

Universidad Nacional del Litoral  
Facultad de Ciencias Económicas



Trabajo Final de la Licenciatura en Economía

**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)**

**Alumno**

JORGE REYNOLDS, Ramiro Emmanuel

**Directora**

Dra. VICENTIN MASARO, Jimena

Santa Fe – Febrero 2020

## **DEDICATORIA**

A la memoria de mi padre, guía y pilar fundamental en mi vida.

## **AGRADECIMIENTOS**

A mi madre por el apoyo de siempre y los valores transmitidos.

A Micaela por su invaluable incondicionalidad y afecto.

A Jimena por su paciencia y colaboración indispensable en la elaboración de esta tesina.

A mis amigos de toda la vida.

## RESUMEN

La disponibilidad de información en general, y económica en particular, es sumamente necesaria para la toma de decisiones sobre recursos escasos. La oportunidad de dicha información en el proceso decisorio permite al agente estar preparado y formar sus expectativas en torno a un hecho. El avance de la tecnología y la cantidad de información disponible resultan un desafío en los tiempos modernos; muchos investigadores comienzan a indagar acerca de la utilidad que tiene este gran caudal de información. En este contexto, el uso generalizado de los buscadores en Internet y la posibilidad de conocer patrones a partir de las estadísticas que de éstos se genera se ha vuelto un tópico que está en boga actualmente, dando lugar a que muchos autores recomienden su uso para fines predictivos. La presente tesina tiene como objetivo evaluar la posibilidad de utilizar información gratuita de las herramientas de motores de búsqueda para predicción de un indicador de actividad económica, como es el ICASFe. Así, se pretende conocer si es posible disminuir los tiempos de espera en su publicación, modelando la evolución a partir de dos métodos: uno univariado y otro multivariado, este último con utilización de datos de los motores de búsqueda. A partir de la modelación en series de tiempo, los resultados no muestran que se hayan logrado mejoras significativas en la predicción del indicador a partir de la incorporación de información proveniente de Google; no se encuentra una reducción significativa el error promedio de estimación respecto al modelo univariado, aunque las predicciones del modelo multivariado son más estables y aproximan mejor en el muy corto plazo, dándole un valor adicional al modelo multivariado. Sin embargo, la falta de una metodología objetiva de selección de la información sobre los patrones de búsqueda y la no incorporación de endogeneidad se convierten en limitantes que deben ser tratadas en futuros trabajos.

**Palabras claves:** información asimétrica, herramientas de Google, predicción económica

## ABSTRACT

The availability of general information, and economic information as well, is extremely necessary for making decisions about scarce resources. The opportunity of such information in the decision-making process enables the agent to be prepared to form his expectations around a fact. The technology and the volume of information available are a challenge in modern times; many researchers begin to inquire about the usefulness of this large amount of information. In this context, the widespread use of Internet search engines and the possibility of knowing patterns based on the statistics generated from them, has become a topic traveling. Many authors recommend their use to predictive purposes. This thesis aims to evaluate the possibility of using free information from search engine tools to forecast an economic activity index, such as ICASFe. Thus, the idea is to know whether it is possible to reduce the waiting times in its publication, modeling its evolution from two methods: a univariate and another multivariate, the latter with search engine data. Using the time series modeling, results do not show significant improvements in the prediction the indicator forecast that has Google information; There is no significant reduction in the average estimation error respect to the univariate model, although the predictions of the multivariate model are more stable and approximate better in the very short term, giving an additional value to the multivariate model. However, the lack of an objective methodology to words selection and the non-incorporation of endogeneity become limitations that should be treated in future work.

**Keywords:** asymmetric information, Google Tools, economic forecast

## INDICE DE CONTENIDOS

Dedicatoria .....	II
Agradecimientos .....	III
Resumen .....	IV
Abstract .....	V
Capítulo 1 .....	1
1.1. Problema de investigación .....	1
1.2. Objetivos .....	4
1.2.1. Objetivo general .....	4
1.2.2. Objetivos específicos .....	4
1.3. Hipótesis .....	4
1.4. Estructura de la tesis .....	5
Capítulo 2 .....	6
2.1. Introducción .....	6
2.2. La información imperfecta como regularidad en los mercados .....	6
2.3. Evolución de uso de <i>World Wide Web</i> como instrumento .....	7
2.3.1. <i>Google Trends</i> .....	9
2.3.2. <i>Google Correlate</i> .....	10
2.4. Antecedentes en predicción de variables económicas mediante herramientas de Google .....	11
Capítulo 3 .....	15
3.1. Introducción .....	15
3.2. Modelos para la estimación .....	15
3.2.1. Univariado .....	16
3.2.1.1. Pruebas de raíces unitarias .....	17
3.2.2. Multivariados: Modelo Autorregresivo con Rezagos Distribuidos .....	19
3.3. Variables y fuentes .....	20
3.3.1. Variable de interés predictivo: ICASFe .....	20
3.3.2. Variables exógenas obtenidas de herramientas de Google .....	22
3.4. Modelos econométricos para predicción del ICASFe .....	26
3.5. Implementación .....	26
Capítulo 4 .....	28
4.1. Introducción .....	28

4.2. Modelo univariado para la predicción .....	28
4.2.1. Análisis de las propiedades del ICASFe .....	28
4.2.2. Selección del modelo univariado.....	30
4.3. Selección del modelo multivariado para la predicción .....	33
4.4. Predicción a corto plazo del ICASFE con modelos seleccionados .....	37
Capítulo 5 .....	40
5.1. Discusión .....	40
5.2. Conclusiones .....	42
Anexo A .....	45
Anexo B .....	47
Bibliografía.....	54

## INDICE DE FIGURAS

Figura 1. Evolución del ICASFe y filtro de Hodrick & Prescott .....	22
Figura 2. Tasa de cambio mensual logarítmica del ICASFe y su valor promedio.....	29
Figura 3. Evolución de la tasa de cambio mensual logarítmica del ICASFe y su valor estimado mediante el modelo univariado.....	32
Figura 4. ICASFe y su valor estimado mediante modelo univariado seleccionado.....	33
Figura 5. Evolución de la tasa de cambio mensual logarítmica del ICASFe y su valor estimado mediante el modelo multivariado .....	36
Figura 6. ICASFe y su valor estimado mediante modelo multivariado seleccionado .....	36
Figura 7. Predicción del ICASFe con modelo univariado.....	38
Figura 8. Predicción del ICASFe con modelo multivariado .....	38

