en un momento determinado pierden precisión con el correr de los meses, siendo recomendable realizar nuevas estimaciones de los mismos a medida que se dispone de nuevos datos. En líneas generales, una actualización trimestral de los coeficientes podría dar buenos resultados, ya que para las primeras tres predicciones el modelo ARIMA evidencia un elevado grado de precisión para los destinos turísticos mencionados.

Contrariamente, para los casos de Jujuy y Santa Fe, el error es menor para los meses más lejanos, mientras que para Mendoza y Pinamar el mismo se mantiene a lo largo de la predicción de cada uno de los doce meses de 2023.

Habiendo evaluado el grado de precisión de los modelos ARIMA, se realiza lo propio con los modelos VAR, con el objetivo de poner a prueba la hipótesis central de este documento, que propone que la información proveniente de las búsquedas en Google resultan variables endógenas relevantes para explicar la evolución de las pernoctaciones en los diferentes destinos seleccionados.

### 4.3. Analizando la eficacia de las búsquedas en Google como variables predictoras

De manera análoga al caso de los modelos univariados, en primer lugar se constata la estacionariedad de las series de las pernoctaciones en cada uno de los destinos turísticos y de las búsquedas realizadas por los usuarios en el motor de búsquedas de Google<sup>1</sup>. Posteriormente, se seleccionan los parámetros  $p$  más adecuados para la estimación del modelo VAR( $p$ ) y se define qué búsquedas incluir y cuáles no en cada caso, en concordancia con la metodología y los procedimientos establecidos en la sección 3.4.

El Cuadro 2 expone los valores de dichos parámetros para cada serie, los cuales fueron seleccionados

<sup>1</sup>Los resultados de las pruebas de raíces unitarias se encuentran en las ya mencionados cuadros A1 y A2 del Anexo A.

considerando el criterio de información de Akaike (AIC). Por su parte, en las figuras que van desde B8 a B14 del Anexo B se pueden visualizar los residuos de cada variable endógena en cada uno de los modelos.

Cuadro 2: Selección de parámetros  $p$  y variables endógenas para modelos multivariados

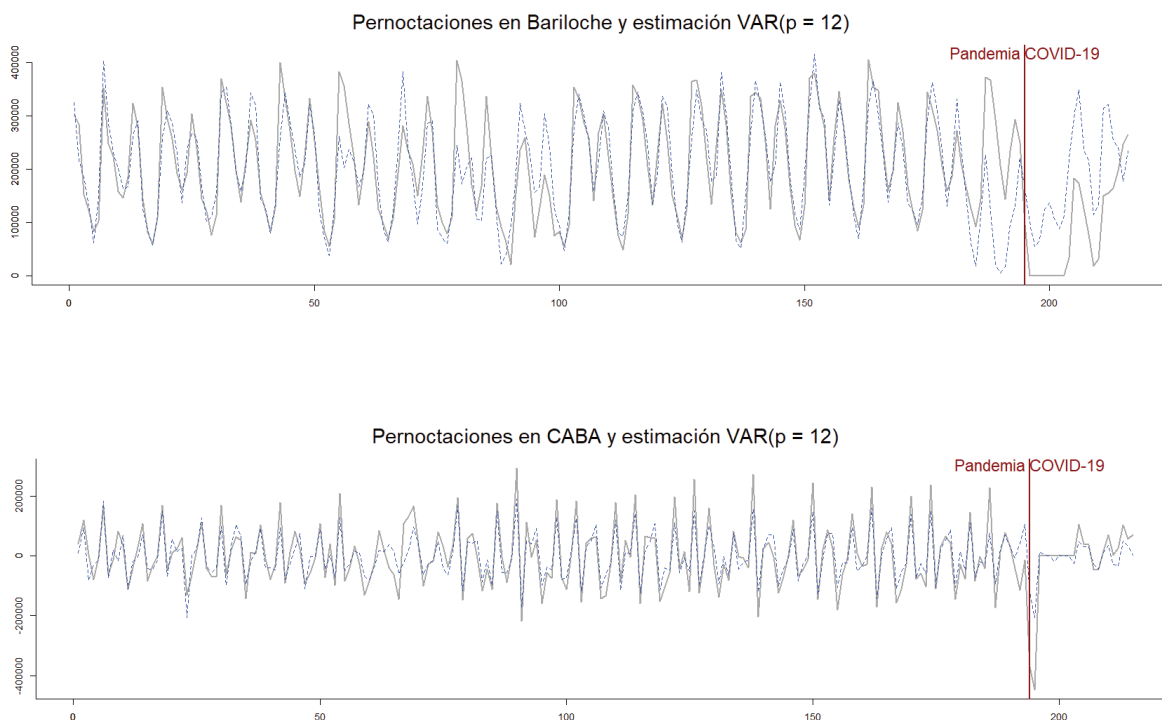
Ubicación	Modelo	Variables	AIC
Bariloche	VAR(12)	Pernoctaciones, Clima y Esquiar	8.265,71
Ciudad Autónoma de Buenos Aires	VAR(12)	Dif(Pernoctaciones), Dif(Clima) y Dif(Hoteles)	7.645,07
Ciudad de Jujuy	VAR(12)	Dif(Pernoctaciones), Dif(Clima), Dif(Hoteles) y Dif(Pasajes)	7.043,10
Ciudad de Mendoza	VAR(12)	Pernoctaciones y Hoteles	6.393,84
Pinamar	VAR(12)	Pernoctaciones, Clima, Hoteles, Pasajes y Playas	10.709,79
Ciudad de Santa Fe	VAR(5)	Dif(Pernoctaciones) y Dif(Puente)	4.266,56
Villa Carlos Paz	VAR(12)	Pernoctaciones y Clima	6.718,81

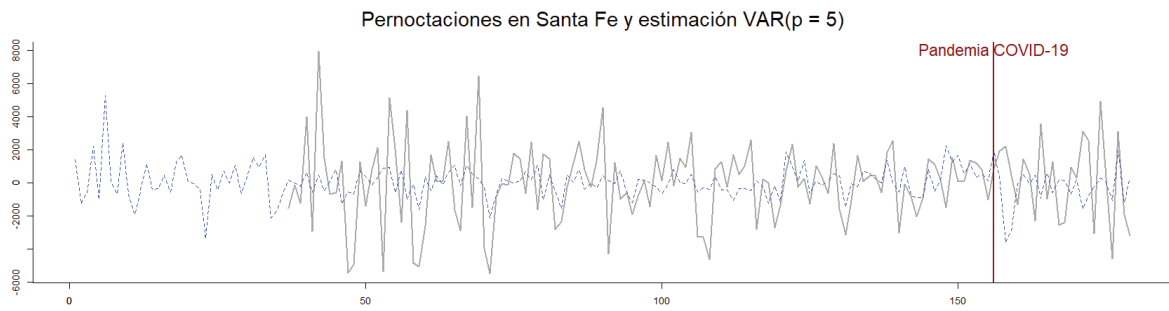
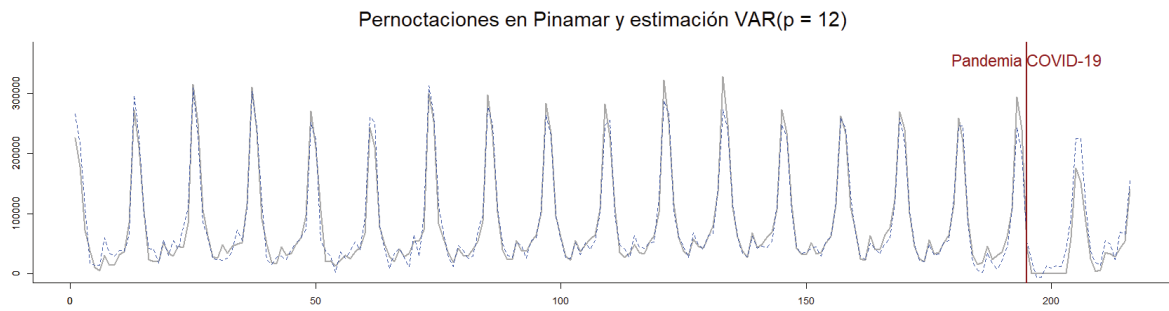
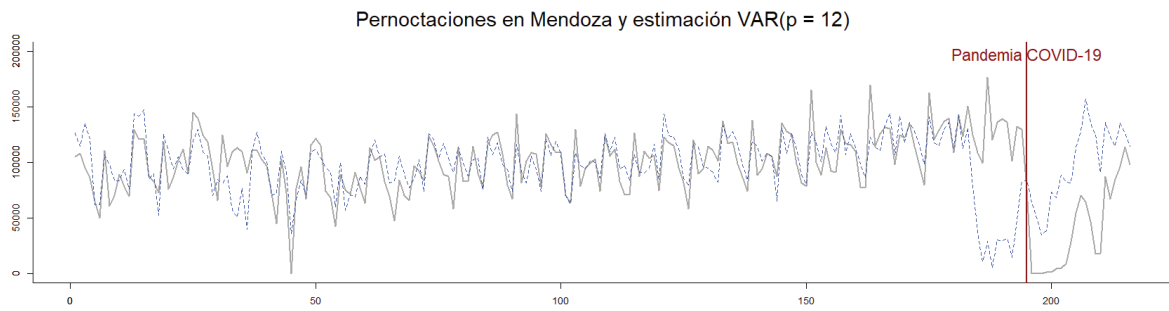
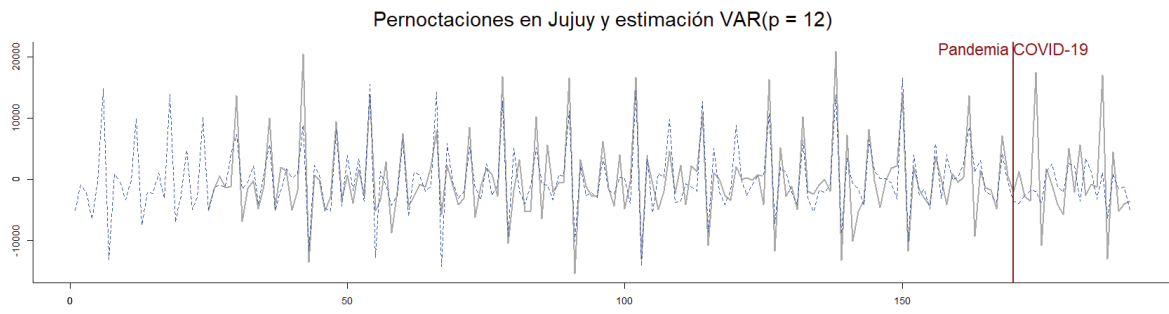
Fuente: elaboración propia en base a datos del Instituto Nacional de Estadística y Censos (INDEC).

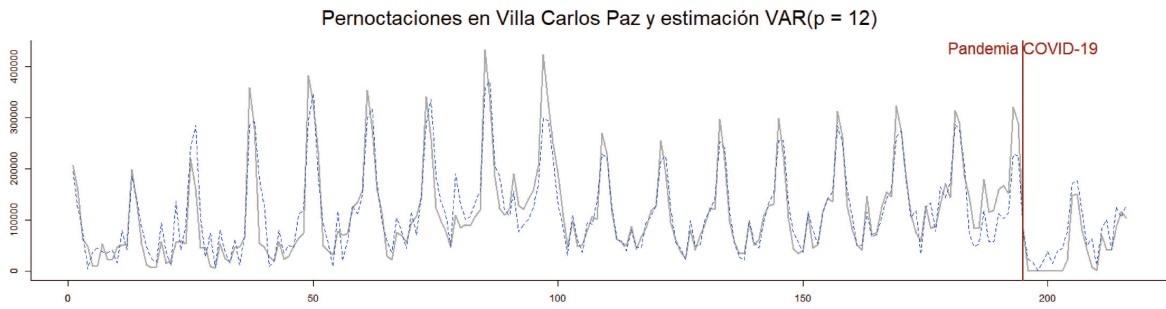
Como se puede observar, para los casos de CABA, Jujuy y Santa Fe, se decide incluir, tanto las pernoctaciones como las búsquedas, en sus primeras diferencias y no así las series originales. Esto se debe a que las series correspondientes a las pernoctaciones en esos destinos no resultan estacionarias, y con el objetivo de que la totalidad de las variables endógenas del modelo VAR se encuentren en iguales niveles de integración, se opta por esta metodología. Es importante destacar que, si bien se adopta el procedimiento mencionado, la metodología más utilizada para la estimación multivariada de series no estacionarias son los modelos de vectores de corrección del error (VECM).

Con los modelos seleccionados para cada caso, en primer lugar se realizaron las predicciones dentro de la muestra. Las mismas se exponen en la Figura 6.