## INDICE DE TABLAS

Tabla 1. Variables que componen el ICASFe .....	21
Tabla 2. Lista de palabras cuyas consultas están mayormente correlacionadas con el ICASFe .....	23
Tabla 3. Consultas relacionadas a los términos de <i>Google Trends</i> .....	24
Tabla 4. Parámetros de las series obtenidas mediante <i>Google Trends</i> .....	25
Tabla 5. Pruebas de raíces unitarias del ICASFe para diferentes modelos de la variable en nivel... 29	
Tabla 6. Pruebas de raíces unitarias de la tasa de cambio logarítmica del ICASFe para diferentes modelos.....	30
Tabla 7. Estadísticos de bondad de ajuste de diferentes modelos univariados .....	31
Tabla 8. Coeficientes estimados del modelo multivariado seleccionado .....	32
Tabla 9. Principales estadísticos de ajustes de los modelos multivariados .....	34
Tabla 10. Estadísticos de adecuación del modelo multivariado seleccionado a los supuestos .....	34
Tabla 11. Coeficientes estimados del modelo multivariado .....	35
Tabla 12. Predicciones del ICASFe y error de estimación de ambos modelos .....	37

# Capítulo 1

## Introducción

### 1.1. Problema de investigación

En un mundo globalizado, donde los contextos de incertidumbre se han vuelto más volátiles y las distancias más cortas, está disponible un amplio caudal de información, y todo ello pareciera traducirse en mercados que se vuelven más inestables de lo que solían ser. Actualmente, el conocimiento de los hechos ocurridos en diferentes partes del mundo se da en tiempo real, precipitando una constante modificación del estado de situación y provocando la necesaria toma de decisiones para responder a dichas alteraciones.

La información oportuna es aquella que se encuentra disponible en un momento conveniente y preciso para la toma de decisiones y por lo tanto, resulta útil para lograr un objetivo determinado. En contextos de incertidumbre contar con información oportuna y veraz tiene un gran valor para los agentes económicos, ya que les permite acotar la imprevisibilidad y evaluar escenarios posibles, de modo que los hechos futuros se puedan prever. Pero por otro lado, recabar más información o evaluar escenarios implica invertir tiempo y recursos, lo que la convierte a la información en un bien costoso. Varian (2010) menciona que existen muchos mercados en el mundo real en los que puede ser muy costoso o incluso imposible obtener información precisa y oportuna.

Diversos motivos influyen en esta falta de oportunidad, desde la incapacidad de procesamiento de los organismos que disponen de la información, hasta la manipulación por parte de determinados agentes que retrasan la difusión por intereses personales, lo que

conlleva a que el valor de la misma se pierda parcial o totalmente. La oportunidad y veracidad son elementos fundamentales, pues permiten que los agentes económicos puedan tomar una decisión objetiva, en el momento adecuado, reduciendo sus costos en términos de eficiencia.

La expansión de las herramientas cibernéticas, el uso de Internet, y la capacidad de almacenamiento, hicieron posible que las instituciones del sector público (e.g. observatorios universitarios, institutos de estadísticas, áreas ministeriales) logren una mejor difusión, posibilitando a los usuarios una mayor accesibilidad a los datos. En igual dirección se desempeñan diversas instituciones del sector privado (e.g. consultoras, centros de estudios de universidades privadas y de entidades civiles, entre otros). Aunque el registro y almacenamiento no es un tratamiento exclusivo de las instituciones, el alto costo de búsqueda y difusión conlleva a que éstas se especialicen en ofrecer información relevante a determinados agentes como un bien necesario para reducir la incertidumbre.

En general, los objetivos que persiguen los oferentes de información a la hora de poner a disposición dicho bien, oscilan entre ofrecer un servicio con ánimo de lucro y la difusión por motivos científicos y de desarrollo del conocimiento. En el primer grupo, innumerable son las consultoras de índole privado que se encargan de la comercialización de los datos procesados, tal es el caso de Bain & Company, McKinsey & Company, The Boston Consulting Group, por citar algunas. Todas ellas ofrecen información generada a partir de análisis específicos, como también datos crudos sin analizar. Así, también en la Argentina existen organismos encargados de poner a disposición de los agentes información de todo tipo, i.e. económica, política, social, de gestión pública, etc. Muchas de ellas son entidades del Estado de orden nacional, provincial y municipal, y estas en general están en mejores condiciones de recabar información con respecto a personas particulares. Pero también, de forma paralela, existen instituciones civiles que realizan publicaciones periódicas de datos económicos relevantes, de manera totalmente gratuita.

En este sentido, la Bolsa de Comercio de Santa Fe a través de su Centro de Estudios y Servicios (CES-BCSF), pone a disposición del público bases de datos, como así también informes descriptivos de diversas temáticas relacionadas con actividad económica, finanzas públicas, producción agrícola, mercado de capitales, entre otros. En el caso particular de su

indicador de actividad económica, denominado Índice Compuesto Coincidente de Actividad Económica de la Provincia de Santa Fe (ICASFe), la publicación se realiza a través de informes mensuales que sintetizan los resultados del índice y sus principales componentes. El indicador en cuestión es de gran importancia debido a que sintetiza la evolución y cambios que se producen en la actividad económica de la provincia de Santa Fe, y esto lo consolida como una fuente de consulta a la hora de la toma de decisiones, para agentes tanto del sector público como del privado. Así, por ejemplo, sirve al Estado para analizar la evolución de la actividad económica e implementar políticas de reactivación cuando sea necesario. De igual manera, los agentes del sector privado pueden fundamentar sus decisiones de producción y consumo de acuerdo a la fase del ciclo económico en el que se encuentren.

Aun considerando la relevancia y utilidad del ICASFe, el mismo enfrenta una limitación común a este tipo de indicadores, a saber, la falta de una oportuna disponibilidad de los datos que se insumen en su elaboración y esto ocasiona que la publicación se realice con un rezago de dos meses. Aun así, el índice reduce los tiempos de espera en relación a los indicadores de nivel de actividad nacional que presentan una evolución significativamente similar a la del índice provincial. Sin embargo, los dos meses de rezago siguen siendo una falencia del indicador en su objetivo de ser una herramienta útil y oportuna en la toma de decisiones.

Es por esto que el presente estudio procura reducir dichos tiempos de espera en la elaboración del indicador, tratando de predecir su evolución con la mayor fiabilidad posible y utilizando para ello información regional disponible en Internet. Mediante las herramientas totalmente gratuitas que brindan los motores de búsqueda, se pretenden abordar dos cuestiones: 1- determinar cuán útiles son dichos motores para predecir los acontecimientos económicos analizando un caso concreto y 2- utilizar una aplicación específica para conocer si, efectivamente existe reducción de costo de búsquedas mediante la utilización de Internet.