Figura 6: Valores observados y estimados dentro de la muestra mediante modelos VAR







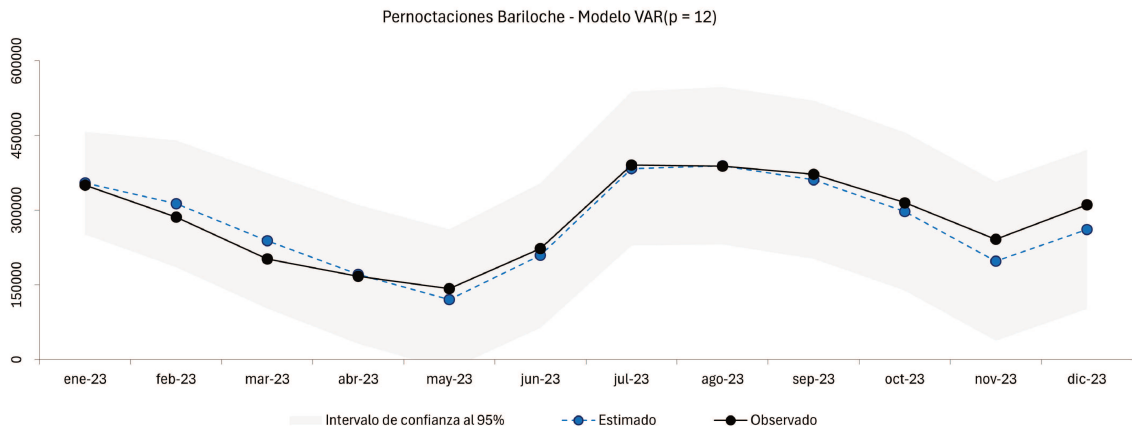
**Fuente:** elaboración propia en base a datos del Instituto Nacional de Estadística y Censos (INDEC) y Google.

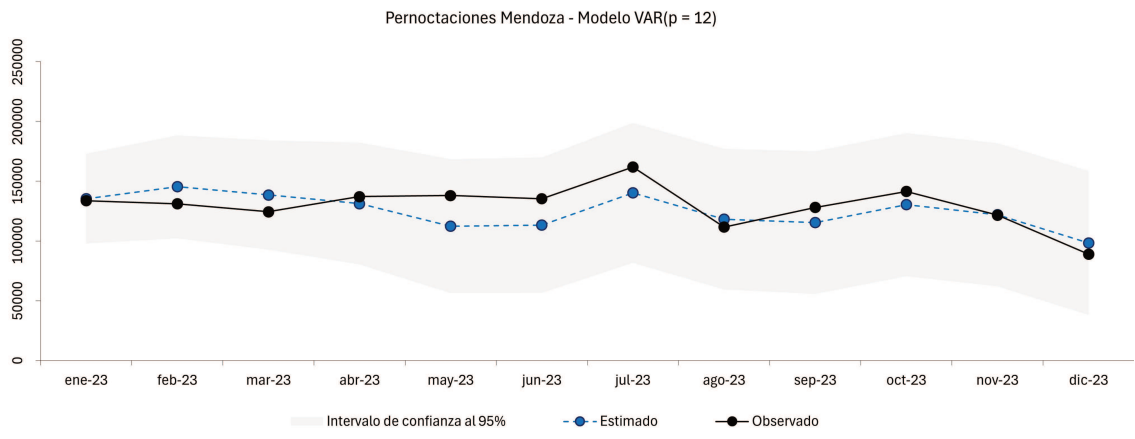
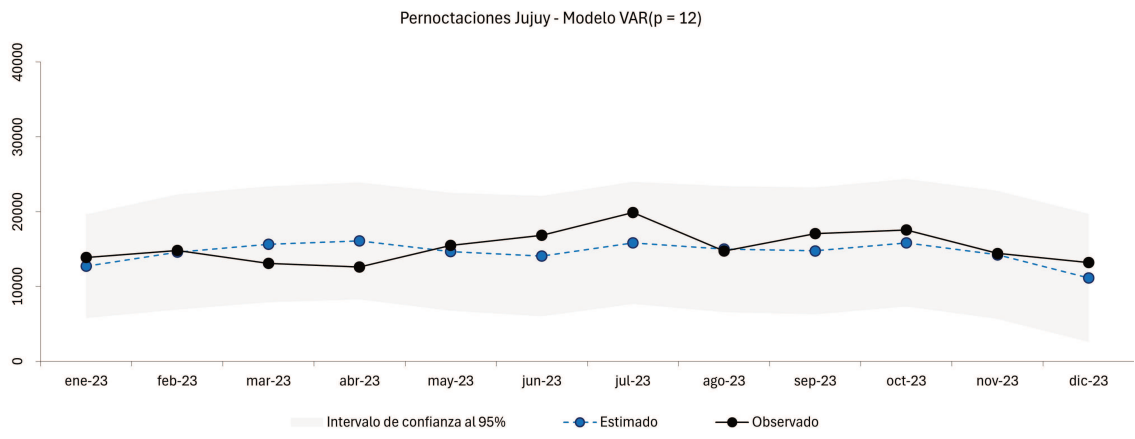
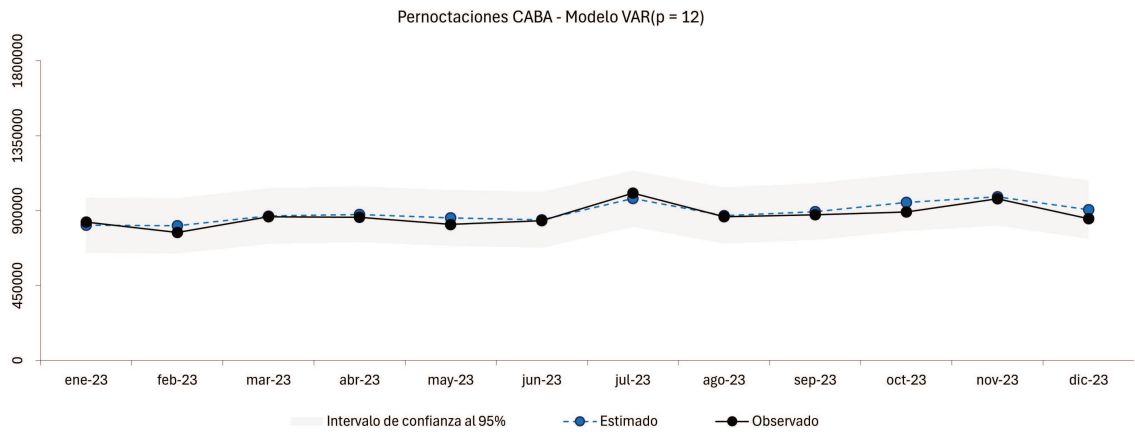
Las estimaciones dentro de la muestra replican de manera adecuada los verdaderos valores de las pernoctaciones en los destinos seleccionados, tanto para los casos en los que las mismas se estiman en niveles como en sus primeras diferencias —CABA, Jujuy y Santa Fe—.

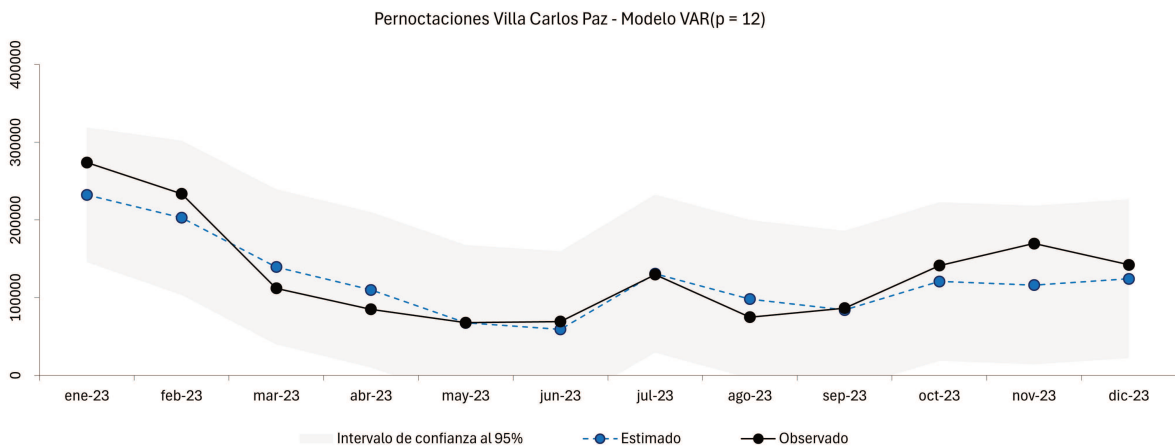
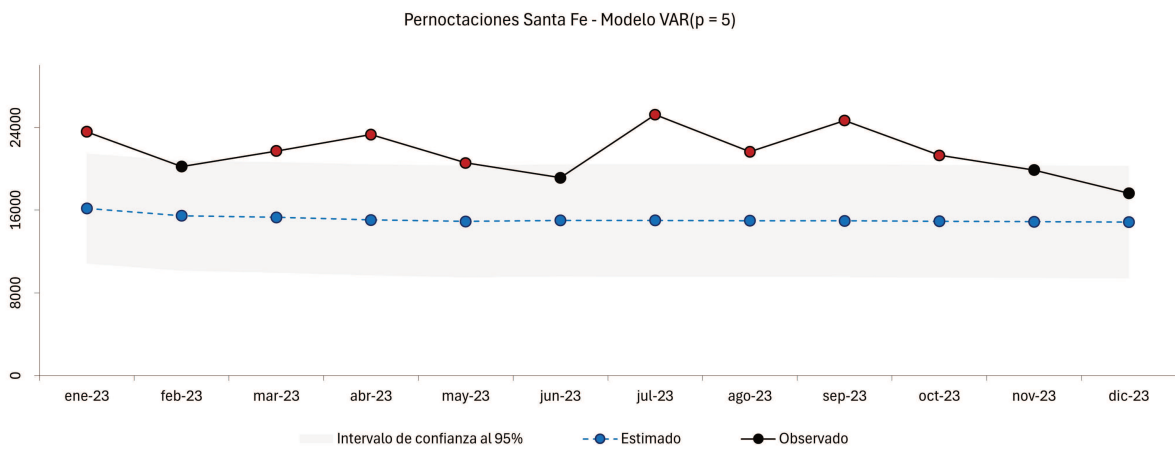
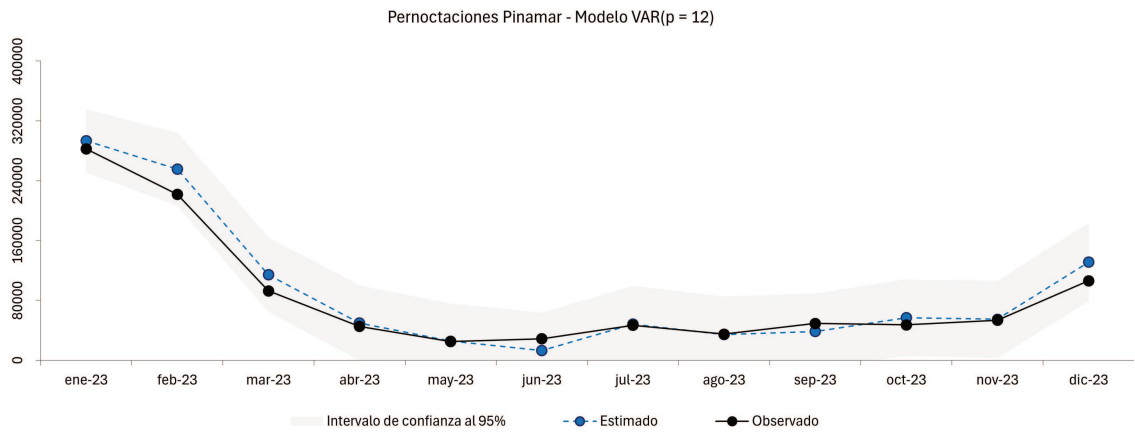
#### 4.3.1. Predicciones fuera de la muestra con modelos multivariados

De manera similar al caso del análisis con versiones univariadas, se predicen las pernoctaciones fuera de la muestra y se observa que el grado de precisión de las estimaciones a través de modelos VAR es heterogéneo, destacándose el caso de la Ciudad de Santa Fe, para la cual el modelo seleccionado (con 5 rezagos e incluyendo las diferencias en pernoctaciones y búsquedas de “Puente Colgante Santa Fe”) arroja predicciones cuyo intervalo de confianza al 95 % no incluye a los verdaderos valores de la serie en 8 de los 12 meses del año 2023 (ver Figura 7).