En este sentido, se espera que la presente investigación permita desarrollar un modelo adecuado para la predicción del ICASFe utilizando información proveniente de los motores de búsqueda de Internet y reduciendo el rezago que afecta la publicación de dicho indicador

económico. Esto implicaría un doble beneficio, en primer lugar, la reducción del tiempo de espera, y en segundo lugar, el desarrollo de un método predictivo que insume información gratuita, oportuna y de fácil disponibilidad.

Sabiendo de la existencia de las herramientas mencionadas y el potencial de las mismas para realizar estimaciones, el problema de investigación de este documento se sintetiza en la siguiente consigna: ¿Es posible, a través de las herramientas *Google Trends* y *Google Correlate* realizar estimaciones y predicciones confiables del Índice Compuesto de Actividad Económica de la Provincia de Santa Fe (ICASFe)?

## **1.2. Objetivos**

### **1.2.1. Objetivo general**

Analizar la utilidad de las herramientas de *Google Trends* y *Google Correlate* para la predicción y adelanto del índice Compuesto de Actividad Económica de la Provincia de Santa Fe (ICASFe).

### **1.2.2. Objetivos específicos**

- Utilizar las herramientas de Google para buscar las palabras claves cuyo patrón de búsquedas estén asociadas con el ICASFe en el período 2004-2019.
- Seleccionar las palabras claves que tengan la mayor correlación con el ICASFe.
- Definir un modelo univariado del ICASFe para la predicción de corto plazo.
- Estimar un modelo que incorpore a las palabras claves obtenidas de las herramientas de Google como variables predictivas.
- Analizar los resultados obtenidos con ambos modelos y hacer predicciones de corto plazo.
- Exponer los principales resultados y conclusiones de los modelos predictivos, y comparar las metodologías de predicción univariada y multivarida.

## **1.3. Hipótesis**

En línea con el problema y los objetivos propuestos, la hipótesis fundamental es que estas herramientas proporcionan información útil para realizar una buena estimación del

ICASFe, ganando oportunidad, y por lo tanto, reduciendo la incertidumbre de los agentes en la toma de decisiones. Es decir, que el modelo multivariado donde se incorporan como variables exógenas información de las herramientas de Google permiten realizar predicciones de corto plazo con menor error que las del modelo univariado.

#### **1.4. Estructura de la tesis**

La estructura de este trabajo consta de cinco capítulos. En el presente primer capítulo, de carácter introductorio, define los lineamientos generales de la investigación, comprendiendo el planteo de la problemática, los objetivos e hipótesis, como así también una fundamentación de los motivos por los cuales resulta pertinente desarrollar este estudio. El segundo capítulo, comprende la exposición de las definiciones y conceptos que enmarcan la problemática y que definen el problema a analizar. Se enuncia el marco teórico sobre el que se fundamenta el análisis. Adicionalmente, se realiza una breve descripción del estado de situación y la evolución de la información que brindan los motores de búsqueda. El tercer capítulo define los materiales y métodos utilizados en esta investigación. En dicho apartado se exponen las fuentes, el procedimiento metodológico y pruebas realizadas. En el capítulo cuatro se muestran los resultados. Finalmente, en el capítulo cinco se expone la discusión del estudio, y se esbozan conclusiones y futuras líneas de trabajo.

# Capítulo 2

## Contextualización y Marco Teórico

### 2.1. Introducción

En este capítulo se realiza una breve descripción del marco teórico y estado del arte sobre el que se fundamenta el presente análisis. Se exponen los problemas de información imperfecta y la utilización de Internet como fuente de datos que podría ayudar en la actualidad. Particularmente, se considera la importancia de los motores de búsqueda y las herramientas que estos brindan a los usuarios, las cuales facilitan la labor de recopilación y acceso a la información contenida en la *web*. También se exponen algunos antecedentes referidos a la utilización de herramientas del mencionado motor de búsquedas y aplicaciones relacionadas, como instrumentos para la realización de estimaciones económicas.

### 2.2. La información imperfecta como regularidad en los mercados

Los pioneros en exponer los problemas de información son Akerlof (1970), Spence (1974), Stiglitz (1981). Los escritos de estos autores representan algunas de las contribuciones más importantes en el estudio de las consecuencias de la información incompleta, haciéndolos merecedores del Premio Nobel de Economía del año 2001. Si bien sus trabajos se centran en el caso puntual de los problemas de información asimétrica, sus estudios dan lugar a un creciente interés por el análisis de la “economía de la información”. Ésta se ocupa del estudio de las relaciones entre los agentes económicos en situaciones donde existen asimetrías de información y la modelización de procesos en los cuales hay que tomar

decisiones cuando la información disponible es incompleta (Vilaseca, Torrent, & Lladós, 2001).

La perfecta información en todos los mercados simultáneamente es una utopía en la realidad. Según el *mainstream*, la falta de información por parte de todos los agentes no resulta un problema en sí mismo, pero da lugar a situaciones en las que existe la posibilidad de que algunos cuenten con mayor información que otros, proporcionándole a los primeros la oportunidad de seguir comportamientos oportunistas que le permitan lograr condiciones ventajosas en un intercambio. Todo ello conlleva a un equilibrio que no es de competencia perfecta y, por ende, a un menor nivel de bienestar general. Así, Varian (2010) señala que la información asimétrica puede ser un obstáculo para el funcionamiento eficiente de un mercado. El elevado costo de obtención de la información puede dar lugar a que productos de mala calidad sean adquiridos o los agentes tomen decisiones erróneas.

En el ámbito del avance tecnológico, Internet juega un papel fundamental. Esta red informática global revoluciona a la economía de la información, al posibilitar el acceso a un universo de datos que algunas décadas atrás habría resultado sencillamente impensado. El notable desarrollo tecnológico de los últimos años, posibilita un incremento considerable de la disponibilidad de datos y la consecuente reducción en los costos de obtención de los mismos. Como destaca Varian (2010), el nivel de conectividad y la reducción de los costos de búsqueda es tan importante que ha llegado a ser considerado como una revolución de la información.

Sin embargo, la pregunta es si el avance de la era de la información decanta en un beneficio o en una desventaja en el proceso de reducción de las barreras en el acceso. La disponibilidad de datos no significa disponibilidad de información. Es necesario tener el conocimiento adecuado para el procesamiento y consiguiente transformación en información.

### **2.3. Evolución de uso de *World Wide Web* como instrumento**

Internet comienza a utilizarse a finales de la década de 1960 y se consolida a partir de 1980 a nivel mundial, propiciando el ambiente para un desarrollo tecnológico sin precedentes. En

Argentina, el servicio recién se comercializa de manera residencial a partir del año 1995. Para el año 2000, según datos de la Unión Internacional de Telecomunicaciones (UIT), el 7,0% de la población argentina tenía acceso a Internet. Una década más tarde, en 2010, la cifra ascendía a 45,0% y en 2017 el 74,3% de los argentinos contaba con este servicio, ya sea a través de computadoras o teléfonos celulares.

La difusión de Internet favorece el acceso de la información por parte de los usuarios, ya sean públicos o privados. Teniendo en cuenta que los agentes económicos valoran la información por considerarla un bien útil, pero que, a su vez tiene como contrapartida el costo de búsqueda, Internet constituye una herramienta que permite una mejora en la situación de los usuarios, ya que reduce los costos de acceso e incrementa el caudal disponible. El acceso a dicha información es posible desde los hogares, gracias al uso de Internet. De igual manera, las diferentes instituciones, y los agentes económicos en particular tienen hoy a disposición una mayor cantidad de datos que antes de la expansión de Internet.

En paralelo a la expansión de Internet se fueron desarrollando diversas herramientas, programas y productos que alimentan endógenamente la dependencia de los usuarios. Algunos de ellos son los motores de búsqueda cuyo algoritmo de funcionamiento brinda la posibilidad de arribar a sitios por medio de palabras claves. Su principal función es encontrar archivos almacenados y direcciones útiles que posibilitan a los usuarios acceder a los sitios de interés vinculados a su consulta. En la actualidad existen numerosos buscadores, cada uno con sus particularidades de funcionamiento y algoritmo de búsqueda. Algunos de los motores de búsqueda disponibles en la actualidad son *Aol*, *Ask*, *Bing*, *Google*, *Yahoo Search*, *Flickr*, *DuckDuckGo*, *StartPage*, *Ecosia*, *Dogpile*, entre otros; aunque el más importante en los países occidentales es Google, con una cuota de mercado que supera el 70,0% en Estados Unidos, Alemania, Gran Bretaña, Francia y Australia (Haucap & Heimeshoff, 2013).

El desarrollo que tiene Google en torno a su funcionamiento permite su primacía. Posee un proceso de desarrollo en investigación que lo hace posicionarse como el primero y, además de las funciones de motor de búsqueda, anexa diferentes aplicaciones que permiten a sus

usuarios reducir los costos de búsqueda. Algunas de las herramientas con las que cuenta la firma son:

- *Google Adwords* que permite crear anuncios,
- *Google Adsence* publica anuncios de Google Adwords,
- *Google Analytics* ofrece información del tráfico de los sitios web según la audiencia, la adquisición, el comportamiento y las conversiones que se llevan a cabo en los mismos,
- *Google Blogger* permite publicar un blog o bitácora en la red,
- *Google Correlate* que muestra la correlación entre una serie de referencia y las búsquedas de una determinada palabra,
- *Google Trends* expone los términos de búsqueda más populares y sus tendencias, entre otras.

En particular, se destacan las últimas dos, dado que ponen a disposición de los usuarios datos gratuitamente que pueden ser procesados para generar información útil. A continuación, se describe el modo en que funcionan las mismas.

### **2.3.1. *Google Trends***

Es una herramienta de libre acceso y gratuita que permite conocer el nivel de búsqueda de un determinado término o palabra clave (*keyword*) a lo largo de un período de tiempo determinado, lo que posibilita identificar las variaciones en las búsquedas en dicho período. La empresa comienza a publicar los datos en el año 2004, y están disponibles hasta la actualidad.

Esta herramienta brinda la posibilidad de obtener información de las búsquedas con diferentes periodicidades, i.e. anual, mensual, semanal, y hasta por hora. Posee un mecanismo que resulta bastante sencillo para el usuario. Éste 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 mínima es por provincia), el período de tiempo, la categoría (arte, ciencias, compras, deportes, etc.) y el tipo de búsqueda (en la web, en noticias, imágenes).

Sin embargo, la información obtenida respecto al nivel de búsqueda de la palabra clave en cuestión no se muestra en términos absolutos, sino que es un indicador, con escala de 0 a 100, donde 100 representa el punto más alto en niveles de búsquedas de la consulta. También expone un cuadro de “consultas relacionadas” que lista los términos que también fueron buscados por usuarios que ingresaron la palabra en cuestión. En el Anexo A se explicitan los resultados de una búsqueda ejemplo en *Google Trends*.

### **2.3.2. *Google Correlate***

*Google Correlate* utiliza un método automatizado para la selección de consultas relacionadas a una serie de referencia. Lo hace a través de un algoritmo que, a partir de la aplicación de coeficientes de correlación, devuelve un conjunto de palabras cuyas búsquedas están mayormente correlacionadas con una serie determinada. En dicho proceso de identificación tiene en cuenta las dimensiones del tiempo y lugar.

Emplea un algoritmo de aproximación sobre millones de consultas en un árbol de búsqueda en línea con el objetivo de arribar a resultados similares al enfoque empleado por *Google Trends* pero utilizando un proceso inverso (Mohebbi et al., 2011). Puntualmente, dado un patrón de interés temporal o espacial, se determina qué consultas imitan mejor los datos. Estas búsquedas pueden servir para construir una estimación del valor del fenómeno (*proxy*).

En este caso, *Google Correlate* pone a disposición un indicador de varianza en relación a la palabra clave, estandarizado con una media de 0 y un desvío estándar de 1. Es decir, expone la evolución de un indicador que se expresa en desviaciones estándar por encima y por debajo de la media.

Cada base de datos contiene decenas de millones de consultas, las cuales provienen de los registros anónimos de búsquedas en la web de Google. Al igual que *Google Trends*, los datos se encuentran disponibles desde enero de 2004, pero la actualización fue discontinuada en marzo de 2017.

## **2.4. Antecedentes en predicción de variables económicas mediante herramientas de Google**

La perfecta información de los agentes es una utopía, pero no es la falta de información generalizada un problema del *mainstream*, sino la asimetría de la misma. Existe información asimétrica desde el momento en que ciertos agentes tienen mayor acceso que otros. Esto les permite obtener ventajas sobre los menos informados. Mucho se argumenta sobre el efecto positivo del avance de las herramientas informáticas y las comunicaciones a través de Internet registrado en las últimas décadas, permitiendo aumentar caudal de datos disponible y reduciendo los costos de acceso a los mismos (Choi & Varian, 2009; Blanco, 2014; Vilaseca et al, 2001).

Pero la disponibilidad de datos no garantiza contar con información válida. La transformación de datos en información valiosa requiere de oportunidad en su publicación y capacidad de manipulación. La capacidad metodológica se transforma ahora en una nueva fuente de generación de asimetrías.