Figura 7: Valores observados, estimación fuera de la muestra e intervalos al 95 % de confianza obtenidos mediante modelos VAR







**Fuente:** elaboración propia en base a datos del Instituto Nacional de Estadística y Censos (INDEC) y Google.

Se puede destacar además que, para los siete destinos turísticos analizados, el error de estimación se mantiene a medida que nos alejamos del momento inicial de la predicción. Esto podría indicar que no resulta necesario un proceso recursivo y que los resultados obtenidos en términos de parámetros, poseen utilidad en las predicciones de los próximos doce meses sin la necesidad de mayores ajustes. El Cuadro 3 expone los mencionados resultados.

Cuadro 3: Error de estimación para cada mes de 2023 y promedio anual de los modelos VAR

	Bariloche VAR (p = 12)	CABA VAR (p = 12)	Jujuy VAR (p = 12)	Mendoza VAR (p = 12)	Pinamar VAR (p = 12)	Santa Fe VAR (p = 5)	VCP VAR (p = 12)
ene-23	1,2%	2,2%	8,3%	1,2%	3,8%	31,5%	15,2%
feb-23	9,3%	5,2%	1,5%	10,9%	15,3%	23,5%	13,2%
mar-23	18,3%	0,5%	19,4%	11,3%	23,2%	29,6%	24,4%
abr-23	2,3%	1,9%	27,9%	4,3%	10,2%	35,4%	29,5%
may-23	15,4%	4,9%	5,4%	18,6%	2,0%	27,5%	0,6%
jun-23	6,3%	0,6%	16,5%	16,2%	55,0%	21,6%	14,5%
jul-23	1,9%	3,2%	20,4%	13,4%	3,6%	40,5%	1,0%
ago-23	0,1%	0,9%	1,8%	5,9%	1,2%	30,8%	31,0%
sep-23	3,0%	2,2%	13,5%	9,9%	21,8%	39,3%	2,7%
oct-23	5,6%	6,4%	9,8%	7,8%	20,6%	30,0%	14,5%
nov-23	18,2%	1,2%	1,3%	0,3%	2,3%	25,1%	31,5%
dic-23	16,0%	6,4%	15,5%	10,6%	23,4%	15,8%	12,7%
Promedio 2023	8,1%	3,0%	11,8%	9,2%	15,2%	29,2%	15,9%

Fuente: elaboración propia en base a datos del Instituto Nacional de Estadística y Censos (INDEC) y Google.

#### 4.4. Comparando estrategias de predicción: ¿son las búsquedas de Google útiles para predecir la demanda turística en Argentina?

El análisis de los resultados de los modelos univariados y multivariados permite observar que, exceptuando el caso de la Ciudad de Santa Fe, los modelos VAR estiman de manera acertada las pernoctaciones en los diferentes destinos turísticos, conteniendo siempre dentro de sus intervalos de confianza al verdadero valor de la variable bajo estudio.

La comparación de los errores de estimación entre ambas alternativas se expone en el Cuadro 4, y permite apreciar que, en términos generales, las predicciones realizadas mediante modelos multivariados mejoran a las obtenidas a través de las versiones ARIMA. Para Bariloche, CABA, Mendoza y Villa Carlos Paz, los resultados obtenidos a través de modelos VAR dan lugar a disminuciones de hasta 16,3 puntos porcentuales en el error de estimación promedio para el año 2023 (como es el caso de Bariloche), respecto a los obtenidos con modelos univariados.

Cuadro 4: Comparación de errores de estimación entre modelos ARIMA y modelos VAR

	Bariloche		CABA		Jujuy		Mendoza		Pinamar		Santa Fe		VCP	
	ARIMA (1,0,0)	VAR (p = 12)	ARIMA (0,1,0)	VAR (p = 12)	ARIMA (1,0,2)	VAR (p = 12)	ARIMA (4,0,1)	VAR (p = 12)	ARIMA (2,0,3)	VAR (p = 12)	ARIMA (1,0,0)	VAR (p = 5)	ARIMA (1,0,0)	VAR (p = 12)
ene-23	4,0%	1,2%	1,0%	2,2%	23,8%	8,3%	11,1%	1,2%	3,6%	3,8%	34,0%	31,5%	0,5%	15,2%
feb-23	5,1%	9,3%	13,0%	5,2%	19,1%	1,5%	7,5%	10,9%	2,2%	15,3%	20,9%	23,5%	5,0%	13,2%
mar-23	1,3%	18,3%	8,0%	0,5%	11,9%	19,4%	4,5%	11,3%	3,8%	23,2%	22,7%	29,6%	4,6%	24,4%
abr-23	21,5%	2,3%	15,8%	1,9%	7,3%	27,9%	24,3%	4,3%	10,2%	10,2%	26,1%	35,4%	11,6%	29,5%
may-23	36,2%	15,4%	20,5%	4,9%	12,4%	5,4%	41,4%	18,6%	10,9%	2,0%	12,9%	27,5%	42,4%	0,6%
jun-23	45,9%	6,3%	15,7%	0,6%	19,3%	16,5%	47,8%	16,2%	27,0%	55,0%	8,3%	21,6%	42,7%	14,5%
jul-23	22,9%	1,9%	10,4%	3,2%	14,2%	20,4%	21,7%	13,4%	4,1%	3,6%	24,5%	40,5%	16,9%	1,0%
ago-23	26,5%	0,1%	18,2%	0,9%	7,3%	1,8%	22,3%	5,9%	8,5%	1,2%	16,7%	30,8%	18,6%	31,0%
sep-23	29,9%	3,0%	20,4%	2,2%	4,8%	13,5%	29,8%	9,9%	34,7%	21,8%	27,6%	39,3%	26,6%	2,7%
oct-23	34,1%	5,6%	22,9%	6,4%	0,0%	9,8%	27,9%	7,8%	8,9%	20,6%	15,2%	30,0%	31,0%	14,5%
nov-23	37,2%	18,2%	14,1%	1,2%	7,7%	1,3%	19,4%	0,3%	0,2%	2,3%	10,7%	25,1%	36,7%	31,5%
dic-23	29,0%	16,0%	22,2%	6,4%	11,0%	15,5%	13,6%	10,6%	4,6%	23,4%	5,3%	15,8%	25,6%	12,7%
Promedio 2023	24,5%	8,1%	15,2%	3,0%	11,6%	11,8%	22,6%	9,2%	9,9%	15,2%	18,7%	29,2%	21,8%	15,9%
Mejora modelo VAR	<b>16,3%</b>		<b>12,2%</b>		<b>-0,2%</b>		<b>13,4%</b>		<b>-5,3%</b>		<b>-10,5%</b>		<b>5,9%</b>	

Fuente: elaboración propia en base a datos del Instituto Nacional de Estadística y Censos (INDEC) y Google.

En términos agregados, la implementación de modelos VAR reduce el error de estimación en un promedio de 4,6 puntos porcentuales para los siete destinos turísticos durante los doce meses del año 2023. Por este motivo, se puede concluir que la incorporación de información proveniente de las búsquedas en Google por parte de los usuarios mejora, al menos en algunos casos, las estimaciones de las pernoctaciones de turistas en diferentes destinos de Argentina.

## Capítulo 5

# Discusión, Conclusiones y Recomendaciones

En esta tesis se ha realizado un análisis metodológico para predecir una variable económica de interés: la demanda turística en centros urbanos de Argentina.

Según datos del INDEC, el sector turístico, medido a través del rubro de hoteles y restaurantes, generó aproximadamente USD 13.332,2 millones durante el año 2023, lo que representa cerca del 2,1 % del PBI de ese año. Dada la importancia de este rubro, es crucial abordar uno de sus principales desafíos: la falta de información oportuna para la toma de decisiones. Aunque los datos oficiales sobre la evolución de las variables relacionadas con esta actividad —incluyendo oferta disponible, composición de la demanda, tendencias, eventos disruptivos y patrones estacionales— permiten conocer el desempeño de este sector, su publicación ocurre normalmente con un rezago de casi dos meses, lo que limita significativamente la capacidad de los agentes del sector para tomar decisiones. Por lo tanto, resulta de interés desarrollar mecanismos que proporcionen a los actores económicos información, en lo posible más actualizada, lo que facilitaría la implementación de medidas para mitigar las fluctuaciones de la demanda y mejorar la prestación de servicios turísticos.

Para generar información oportuna, existen diversas alternativas metodológicas. Uno de los enfoques más utilizados para la predicción de variables son los modelos univariados, debido a su simplicidad en el diseño y su facilidad de implementación, así como a la economía en la disponibilidad de datos. Sin embargo, con el avance de la tecnología, se ha generado un gran volumen de datos que crece exponencialmente y tiene una penetración significativa en la vida cotidiana de las personas. Este fenómeno presenta un notable potencial para ser considerado como una fuente valiosa de información, permitiendo conocer patrones de consumo y comportamiento, al mismo tiempo que facilita la recolección de datos, a costos cada vez menores.

Por ello, en esta tesis se propuso evaluar la utilidad de la información disponible relacionada con búsquedas, para aplicar modelos econométricos que la incorporen. Así, los modelos VAR integran información sobre las búsquedas en Google para predecir la evolución de las pernoctaciones en diversos destinos turísticos de Argentina, comparando su precisión con la de los modelos ARIMA que utilizan datos históricos como único insumo para las estimaciones. La finalidad de incorporar el mencionado método multivariado es, por un lado, obtener información anticipada sobre el sector mediante datos con rezagos prácticamente nulos que permitan a los agentes tomar decisiones de diferentes índoles —ya sean a corto o largo plazo— y, por otro lado, evaluar una fuente de información cuya actualización es casi inmediata y que, si resulta efectiva, potenciaría la capacidad de análisis comportamental.

Durante el desarrollo del mencionado proceso, uno de los desafíos más significativos que se ha identificado en la bibliografía —y que la presente tesis no resuelve— es la falta de una metodología objetiva para seleccionar las palabras clave que podrían ser de utilidad para explicar el comportamiento turístico. Esta carencia, destacada por Blanco (2014) y Naccarato et al. (2018), es especialmente relevante, ya que las palabras clave pueden variar considerablemente entre diferentes destinos debido a sus características particulares y a las preferencias cambiantes de los turistas.

En relación con este punto, Askitas y Zimmermann (2009) utilizan Google Insights<sup>1</sup> como fuente para la selección de palabras clave. Sin embargo, los autores también advierten sobre las limitaciones del servicio, como el método de muestreo utilizado y las variaciones diarias en los datos.

En investigaciones previas del autor, el mencionado proceso de selección se realizó mediante la utilización de la herramienta Google Correlate, la cual dejó de estar disponible a finales del año 2019<sup>2</sup>. Por no contarse dicho instrumento, en esta tesis se seleccionaron los términos mediante un criterio completamente subjetivo. Este procedimiento puede llevar a resultados sesgados o que no reflejen de la mejor manera la problemática real, limitando la capacidad de que los mismos se puedan extrapolar a otras aplicaciones de interés. Por este motivo, se considera necesario que futuras versiones de este documento incorporen una metodología superadora en lo que respecta la selección de palabras claves, utilizando técnicas tales como la reducción de dimensiones a través de análisis de componentes principales (PCA, por sus siglas en inglés) o análisis discriminante lineal (LDA, por sus siglas en inglés) de rango reducido, solo por mencionar algunos ejemplos.

Otro aspecto que merece atención es la implementación, propiamente dicha, de los modelos predictivos multivariados. En esta tesis se utilizaron modelos VAR para todas las variables, y en los casos de series no estacionarias, se aplicaron dichos modelos a las series en primeras diferencias. En futuros trabajos, sería recomendable evaluar la posibilidad de implementar vectores de corrección del error (modelos VECM), que pueden aplicarse directamente sin requerir la estacionariedad de las series bajo análisis. Esto podría facilitar un enfoque más adecuado para modelar variables no estacionarias y mejorar la precisión predictiva.

Un interrogante adicional que surge del proceso de implementación refiere a la conveniencia de modelar las variables por separado o intentar un enfoque más integrado, como un modelo pool. Al respecto, se puede destacar que modelar por separado puede permitir una mayor flexibilidad y adaptación a las características específicas de cada variable, pero también puede resultar en una pérdida de información valiosa sobre las interrelaciones entre ellas. Por otro lado, un modelo pool podría ofrecer una visión más global y a la vez simplificada, aunque se corre el riesgo de no capturar adecuadamente las dinámicas individuales.

Aun teniendo en cuenta las limitaciones y consideraciones mencionadas, la inclusión de información proveniente de Google resulta pertinente, y puede potencialmente contribuir a la obtención de resultados más precisos que los obtenidos mediante análisis univariado, ya que como mencionan Carrière-Swallow y Labbe (2011), estos datos se derivan del accionar de micro usuarios y tienen una buena representatividad, se publican con alta frecuencia y en intervalos regulares.