Así, por ejemplo, a nivel internacional existen investigaciones que utilizan herramientas de Internet para realizar estimaciones de diferentes variables económicas con resultados precisos en relación a los valores reales. En dicho sentido, Azkitas y Zimmermann (2009) hacen uso de datos de búsqueda en Internet (*Google Trends*) para modelar y predecir las tasas de desempleo en Alemania. Estos autores encuentran una fuerte correlación entre las búsquedas de palabras clave y las tasas de desempleo de ese país. Por su parte, Choi y Varian (2009) comparan metodologías clásicas de predicción con otras donde se incorpora información disponible en herramientas de búsqueda, afirmando que esta última mejora la predicción de “ventas de motores de vehículos y autopartes” y “comienzo de nuevas viviendas” dentro de las actividades económicas analizadas.

En el mismo sentido, Carrière-Swallow y Labbé (2011) utilizan *Google Trends* para elaborar un índice de las compras de automóviles en Chile, denominado *Google Trends Automotive Index* (GTAI). Destacan la necesidad de acortar los tiempos de espera de las estadísticas oficiales para la toma de decisiones de los sectores privados y públicos, y proponen este indicador construido a partir de datos de Internet que, si bien posee errores,

permite observar las tendencias. Los autores mencionan 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.

También Bortoli y Combes (2015) hacen uso de información de búsquedas *online* para predecir la evolución del gasto en bienes de consumo de los hogares de Francia a través de *Google Trends* y, constatan la utilidad de la herramienta en la predicción sólo en los casos de un tipo particular de bienes, a saber, los durables. Otra de las características en la que estos autores insisten son las ventajas relacionadas a la oportunidad y accesibilidad de los datos que estas herramientas brindan.

Con igual criterio pero para estimaciones en el rubro turístico, Artola y Galán (2012) utilizan herramientas de Google para analizar los flujos de turistas británicos que ingresan a España. Estos autores también destacan las ventajas de la información de Internet, debido a su volumen, disponibilidad en tiempo real y gratuidad.

Artola y Galán (2012), así como Park et al. (2016) tienen objetivos similares analizando el rubro turístico. Este último se propone 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. Este trabajo tiene la particularidad de que incorpora una variable *dummy* que permite modelar la ocurrencia de eventos que potencialmente podrían atraer turistas, como pueden ser eventos deportivos, políticas de incentivos en ese sector, ocurrencia de crisis económicas que afecten negativamente el caudal de visitantes, entre otros. En ambos trabajos se concluye que los modelos que incorporan información proveniente de esta fuente tienen un mejor rendimiento que aquellos que sólo internalizan datos sobre la variable dependiente, como es el caso de los modelos autorregresivos.

Un antecedente más reciente en el sector turismo es el trabajo de Dergiades et al. (2018), quienes utilizan información de motores de búsquedas en Internet para predecir el consumo de productos turísticos en el caso de 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 estas herramientas 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 efectividad de la asignación de recursos limitados.

En el trabajo de Naccarato et al. (2018), con el objetivo de predecir la tasa de desempleo juvenil, también se utiliza información proveniente de *Google Trends* en modelos uni y multivariados. Los resultados confirman dos hechos, el primero es que existe una estrecha conexión entre el desempleo juvenil y la búsqueda de empleo en Internet; en segundo lugar, el uso de datos de *Google Trends* modelados a través de vectores autorregresivos produce mejores resultados en términos de errores de pronóstico que el autorregresivo integrado de medias móviles.

Otro antecedente reciente es el aporte de Jun et al. (2017), quienes realizan una descripción del estado del arte analizando 657 trabajos científicos que, con diferentes objetivos, utilizan *Google Trends* en 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 los distintos autores mencionan, dentro de cuales las más importantes son la gran disponibilidad de datos, la oportunidad y gratuidad de éstos; en cuanto a las limitaciones, se destacan las generalizaciones excesivas basadas en las predicciones con datos de esta fuente que no siempre coinciden con los fenómenos del mundo real.

También se han encontrado algunos trabajos iniciales relevantes a nivel nacional. En este sentido, Blanco (2014) propone pronosticar las ventas de automóviles utilizando búsquedas individuales de ciertas palabras clave en Internet y, por otra parte, construir un indicador mensual de consumo anticipado utilizando las categorías preestablecidas de Google. Los datos de *Google Trends* complementan a los tradicionales y se concluye que el indicador construido mejora el pronóstico y supera al modelo autorregresivo simple y al ampliado con el Índice de Confianza del Consumidor (elaborado por la Universidad Torcuato Di Tella). Los resultados refuerzan la hipótesis de que un indicador de búsqueda de términos y

por categorías mejora la capacidad de pronóstico de modelos simples de serie de tiempo. Sin embargo, la principal dificultad inherente es el establecimiento de un criterio objetivo para seleccionar las palabras claves más adecuadas dentro de aquellas con mayor potencial predictivo.

En el caso particular de este proyecto, se pretende realizar una estimación del ICASFe, elaborado por el CES-BCSF, utilizando como insumo la información generada por Google – principal motor de búsqueda en la actualidad –, como así también las diversas herramientas complementarias que este pone a disposición. No se han encontrado muchos trabajos que analicen la problemática de adelantar la disponibilidad de series económicas en Argentina, y menos aún en la provincia de Santa Fe, por lo que la presente propuesta resulta original en dicho sentido. Además, provee de una aplicación y evaluación de herramientas que pueden ser de utilidad para la modelación económica, constituyéndose en el puntapié inicial para un análisis más profundo de cómo incorporar en los futuros análisis la información que Internet comienza a recabar de manera creciente las preferencias de los agentes.

# Capítulo 3

## Materiales y Métodos

### 3.1. Introducción

Este apartado comprende una descripción general de los modelos predictivos utilizados en el análisis de series de tiempo, distinguiendo dos grandes grupos: los univariados y los multivariados que son utilizados en el presente trabajo. Dentro de los primeros, se destaca el modelo autorregresivos integrados de medias móviles (ARIMA)<sup>1</sup>, mientras que en el segundo grupo, los autorregresivos de rezagos distribuidos (ARDL)<sup>2</sup>. Conjuntamente, se presentan las condiciones básicas necesarias para la aplicación de los modelos.

De manera adicional, se realiza una exposición de las variables y fuentes utilizadas a lo largo del trabajo, tanto en lo referido al ICASFe como a la información obtenida a través de las herramientas de *Google Correlate* y *Google Trends*. Finalmente, se exponen los modelos predictivos (univariado y multivariado) y los fundamentos principales de la implementación de los mismos, así como la forma en que se implementan en el presente trabajo.

### 3.2. Modelos para la estimación

Gujarati y Porter (2009) enumeran, en términos generales, cinco enfoques de los pronósticos económicos basados en series de tiempo. Estos son: métodos de suavización

---

<sup>1</sup> Por sus siglas en inglés, *Autoregressive Integrated Moving Average*.