En el caso puntual de esta tesis, los resultados obtenidos indican que, para cuatro de los siete destinos analizados, las predicciones generadas por el modelo VAR que incorpora información sobre búsquedas, son más precisas que las producidas por el modelo ARIMA<sup>3</sup>. La eficiencia observada para los modelos que incluyen datos sobre búsquedas se encuentra en línea con lo planteado por Choi y Varian (2012), Askitas y Zimmermann (2009), Bhattacharjee y Kohns (2022) y el propio Blanco (2014), entre otros. No obstante, es

---

<sup>1</sup>Esta herramienta permite ver y comparar las frases más populares entre las consultas de búsqueda en una categoría temática determinada, una región del mundo o país, y en un periodo indicado.

<sup>2</sup>Ver Jorge (2020) y Camusso y Jorge (2021).

<sup>3</sup>Los casos en los que el modelo multivariado mejora los resultados obtenidos por el univariado son Bariloche, CABA, Mendoza y Villa Carlos Paz.

importante señalar que el modelo ARIMA se presenta en todos los casos bajo estudio como más parsimonioso, lo que plantea interrogantes sobre la conveniencia de implementar modelos multivariados debido a que esto incrementa la posibilidad de sobre especificación y es más costosa en términos de grados de libertad, dando lugar a modelos más complejos y difíciles de interpretar y pudiendo generar que el error estándar de los coeficientes estimados se incremente. Esta observación podría ir en contra de la hipótesis inicial, sugiriendo que, aunque la inclusión de datos de búsqueda puede mejorar la precisión en ciertos contextos, no siempre es la opción más eficiente.

Dentro de los autores que forman parte de la revisión bibliográfica realizada en este documento, se destacan las conclusiones de Choi y Varian (2012)<sup>4</sup>, quienes afirman que, aunque los modelos que incorporan variables de búsquedas en Google tienden a superar a aquellos que no lo hacen, la efectividad de estos modelos no es universal. En algunos casos, la mejora en las predicciones es modesta, mientras que en otros puede ser considerable —entre 12% y 18%—. Esto indica que, si bien Google Trends puede ser una herramienta valiosa en ciertas situaciones, su utilidad puede variar según el contexto y la naturaleza de los datos analizados.

Por lo mencionado, y en lo que refiere específicamente a los resultados obtenidos sobre la demanda turística en Argentina, se concluye que la inclusión de información proveniente de Google podría resultar de gran relevancia y contribuye a reducir significativamente el error de estimación en determinados destinos turísticos (como es el caso de Bariloche, CABA, Mendoza y Villa Carlos Paz), mientras que para otros destinos, la incorporación de estos datos no redundará en mejores resultados sino en una mayor complejización del modelo predictivo. Para el primer grupo de destinos, la inclusión de información sobre las búsquedas en Google, al mejorar la precisión de las predicciones, permite anticipar cambios en la demanda (disponer de información casi en tiempo real permite a los agentes económicos anticipar cambios en el comportamiento del consumidor), mejorar la toma de decisiones (con datos actualizados y relevantes, los responsables de la toma de decisiones pueden actuar con mayor rapidez y eficacia), optimizar los recursos (al predecir con mayor precisión la demanda turística, las empresas pueden optimizar sus recursos, desde la gestión del personal hasta el inventario y las reservas) e identificar oportunidades (la utilización de datos de búsqueda permite identificar patrones y oportunidades que podrían no ser evidentes con otro tipo de técnicas).

Después del análisis realizado se pueden mencionar algunas recomendaciones que deberían tenerse en cuenta en futuras investigaciones en aras de la construcción de consensos. En primer lugar, resulta fundamental establecer un proceso más riguroso y objetivo para la selección de palabras clave. Esto podría incluir el uso de herramientas analíticas que identifiquen términos relevantes basados en tendencias actuales del mercado y el comportamiento del consumidor, o como se mencionó precedentemente, mediante la utilización de técnicas de reducción de dimensiones. Además, se recomienda considerar la implementación de modelos econométricos alternativos —como podrían ser los mencionados modelos VECM— que puedan adaptarse mejor a las características específicas de cada destino turístico y su dinámica particular.

Adicionalmente, sería pertinente considerar la inclusión de variables económicas que podrían ser relevantes para predecir la evolución de la demanda turística (precios de traslado y alojamiento, salario real promedio, tipo de cambio, políticas públicas de incentivo a esta actividad, entre otros). Al comparar los resultados obtenidos mediante los modelos presentados en este documento —ARIMA y VAR utilizando información de Google— con aquellos que incorporan estas variables económicas, se podría evaluar la importancia de la información relacionada con las búsquedas en comparación con las demás variables, o incluso combinar ambas fuentes de datos.

Por último, es relevante explorar la posibilidad de adquirir bases de datos que contengan información

---

<sup>4</sup>En su investigación, analizan la afluencia de turistas a diferentes destinos, las ventas minoristas, de automóviles y de inmuebles.

complementaria. Google, por ejemplo, ofrece acceso a ciertos conjuntos de datos pagos, aunque es importante verificar las condiciones y costos asociados a su uso. La integración de estas bases de datos podría enriquecer aún más los análisis y mejorar la capacidad predictiva de los modelos.

En resumen, y aun haciendo las salvedades metodológicas pertinentes, se considera que esta investigación ha aportado evidencia sobre la efectividad del uso de datos provenientes de búsquedas en Google para mejorar las predicciones en el sector turístico, ya que su utilización aún se encuentra en etapas iniciales del proceso analítico. Sin embargo, también ha puesto de manifiesto la necesidad de abordar ciertos desafíos metodológicos que podrían optimizar y enriquecer aún más este enfoque en futuras investigaciones.

# Bibliografía

- AKERLOF, G. A. (1970). The market for “lemons”: quality uncertainty and the market mechanism. *Market Failure or Success*, 66.
- ARTOLA, C. & GALÁN, E. (2012). Tracking the future on the web: construction of leading indicators using Internet searches. *Banco de Espana occasional paper*, (1203).
- ASKITAS, N. & ZIMMERMANN, K. F. (2009). Google econometrics and unemployment forecasting. *Applied Economics Quarterly*, 55(2), 107-120.
- AUSTIN, M. P. A., AUSTIN, P., MARINI, M. M., SANCHEZ, A., SIMPSON-BELL, C. & TEBRAKE, J. (2021). Using the Google Places API and Google Trends data to develop high frequency indicators of economic activity. *International Monetary Fund*.
- BANTIS, E., CLEMENTS, M. P. & URQUHART, A. (2022). Forecasting GDP growth rates in the United States and Brazil using Google Trends. *International Journal of Forecasting*, 39(4), 1909-1924.
- BARTLETT, M. S. (1946). On the theoretical specification and sampling properties of autocorrelated time-series. *Supplement to the Journal of the Royal Statistical Society*, 8(1), 27-41.
- BECKER, G. S. (1976). *The Economic Approach to Human Behavior*. The University of Chicago Press.
- BHATTACHARJEE, A. & KOHNS, D. (2022). *Nowcasting Growth using Google Trends Data: A Bayesian Structural Time Series Model*. (No. 538). National Institute of Economic and Social Research.
- BLANCO, E. (2014). *Herramientas de Big Data: ¿Podemos aprovechar Google Trends para pronosticar algunas variables macro relevantes?*. Asociación Argentina de Economía Política.
- BOEGERSHAUSEN, J., BORAH, A. & STEPHEN, A. (2020). Fields of gold: Web scraping for consumer research. *Marketing Science Institute Working Paper Series*.
- BOLLEN, J., MAO, H. & ZENG, X. (2011). Twitter mood predicts the stock market. *Journal of computational science*, 2(1), 1-8.
- BORTOLI, C. & COMBES, S. (2015). *Contribution from Google Trends for forecasting the short-term economic outlook in France: limited avenues*.
- BOX, G. E. & PIERCE, D. A. (1970). *Distribution of residual autocorrelations in autoregressive-integrated moving average time series models*. *Journal of the American statistical Association*, 65(332), 1509-1526.

- CAMUSSO, F. & JORGE, R. (2021). Google Correlate y Google Trends como herramientas para realizar un nowcast de las ventas minoristas. *Ensayos Económicos*, (76), 26-45.
- CARRIÈRE-SWALLOW, Y. & LABBÉ, F. (2011). *Nowcasting with Google Trends in an emerging market*. Journal of Forecasting. Santiago, Chile.
- CHATFIELD, C. & XING, H. (2019). *The analysis of time series: an introduction with R*. Chapman and hall/CRC.
- CHOI, H. & VARIAN, H. R. (2012). *Predicting the present with Google Trends*. Economic record, 88, 2-9.
- CORREA, A. (2021). *Prediciendo la llegada de turistas a Colombia a partir de los criterios de Google Trends*. Lecturas de Economía, (95), 105-134.
- DERGIADES, T., MAVRAGANI, E. & PAN, B. (2018). *Google Trends and tourist´s arrivals: emerging biases and proposed corrections*. Tourism Management, 66, 108-120.
- DICKEY, D. A. & FULLER, W. A. (1979). *Distribution of the estimators for autoregressive time series with a unit root*. Journal of the American statistical association, 74(366a), 427-431.
- GUJARATI, D. & PORTER, D. (2010). *Econometría (quinta edición)*. México: Editorial Mc. Graw Hill.
- HOLMSTRÖM, B. (1979). *Moral hazard and observability*. The Bell journal of economics, 74-91.
- INSTITUTO NACIONAL DE ESTADÍSTICA Y CENSOS (INDEC). *Accesos a Internet*. <https://www.indec.gov.ar/indec/web/Nivel4-Tema-4-26-153>.
- INSTITUTO NACIONAL DE ESTADÍSTICA Y CENSOS (INDEC). *Encuesta de Ocupación Hotelera (EOH)*. <https://www.indec.gov.ar/indec/web/Nivel4-Tema-3-13-56>.
- JORGE, R. E. (2020). Herramientas de Google para la predicción de variables económicas. Una aplicación al Índice Compuesto Coincidente de Actividad Económica de la Provincia de Santa Fe (ICASFe). In *LV Reunión Anual de la Asociación Argentina de Economía Política (modalidad virtual, 18 al 20 de noviembre de 2020)*.
- JUN, S. P., YOO, H. S. & CHOI, S. (2018). *Ten years of research change using Google Trends: From the perspective of big data utilizations and applications*. Technological forecasting and social change, 130, 69-87.
- MANKIW, N. G. (2012). *Principles of economics*. Cengage Learning.
- ORGANIZACIÓN MUNDIAL DEL TURISMO (2019). *Definiciones de turismo de la OMT*. OMT, Madrid, DOI: <https://doi.org/10.18111/9789284420858>.
- NACCARATO, A., FALORSI, S., LORIGA, S. & PIERINI, A. (2018). *Combining official and Google Trends data to forecast the Italian youth unemployment rate*. Technological Forecasting and Social Change, 130, 114-122.
- PARK, S., LEE, J. & SONG, W. (2017). *Short-term forecasting of Japanese tourist inflow to South Korea*

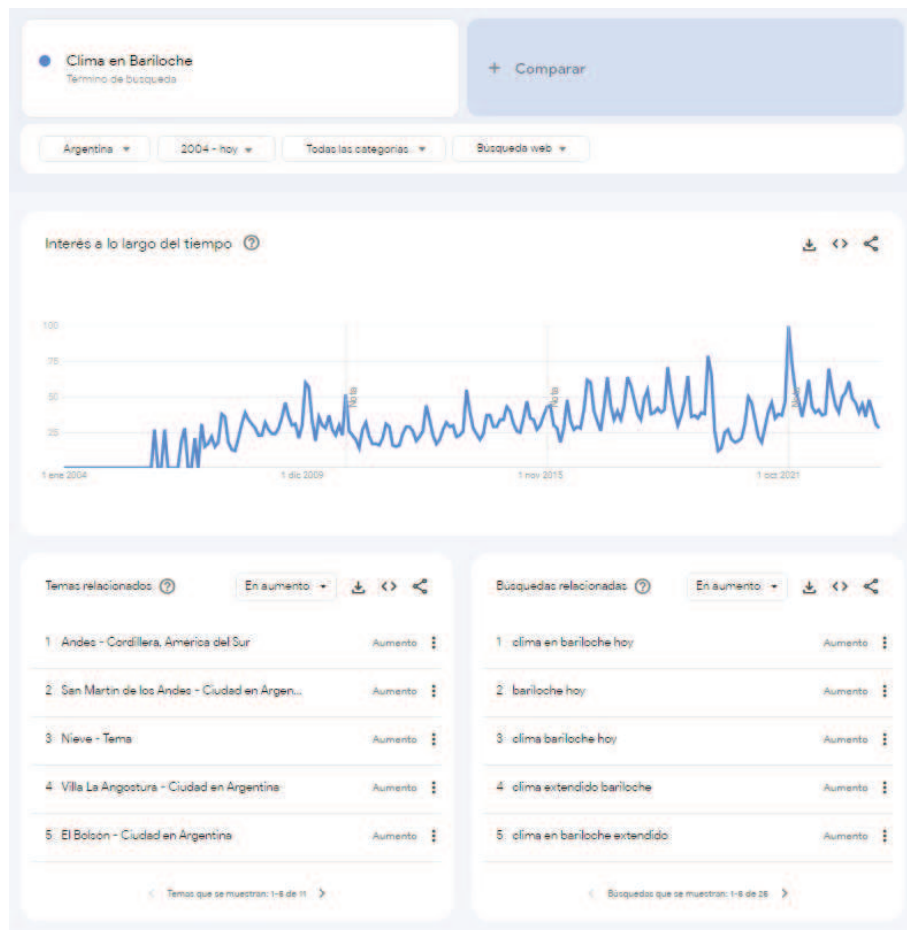
- using Google trends data*. *Journal of Travel Tourism Marketing*, 34(3), 357-368.
- PINDYCK, R. S. & RUBINFELD, D. L. (2018). *Microeconomics (Ninth Edit)*. England: Pearson Education.
- SIGALA, M. (2020). Tourism and COVID-19: Impacts and implications for advancing and resetting industry and research. *Journal of business research*, 117, 312-321.
- SNYDER, C., STEWART, R. & NICHOLSON, W. (2015). *Microeconomic theory: Basic principles and extensions (Europe, Middle East & Africa ed.)*. Andover: Cengage Learning.
- STATISTA GMBH. Cuota de mercado de los principales motores de búsqueda online a nivel mundial de 2021 a 2023.
- STIGLITZ, J. E. (2000). Economics of the public sector. *London-New York: 3th Edition WW Norton*.
- INTERNATIONAL TELECOMMUNICATION UNION (ITU). Individuals using the Internet (from any location), by gender and urban/rural location. <https://www.itu.int/en/ITU-D/Statistics/Pages/stat/default.aspx>.
- VARIAN, H. R. (2010). *Intermediate microeconomics: a modern approach*. New York: W. W.
- VON NEUMANN, J. & MORGENSTERN, O. (1944). Theory of games and economic behavior, Princeton, 1944. *On Decision-making under uncertainty*, 285.
- WOOLDRIDGE, J. M. (2009). *Introductory Econometrics: A Modern Approach*, South Western. *Cengage Learning*.
- XIANG, Z. & FESENMAIER, D. R. (2017). *Analytics in smart tourism design: concepts and methods*. Springer International Publishing Switzerland.

## Anexo A

### Interface y aspectos generales de Google Trends

Google Trends brinda la posibilidad de obtener información de las búsquedas con diferentes periodicidades, i.e. anual, mensual, semanal, e incluso por hora. El mecanismo de funcionamiento resulta intuitivo y de uso sencillo para el usuario. Este ingresa una palabra clave en el buscador, teniendo la posibilidad de aplicar filtros de búsqueda, entre los cuales se puede mencionar el área geográfica (cuya dimensión espacial mínima es la provincia), el período de tiempo, la categoría (arte, ciencias, compras, deportes, etc.) y el tipo de búsqueda (en la web, en noticias, imágenes). En la Figura A1 se expone, a modo de ejemplo, la pantalla de resultados que se obtiene al introducir “Clima en Bariloche” en Google Trends. Como se puede observar en la figura, de manera adicional al gráfico y los datos sobre la evolución de las búsquedas, esta herramienta expone un cuadro de “consultas relacionadas” que lista los términos que también fueron buscados por usuarios que ingresaron la palabra en cuestión.

Figura A1: Interface de la página de Google Trends



Fuente: Google Trends.

## Pruebas de raíces unitarias para modelo univariado

Cuadro A1: Resultados de pruebas de raíces unitarias de las pernoctaciones totales

Serie	Bariloche			CABA			Jujuy			Mendoza			Pinamar			Santa Fe			VCP		
	1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%
Prueba sin intercepto ni tendencia	E	E	E	NE	NE	NE	NE	NE	NE	NE	NE	NE	E	E	E	NE	NE	NE	E	E	E
Prueba con intercepto	E	E	E	NE	NE	E	E	E	E	E	E	E	E	E	E	NE	E	E	E	E	E
Prueba con tendencia	E	E	E	NE	NE	NE	E	E	E	E	E	E	E	E	E	NE	NE	E	E	E	E
Prueba ADF sobre series originales	E	E	E	NE	NE	NE	NE	NE	NE	NE	NE	NE	E	E	E	NE	NE	NE	E	E	E
Prueba ADF sobre series con diferencia	E	E	E	E	E	E	E	E	E	E	E	E	E	E	E	E	E	E	E	E	E

E Estacionaria      NE No Estacionaria

## Pruebas de raíces unitarias para modelo multivariado

Cuadro A2: Resultados de las pruebas de raíces unitarias ADF sobre las búsquedas

Serie	Bariloche			CABA			Jujuy			Mendoza			Pinamar			Santa Fe			VCP		
	1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%	1%	5%	10%
Clima	NE	E	E	E	E	E	NE	NE	NE	E	E	E	E	E	E	E	E	E	E	E	E
Hoteles	NE	NE	E	E	E	E	E	E	E	E	E	E	E	E	E	NE	NE	NE	E	E	E
Pasajes	E	E	E	E	E	E	E	E	E	E	E	E	E	E	E	E	E	E	E	E	E
Esquiar en Bariloche	E	E	E																		
Lugares en CABA				NE	NE	NE															
Cerro de los Siete Colores en Jujuy							NE	NE	NE												
Puente del Inca de Mendoza										E	E	E									
Playas en Pinamar													E	E	E						
Puente Colgante de Santa Fe																E	E	E			
Tren de las Sierras de Villa Carlos Paz																			NE	NE	E

E Estacionaria      NE No Estacionaria

## Selección de parámetros para modelo univariado

Cuadro A3: Selección de parámetros  $p$ ,  $d$  y  $q$  para modelos univariados

Ubicación	Modelo	AIC	BIC	RMSE
Bariloche	ARIMA(1,0,0)	5.196,18	5.206,31	37.604,20
Ciudad Autónoma de Buenos Aires	ARIMA(0,1,0)	5.797,69	5.804,54	83.076,18
Ciudad de Jujuy	ARIMA(1,0,2)	3.643,33	3.659,62	2.934,15
Ciudad de Mendoza	ARIMA(4,0,1)	4.779,55	4.803,18	13.895,21
Pinamar	ARIMA(2,0,3)	4.772,10	4.799,10	13.766,31
Ciudad de Santa Fe	ARIMA(1,0,0)	3.563,35	3.579,64	2.509,96
Villa Carlos Paz	ARIMA(1,0,0)	5.061,96	5.078,83	27.499,59

## Anexo B

### Residuos de los modelos univariados seleccionados

Figura B1: Residuos de modelo univariado sobre las pernoctaciones en Bariloche

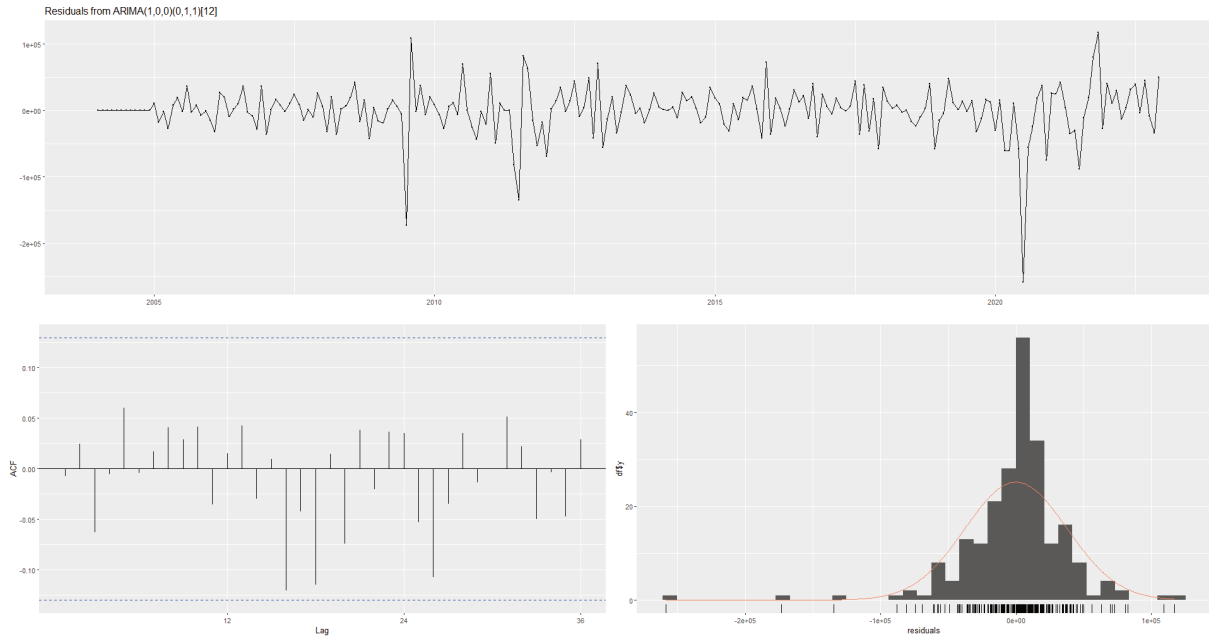


Figura B2: Residuos de modelo univariado sobre las pernoctaciones en CABA

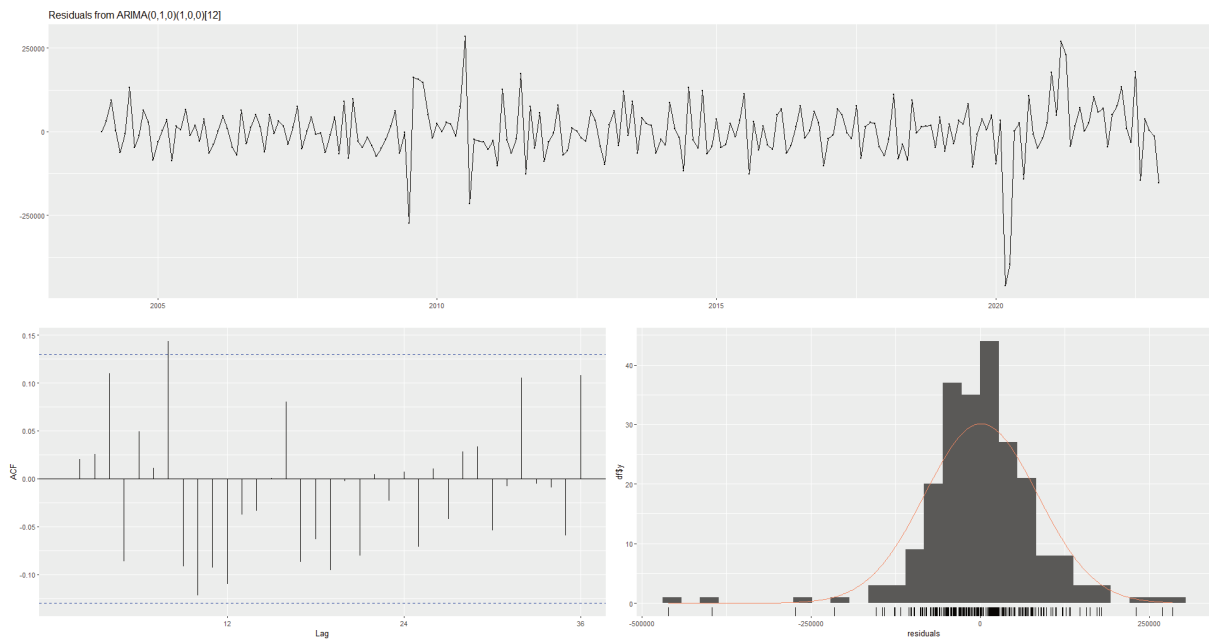


Figura B3: Residuos de modelo univariado sobre las pernoctaciones en Ciudad de Jujuy