<sup>2</sup> Por sus siglas en inglés, *Autoregressive Distributed Lag*.

exponencial, modelos de regresión uniecuacionales, modelos autorregresivos integrados de promedios móviles (ARIMA), modelos de regresión de ecuaciones simultáneas, y modelos de vectores autorregresivos (VAR). En el presente trabajo se evalúan diferentes modelos predictivos, a saber, modelo univariados donde se incluye sólo la variable de interés predictivo y multivariados que contienen información obtenida de los motores de búsqueda para predecir la variable de interés.

### 3.2.1. Univariado

Los modelos univariados se basan en la capacidad predictiva de una variable respecto a la evolución de sí misma, i.e. plantean la existencia de una relación entre su valor pasado y presente. En series de tiempo, los modelos univariados orientados a realizar predicciones más comunes son los denominados ARIMA. La composición de este modelo puede desagregarse en AR-I-MA. En este sentido, la sigla AR denota un proceso autorregresivo, el cual considera que 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  $t$ . Un modelo autorregresivo de orden 1, se denota como AR(1) y se expresa mediante la ecuación:

$$y_t = \phi y_{t-1} + a_t, \quad (1)$$

siendo  $y_*$  el valor de  $y$  en el momento  $*$ ,  $\phi$  representa la proporción en la que  $y_t$  se compone de  $y_{t-1}$ , mientras que  $a_t$  es la perturbación aleatoria de la variable  $y$  en el momento  $t$ . Generalizando, un modelo AR( $p$ ) queda expresado de la siguiente manera:

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

Un proceso de medias móviles (MA) de orden  $q$  puede definirse como:

$$y_t = a_t - \theta_1 a_{t-1} - \theta_2 a_{t-2} - \dots - \theta_q a_{t-q}. \quad (3)$$

Indica que la innovación pasada ( $t - 1, \dots, t - q$ ) influye parcialmente en el presente  $t$ . Se puede afirmar que un proceso MA( $q$ ) es siempre estacionario, ya que es la suma de  $q + 1$  procesos estacionarios.

Al complementar los dos procesos antes mencionados, se obtiene un modelo ARMA( $p, q$ ), donde  $p$  es el orden del modelo AR y  $q$  el del modelo MA. El mismo puede ser expresado como sigue:

$$y_t = \phi_1 y_{t-1} + \dots + \phi_p y_{t-p} + a_t + \theta_1 a_{t-1} + \dots + \theta_q a_{t-q} . \quad (4)$$

Finalmente, una serie de tiempo también puede ser no estacionaria y ello define el nivel de integración. En este sentido, al diferenciar una serie de tiempo  $d$  veces se logra la estacionariedad de la misma, y luego dicha serie estacionaria sigue un proceso ARMA( $p, q$ ), entonces la serie de tiempo original es un ARIMA( $p, d, q$ ) es decir, es una serie de tiempo autorregresiva integrada de promedios móviles (Gujarati & Porter, 2009). La ecuación (5) muestra este tipo de procesos de la variable aleatoria en series de tiempo:

$$y_t = -(\Delta^d y_t - y_t) + \alpha_0 + \sum_{i=1}^p \phi_i \Delta^d y_{t-i} + \sum_{j=1}^q \theta_j a_{t-j} + \varepsilon_t . \quad (5)$$

Se puede observar que la condición de estacionariedad es relevante a la hora de modelar una serie de tiempo. Para conocer esta condición existen varias pruebas de raíces unitarias. A continuación se describe una de la más usada en la literatura de referencia.

### 3.2.1.1. Pruebas de raíces unitarias

Un proceso de caminata aleatoria<sup>3</sup> o estacionaria refiere a una serie temporal que no posee tendencia ni estacionalidad, y puede definirse de la siguiente manera:

$$y_t = \rho y_{t-1} + u_t \quad \text{con} \quad -1 \leq \rho \leq 1 , \quad (6)$$

donde  $\rho$  representa la autocorrelación entre el valor de la variable  $y$  en el presente y su valor en el momento inmediatamente anterior, i.e.  $t - 1$ , mientras que  $u_t$  es un término de error de ruido blanco o *shock* aleatorio. Cuando se dice que la serie de tiempo posee raíz unitaria, el valor de  $\rho$  es exactamente 1, implicando la no estacionariedad de la variable  $y$ .

---

<sup>3</sup> Se entiende por proceso de caminata aleatoria a aquel en el que la variable de interés en el momento  $t$  depende solo de su valor en algún momento previo y de alguna variable aleatoria o choque aleatorio.

Para corroborar la existencia de raíces unitarias y la consecuente no estacionariedad de una serie de tiempo, existen diferentes pruebas. Una primera aproximación 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  $\hat{\rho}_k \sim (0, 1/n)$ . 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 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 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 utilizan simulaciones de Monte Carlo para calcular los valores críticos de la prueba. Dickey-Fuller (DF) (como se cita en Gujarati y Porter (2009)) indican que una serie de tiempo  $y_t$  puede seguir diferentes modelos:

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

$$\Delta y_t = \delta y_{t-1} + u_t . \quad (7)$$

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

$$\Delta y_t = \beta_1 + \delta y_{t-1} + u_t . \quad (8)$$

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

$$\Delta y_t = \beta_1 + \beta_2 t + \delta y_{t-1} + u_t . \quad (9)$$

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

Una versión más compleja de esta prueba es la prueba Dickey-Fuller Aumentada (ADF)<sup>4</sup> que incorpora 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,

$$\Delta y_t = \beta_1 + \beta_2 t + \delta y_{t-1} + \sum_{i=1}^p \alpha_i \Delta y_{t-i} + \varepsilon_t, \quad (10)$$

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  $\delta$ .

El principal inconveniente inherente a la no estacionariedad de una serie de tiempo es que cualquier predicción que con ella se realice da lugar a errores de estimación no estacionarios y que, por lo tanto, tienen problemas de autocorrelación. La autocorrelación de los errores puede ocurrir aun cuando las series utilizadas en la estimación son estacionarias, pero la no estacionariedad de la serie es condición necesaria para la no estacionariedad del error. Otro problema relacionado a la existencia de raíces unitarias es la aparición de regresiones espurias. Una manera muy utilizada para salvar los problemas de no estacionariedad es el cálculo o modelación de las primeras diferencias de una serie. Sin embargo, esta solución está sujeta al orden de la no estacionariedad.