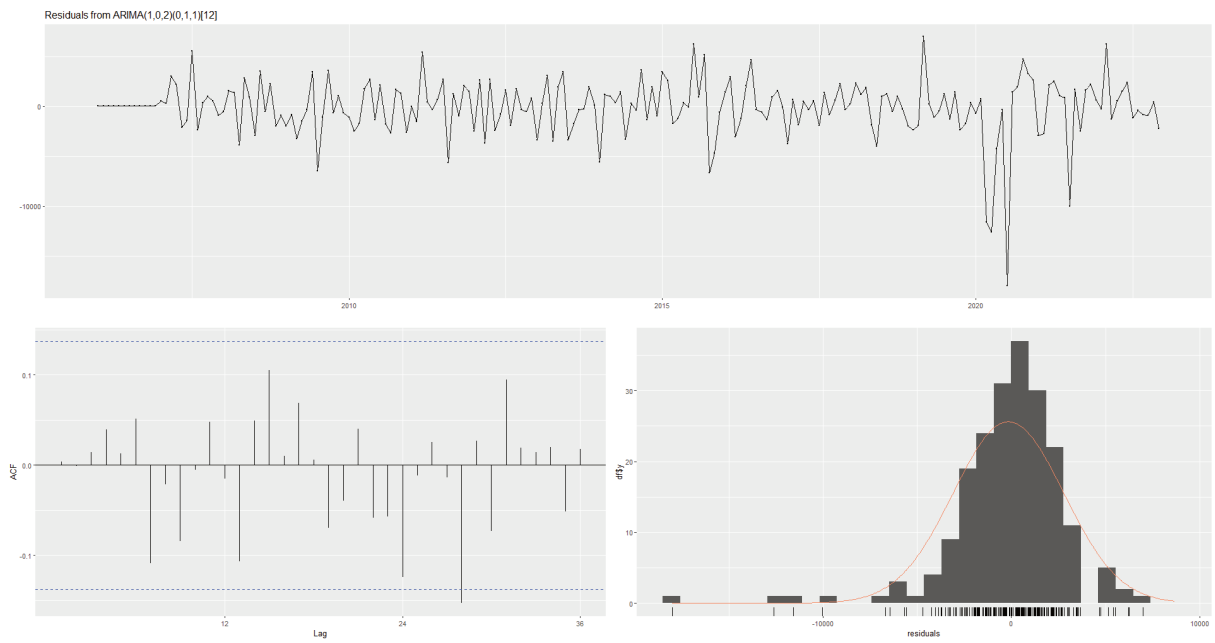


Figura B4: Residuos de modelo univariado sobre las pernoctaciones en Ciudad de Mendoza

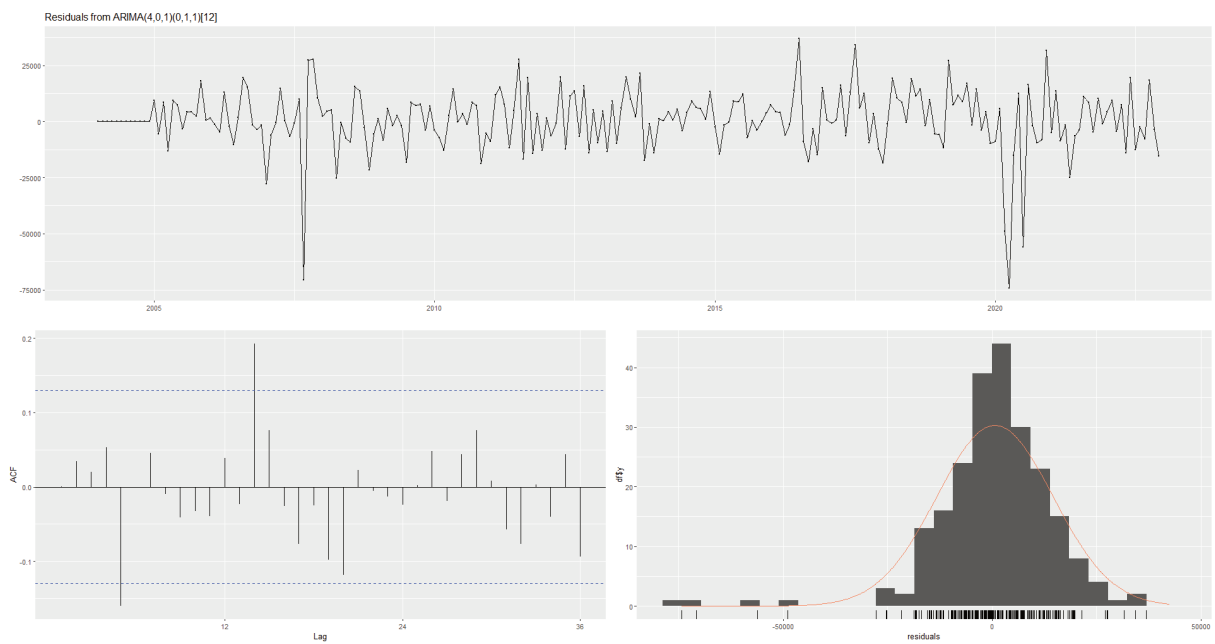


Figura B5: Residuos de modelo univariado sobre las pernoctaciones en Pinamar

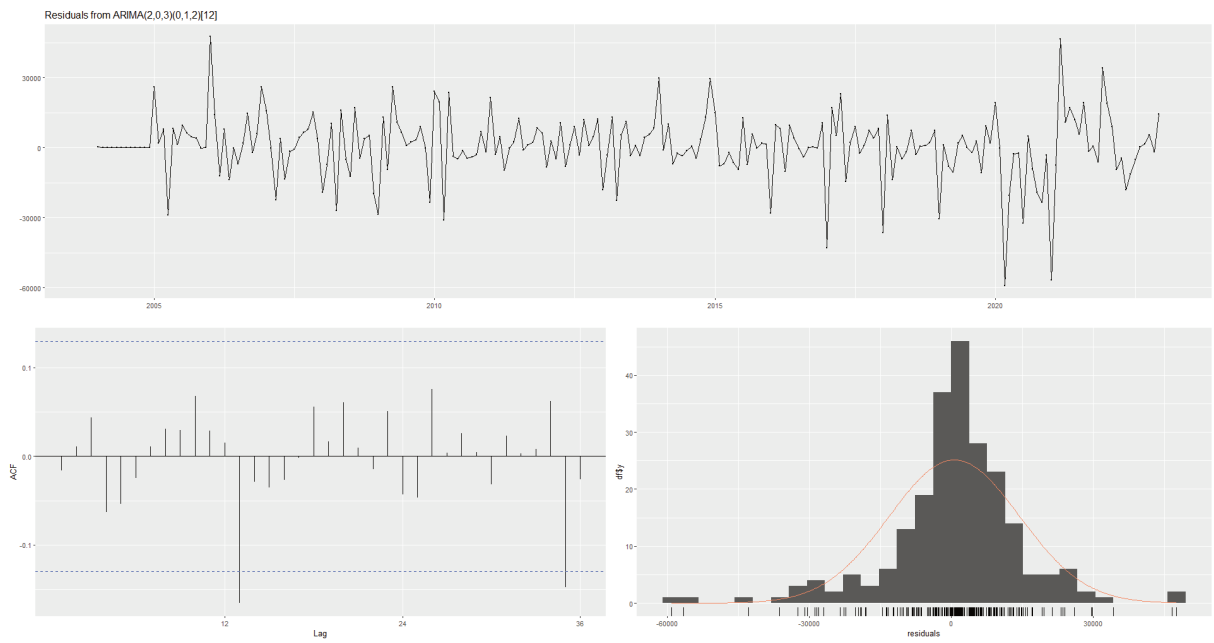


Figura B6: Residuos de modelo univariado sobre las pernoctaciones en Ciudad de Santa Fe

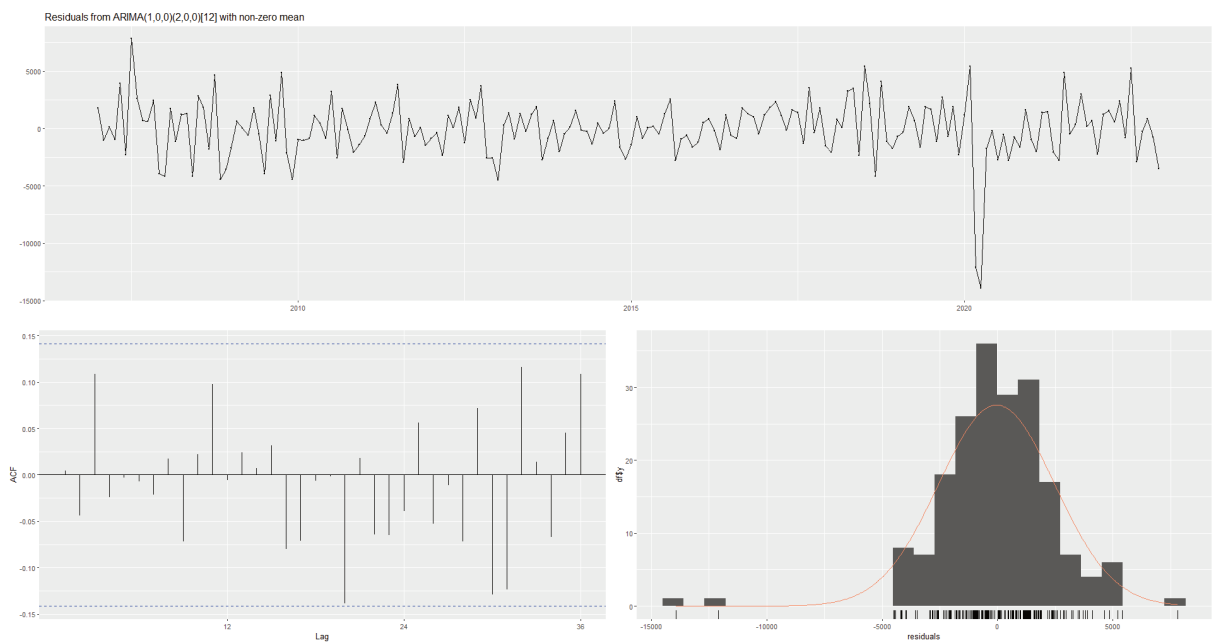
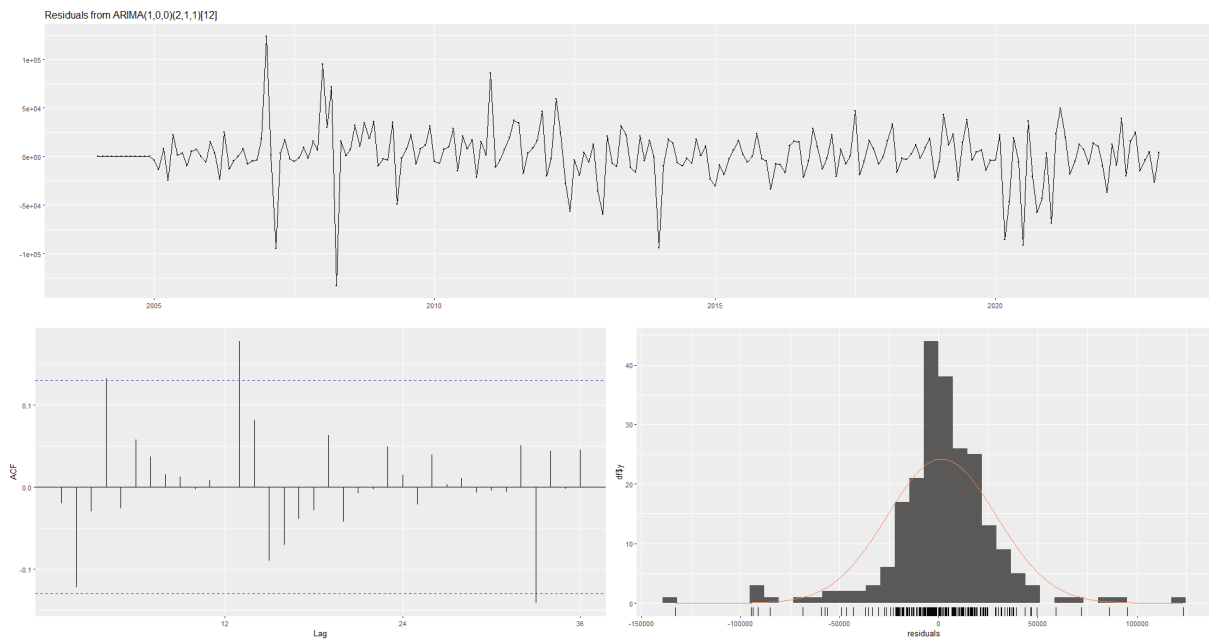


Figura B7: Residuos de modelo univariado sobre las pernoctaciones en Villa Carlos Paz



El análisis de los criterios de especificación de los modelos seleccionados permite afirmar que los residuos de los mismos no parecen presentar estructuras asimétricas<sup>5</sup> ni autocorrelación temporal significativa.

---

<sup>5</sup>Aun cuando los resultados de las pruebas de normalidad de Shapiro-Wilk y Jarque-Bera evidencian existencia de no normalidad, los residuos no parecen tener una estructura marcadamente asimétrica, pudiendo atribuirse los resultados a la existencia de valores atípicos en 2009 (posiblemente por el brote de Gripe A registrado en el invierno de ese año) y 2020 (a causa de la pandemia de COVID-19).

## Residuos de los modelos multivariados seleccionados

Figura B8: Residuos de modelo multivariado sobre las pernотaciones en Bariloche

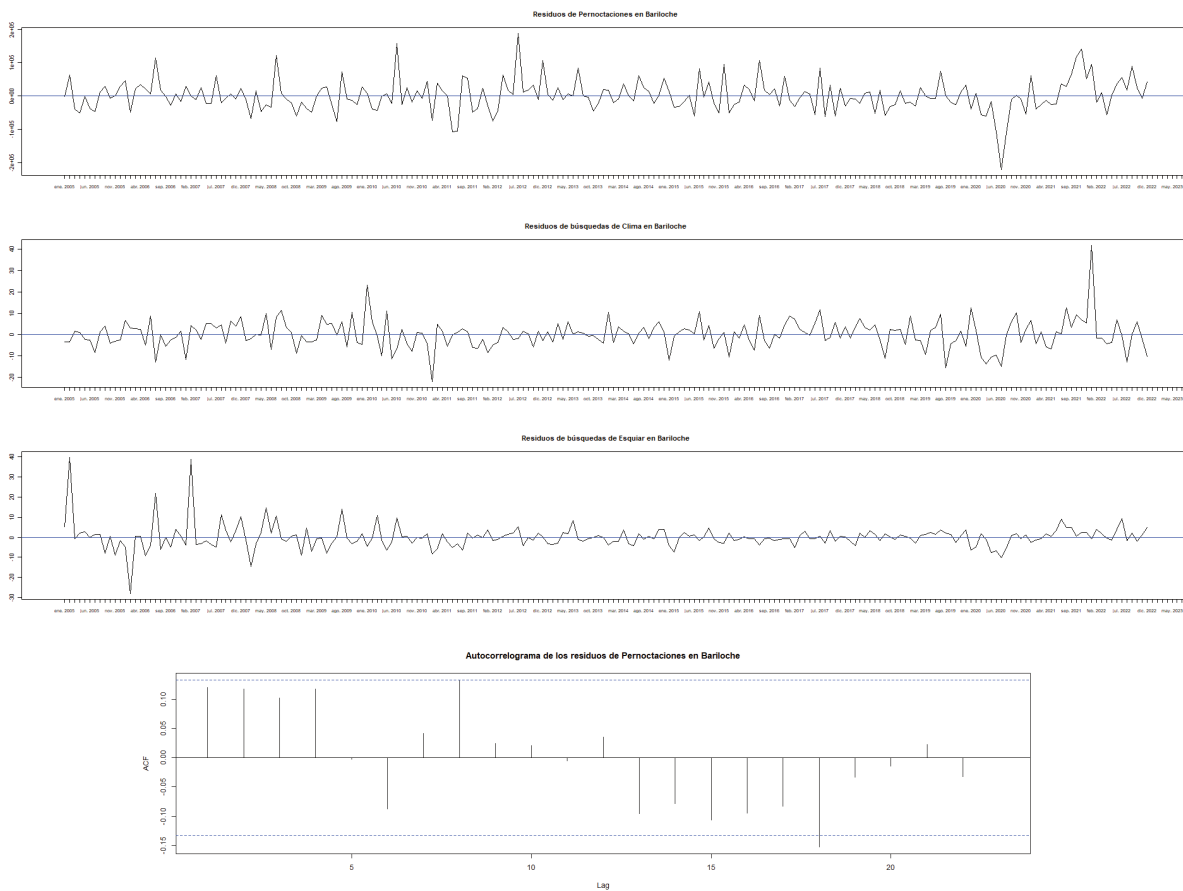


Figura B9: Residuos de modelo multivariado sobre las pernoctaciones en CABA

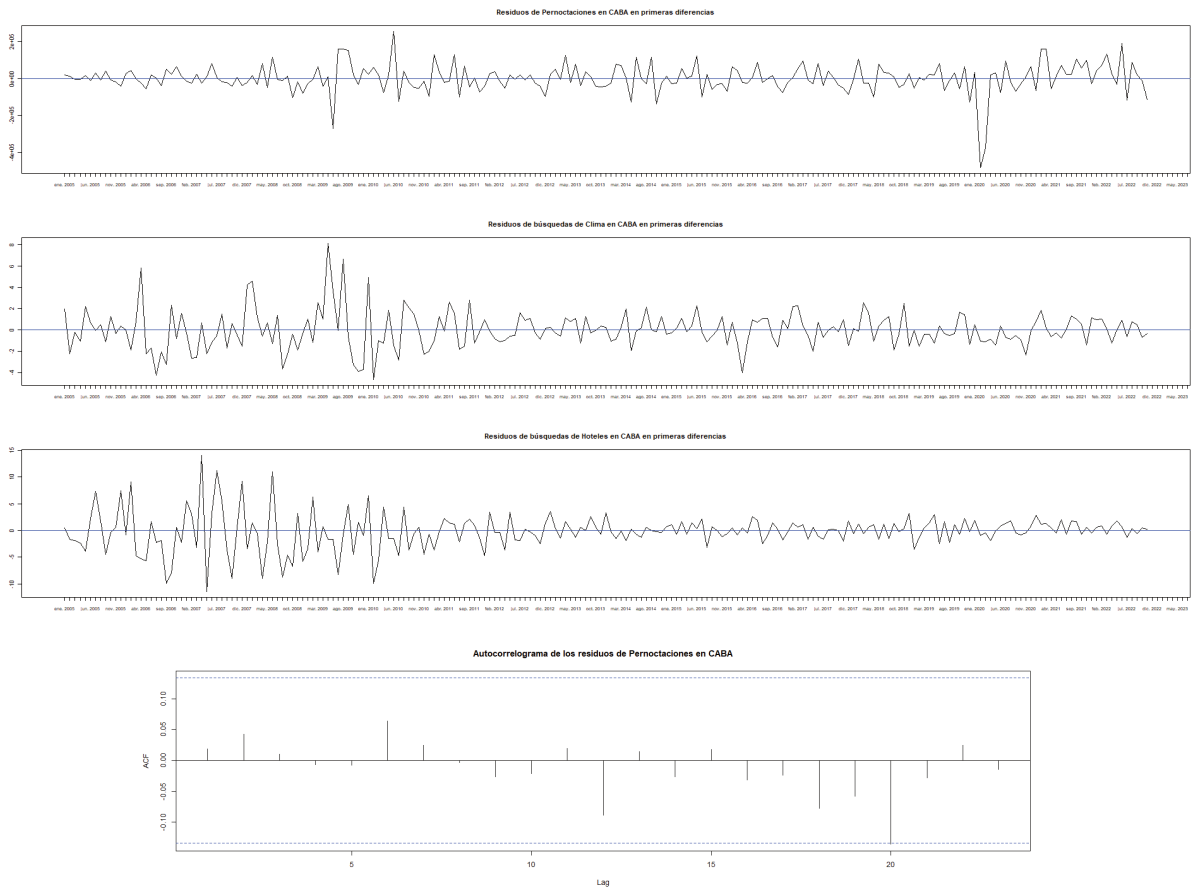


Figura B10: Residuos de modelo multivariado sobre las pernотaciones en Jujuy

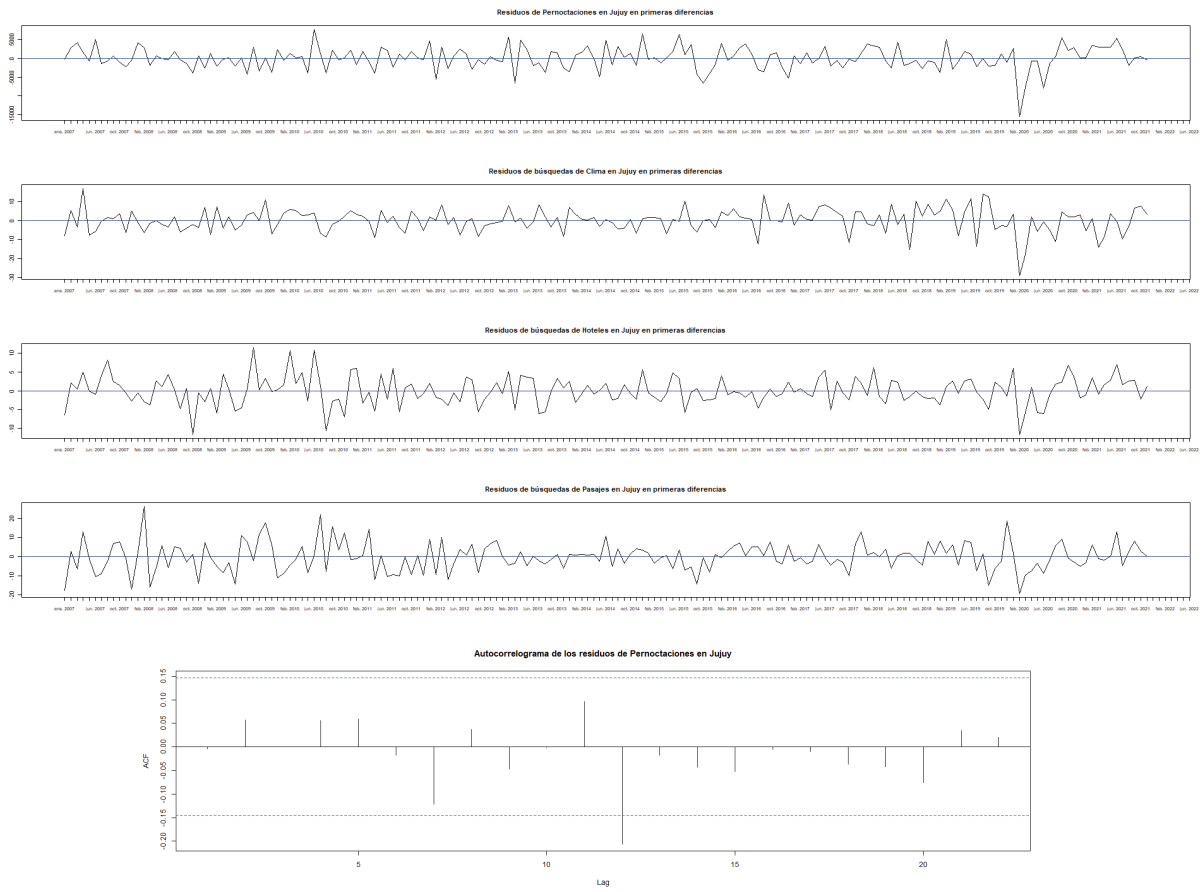


Figura B11: Residuos de modelo multivariado sobre las pernотaciones en Mendoza