### 3.2.2. Multivariados: Modelo Autorregresivo con Rezagos Distribuidos

Los modelos multivariados son aquellos cuya modelación involucra un conjunto de variables. Los modelos autorregresivos con rezagos distribuidos (ARDL) son usados para modelar la relación entre variables de series de tiempo en una sola ecuación y son útiles para la predicción y la separación de relaciones de corto y largo plazo entre las variables de interés. Así, suponiendo que  $y_t$  es la variable de interés predictivo y  $x_t = (x_{1t}, x_{2t}, \dots, x_{nt})$  son variables de series de tiempo exógenas, entonces un modelo ARDL( $p, q$ ) puede representarse de la siguiente manera,

$$y_t = \alpha + \varphi t + \sum_{i=1}^p \gamma_i y_{t-i} + \sum_{j=1}^q \beta_j' x_{t-j} + u_t. \quad (11)$$

---

<sup>4</sup> Por sus siglas en inglés, *Augmented Dickey-Fuller*

Con  $p \geq 1$ , y  $q \geq 0$ , donde  $x_*$  es un vector  $n \times 1$ . La parte de la ecuación (11) referida a la autorregresión ( $y_*$ ) utiliza información del pasado para predecir el valor presente de  $y$ , mientras que los términos que involucran a variables exógenas ( $x_*$ ) refieren a información proveniente de otras variables, tanto en su valor presente como en valores pasados, para predecir  $y$  en el momento  $t$ .

Una de las ventajas de estos modelos es que no tienen una exigencia sobre el orden de integración de las series, pudiéndose aplicar tanto en series  $I(0)$  como  $I(1)$ , o una combinación de ambas. Además, permiten tratar el problema de relaciones espurias por medio de la incorporación de rezagos.

### **3.3. Variables y fuentes**

En total la base de datos consta de 16 variables mensuales (15 de estas obtenidas a través de las herramientas de *Google*, y la restante es el ICASFe). El período disponible y coincidente para todas ellas va desde enero de 2004 hasta julio del 2019, comprendiendo un total de 187 datos mensuales para cada variable. Aunque también se cuenta con datos de agosto a noviembre de 2019, los mismos son dejados por fuera de la modelación para realizar la validación de las predicciones.

#### **3.3.1. Variable de interés predictivo: ICASFe**

La variable de interés predictivo en el presente trabajo es el ICASFe. Éste es un índice coincidente de actividad económica de periodicidad mensual que permite identificar las fases de contracción y expansión económica de la provincia de Santa Fe con un rezago de dos a tres meses. El mismo es elaborado por el CES-BCSF desde 2007.

Producido desde enero de 1994 y con base 1994=100, representa la situación económica de la provincia de Santa Fe y se convierte en una herramienta de información para la toma de decisiones en ámbitos públicos y privados. De acuerdo a sus características, resulta una fuente de consulta para un heterogéneo grupo de usuarios, desde empresarios, docentes, y alumnos, hasta funcionarios del sector público y medios de comunicación, como así también para el público en general.

En particular, insume información proveniente de catorce series temporales de interés económico de alcance provincial. La inclusión de estas variables se fundamenta en que: (1) brindan información referida al espacio geográfico, (2) son representativas de variables con significancia económica, (3) tienen una periodicidad mensual, y (4) poseen una disponibilidad y una fecha de inicio común (D’Jorge *et al.*, 2018)<sup>5</sup>. Las variables componentes del ICASFe se muestran en la Tabla 1.

**Tabla 1.** Variables que componen el ICASFe

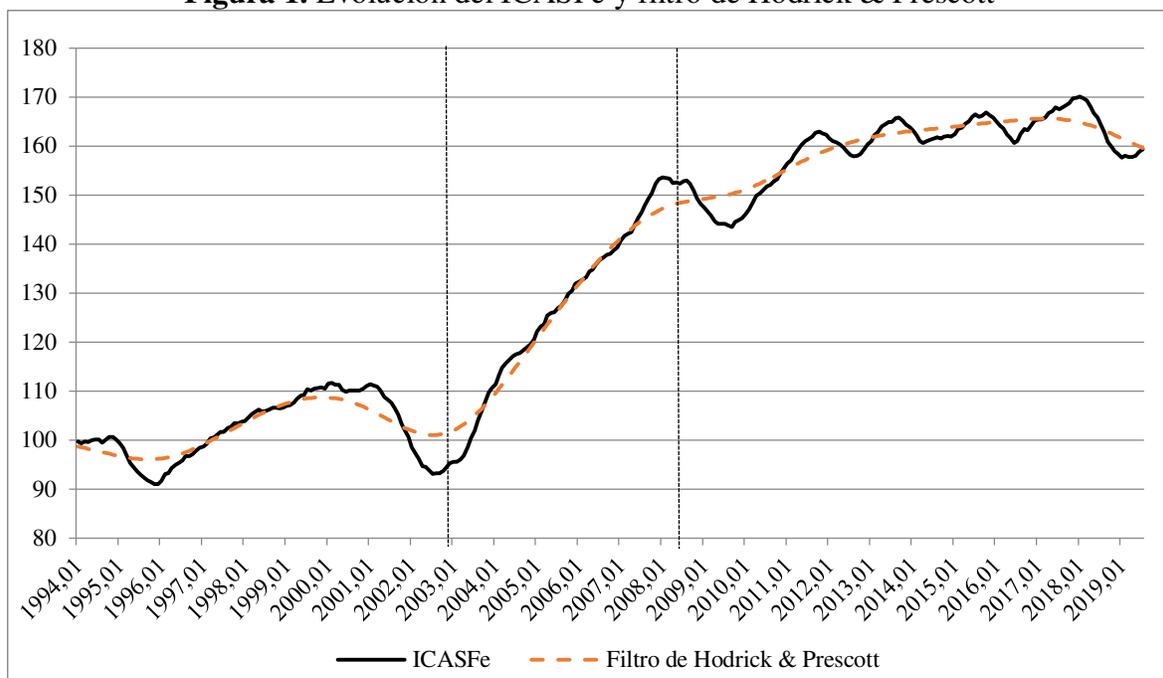
<b>Bloque</b>	<b>Variable</b>
Empleo	- Número de puestos de trabajo registrados en la provincia
	- Índice de demanda laboral
Producción Industrial	- Consumo de energía eléctrica industrial
	- Consumo de gas industrial
	- Consumo de hidrocarburos líquidos
	- Faena de ganado bovino y porcino
	- Producción industrial de lácteos
	- Molienda de oleaginosas
	- Ventas de maquinaria agrícola
Ventas Minoristas	- Ventas reales de supermercados
	- Consumo de cemento Pórtland
	- Patentamiento de vehículos nuevos
Ingreso Disponible	- Recaudación tributaria de la provincia y coparticipación
	- Masa de remuneraciones reales percibida por los asalariados

Fuente: elaboración propia en base a datos del CES-BCSF

En la Figura 1 se puede ver la evolución del ICASFe desde su constitución (enero de 1994) hasta julio de 2019. Este indicador muestra una tendencia creciente (que puede visualizarse a través del filtro de Hodrick & Prescott) y un patrón cíclico recurrente, con mayor frecuencia a partir de 2008. El índice no muestra una estacionalidad marcada ni la existencia de valores irregulares. Dicha característica es precisamente deseable en la construcción de este tipo de indicadores, ya que su propósito es la estimación de la tendencia-ciclo de la serie de actividad económica, y no sus oscilaciones en el corto plazo.