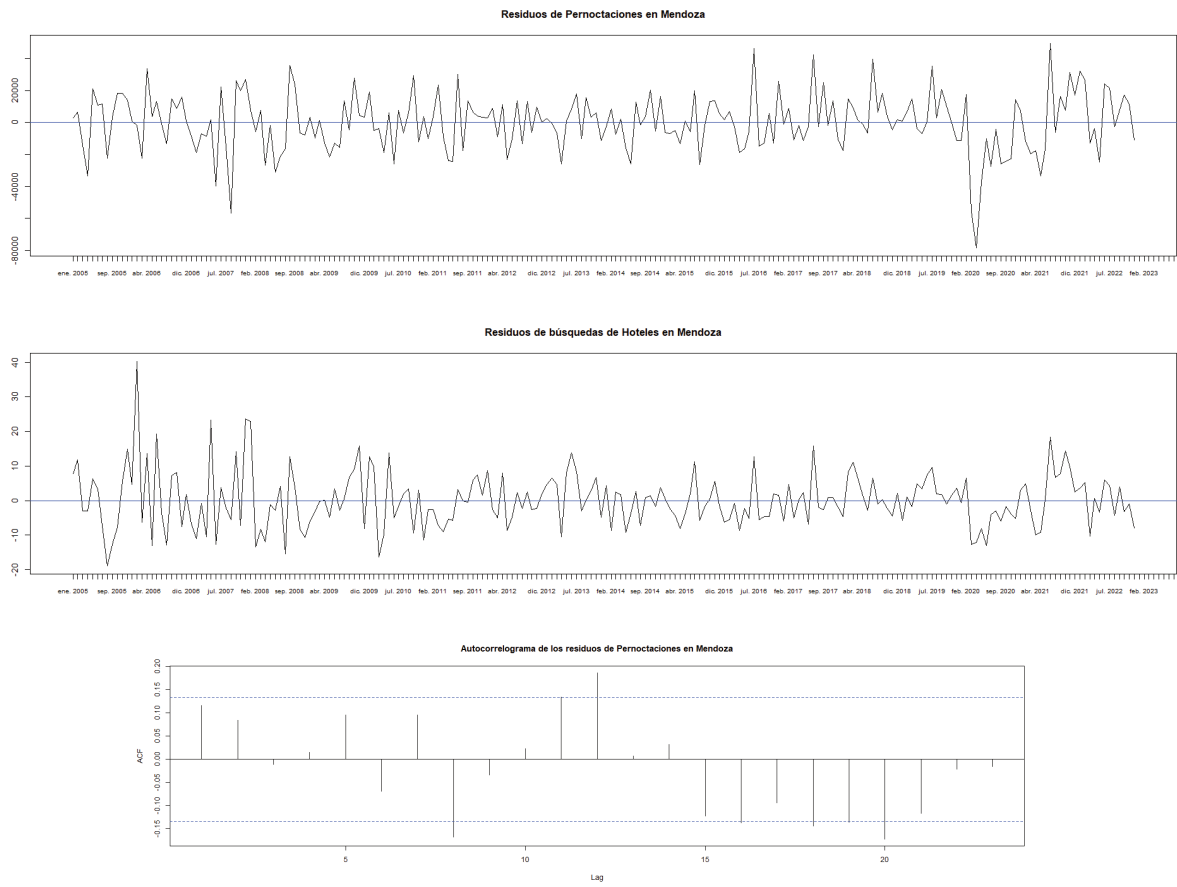


Figura B12: Residuos de modelo multivariado sobre las pernoctaciones en Pinamar

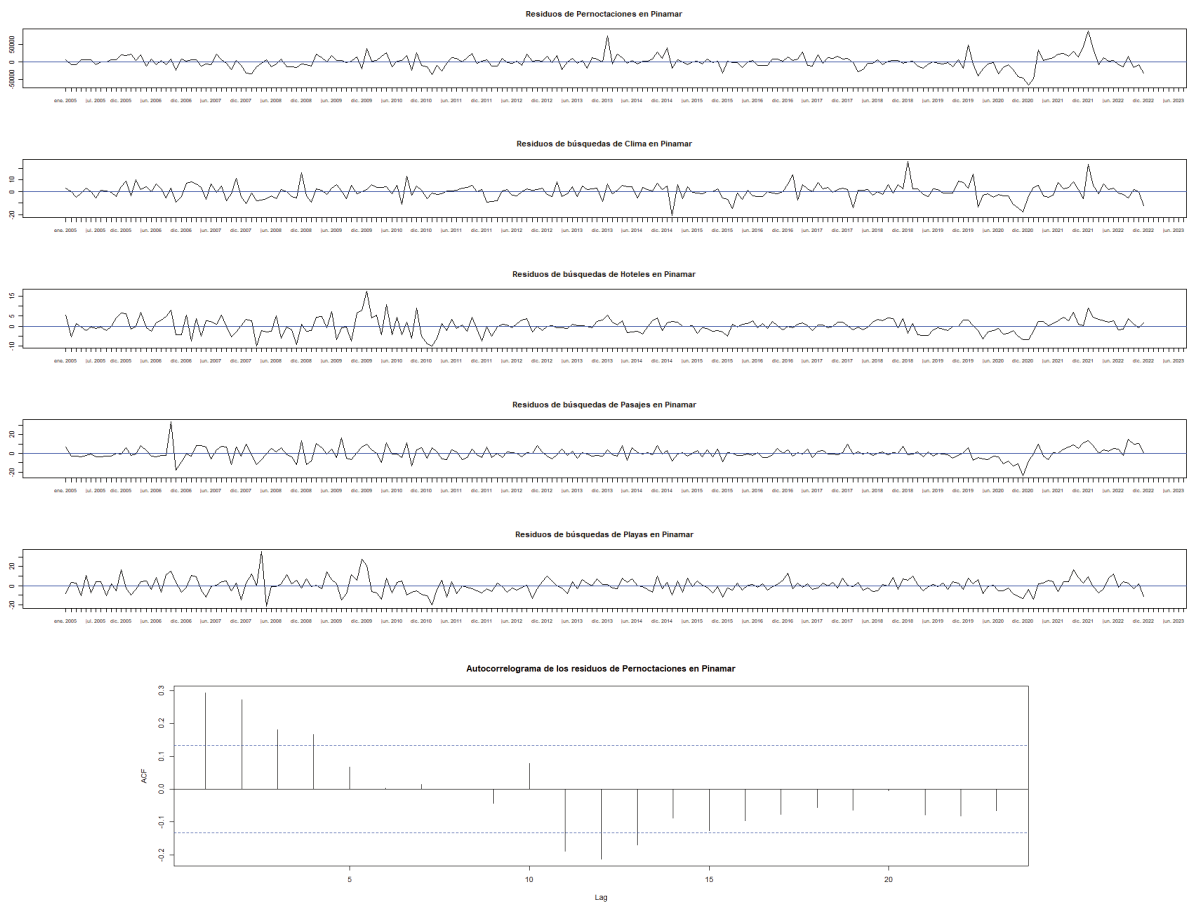


Figura B13: Residuos de modelo multivariado sobre las pernотaciones en Santa Fe

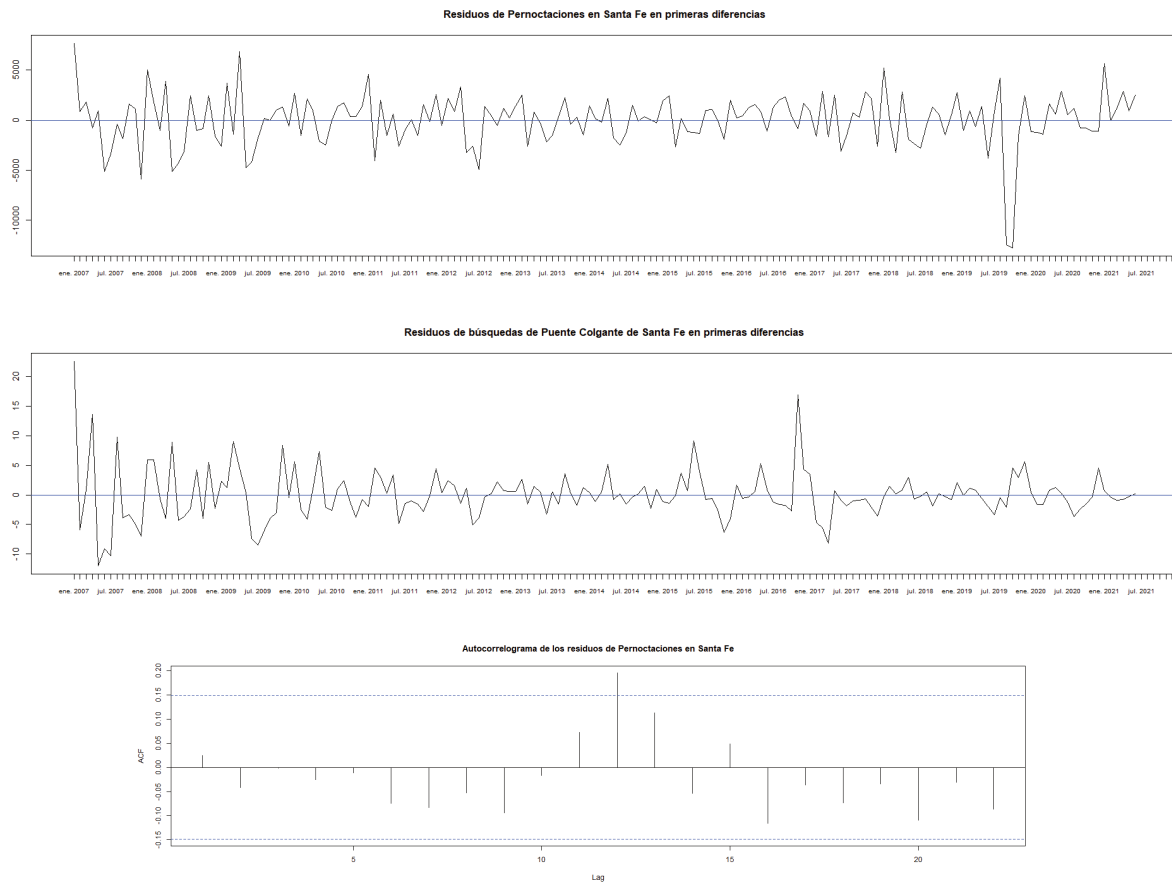
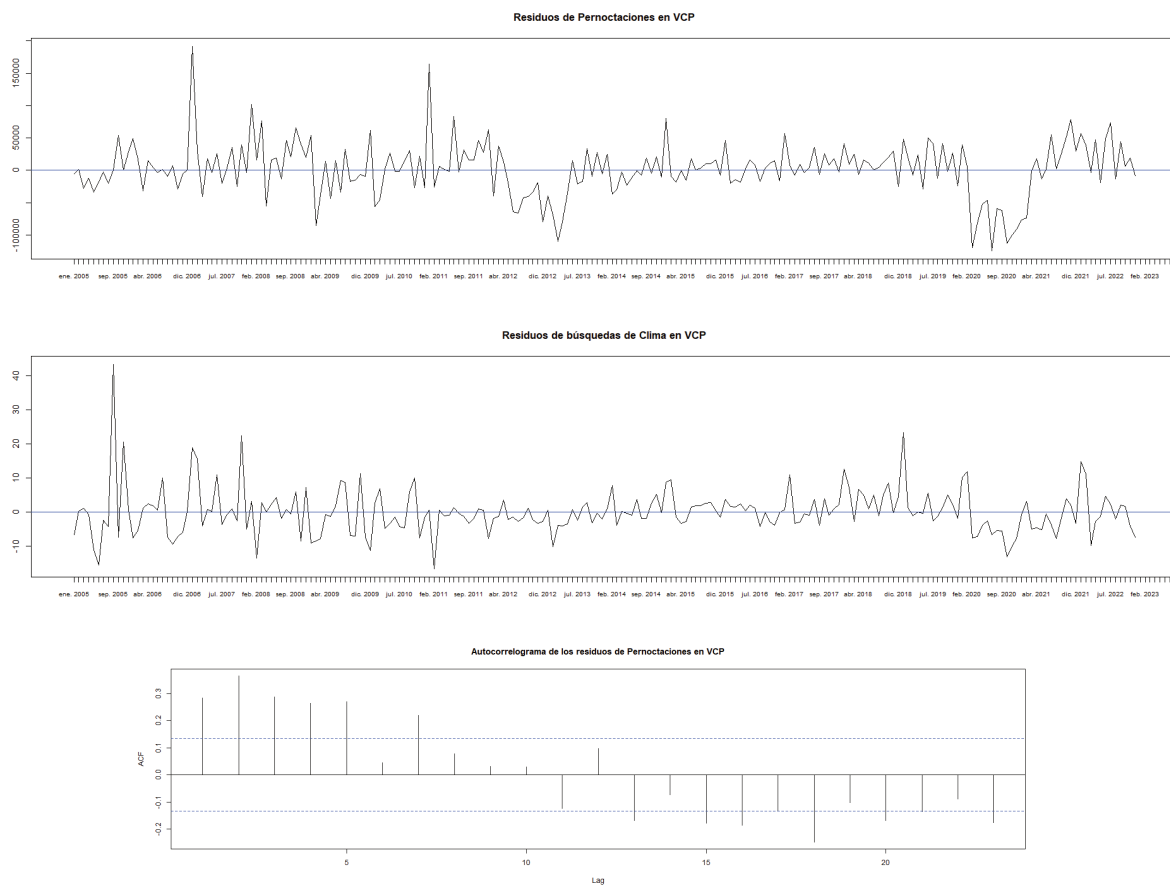


Figura B14: Residuos de modelo multivariado sobre las pernотaciones en Villa Carlos Paz



Los residuos de los modelos VAR no parecen presentar estructuras asimétricas<sup>6</sup>, aunque sí autocorrelación temporal significativa para los casos de Pinamar y Villa Carlos Paz.

<sup>6</sup>Al igual que en el caso de los modelos univariados seleccionados, para las versiones multivariadas no existe evidencia suficiente al 95 % de confianza de ausencia de no normalidad. Aun así, los residuos no parecen tener una estructura marcadamente asimétrica, pudiendo la existencia de valores atípicos explicar los resultados.