<sup>5</sup> Además las series deben respetar ciertas condiciones técnicas, a saber: (1) conformidad: deben fluctuar en forma aproximadamente sincrónica con el ciclo económico de referencia; (2) correspondencia temporal, (3) significancia económica, (4) consistencia estadística, proveniente de fuente confiable, (5) suavidad: no presentan una gran volatilidad intrínseca, y (6) oportunidad con fechas de publicación ajustada a un cronograma estable.

**Figura 1.** Evolución del ICASFe y filtro de Hodrick & Prescott



Fuente: elaboración propia en base a datos del CES-BCSF

En términos desagregados, el período 1994-2003 presenta marcadas oscilaciones en el nivel de actividad, sin denotar una variación significativa en niveles. Entre 2003 y 2008 se observa el mayor crecimiento de la serie en todo el período bajo análisis; es a partir de 2008 que el nivel de actividad económica de la provincia comienza a crecer a una tasa menor, y hacia 2012 pasa a registrar un estancamiento relativo, similar al observado en el primer sub-período mencionado.

### 3.3.2. Variables exógenas obtenidas de herramientas de Google

En el modelo multivariado se relacionan la variable de interés predictivo con otras variables exógenas obtenidas de las herramientas de Google, vinculadas con las búsquedas en este motor. Así, cuando la serie del ICASFe se ingresa a *Google Correlate* se obtiene una lista de las 100 consultas cuyos patrones de búsquedas tienen mayor correlación con la serie de referencia. Esta herramienta permite detectar términos de búsqueda con un comportamiento similar al de dicha serie de referencia, algunos de los cuales de manera intuitiva podrían no ser tenidos en cuenta. En este trabajo, y considerando que es una primera aproximación a la utilización de estas herramientas, se decide considerar las 5 consultas con mayor correlación con la serie del ICASFe presentada en la Figura 1.

La Tabla 2 contiene el listado de dichas palabras, conjuntamente con su nivel de correlación respecto al ICASFe. Se muestran las cinco palabras claves cuya correlación es mayor en el período enero 2004 - marzo 2017. Así, por ejemplo, las búsquedas mensuales de “coopeplus” tuvieron un nivel de correlación de aproximadamente 96,08% con la evolución del ICASFe durante ese período. De manera similar se interpretan las demás variables.

**Tabla 2.** Lista de palabras cuyas consultas están mayormente correlacionadas con el ICASFe

<b>Palabras Correlacionadas a ICASFe</b>	<b>Coefficiente de Correlación con ICASFe</b>
Coopeplus	0,9608
Vallan	0,9495
Cualquier	0,9484
Pasar	0,9478
Bancopatagonia	0,9440

Fuente: elaboración propia en base a Google Inc.

Dentro de los términos con mayor correlación existen dos palabras que puede interpretarse con significado económico, estas son “coopeplus” y “bancopatagonia” (Banco Patagonia). El primer caso corresponde a una tarjeta de crédito emitida por Nueva Card S.A., mientras que Banco Patagonia es un banco privado que surge de la fusión entre Banco Mercantil Argentino y Banco Caja de Ahorro en el año 1999 y que tiene operatoria en la Provincia de Santa Fe. Tanto la tarjeta Coopeplus como el Banco Patagonia tienen alcance nacional y por su significado económico, es de esperar que posean cierto grado de correspondencia con el nivel de actividad del país. Paralelamente, el nivel de actividad de la provincia de Santa Fe, medido a través del ICASFe, tiene un elevado grado de coincidencia con el desempeño económico nacional. Es por esto que se espera exista una relación comprobable entre la evolución de las búsquedas de esas palabras y el índice de actividad provincial.

Posteriormente, los términos obtenidos mediante *Google Correlate* se ingresan a *Google Trends* para conocer qué otras palabras se relacionan con cada uno de ellos. Lo que se busca es conocer qué consultas realizan los usuarios de Google de manera conjunta con cada uno de los términos de la Tabla 2. En la Tabla 3 se observan las dos palabras cuyas

consultas tienen mayor relación respecto a cada uno de los cinco términos obtenidos a través de *Google Correlate*.

**Tabla 3.** Consultas relacionadas a los términos de *Google Trends*

<b>Búsqueda</b>	<b>Consultas Relacionadas</b>
Coopeplus	Coopeplus tarjeta Coopeplus bahía blanca
Vallan	Colectivos Valla
Cualquier	Cualquier coincidencia con la realidad es pura coincidencia Cualquier coincidencia con la realidad
Pasar	Como pasar credito Como pasar
Bancopatagonia	Bancopatagonia e bank Banco

Fuente: elaboración propia en base a datos de Google Inc.

Los elementos presentados en la Tabla 3 se corresponden con búsquedas que se realizan en el período enero 2004 a noviembre de 2019, en Argentina y se consideran los términos “principales”.<sup>6</sup>

Ambas herramientas, además de indicar cuáles son las palabras claves, ponen a disposición índices sobre la evolución de búsqueda de cada palabra durante el período de referencia. El indicador de *Google Correlate* (GC) es una medida de variabilidad alrededor de la media, i.e. muestra la variabilidad de búsqueda alrededor del promedio durante el período enero 2004 a marzo 2017. Por tanto, cada serie de datos asume valores entre -3 y 3. El comportamiento de estas series se asemeja al del ICASFe, dado que mantienen un alto nivel de correlación.

En lo que respecta a las series obtenidas mediante *Google Trends* (GT), se expresan a modo de índice, con un valor mínimo de 0 y un valor máximo de 100. La Tabla 4 contiene estadísticos descriptivos de las 15 series que son tratadas como variables exógenas en los

<sup>6</sup> Se pueden seleccionar “principales” o “en aumento”, los primeros son aquellos que más se consultaron de manera simultánea con la consulta de referencia a lo largo de todo el período analizado, mientras que los que se encuentran en aumento, son las consultas con mayor simultaneidad en los últimos meses.

modelos multivariados (las 5 obtenidas por sugerencia de GC y las 10 consultas relacionadas), para el período enero 2004 a julio 2019.

**Tabla 4.** Parámetros de las series obtenidas mediante *Google Trends*

<b>Palabras Consultadas</b>	<b>Valor mínimo</b>	<b>Valor Máximo</b>	<b>Promedio</b>	<b>Mediana</b>	<b>Desvío Estándar</b>
Coopeplus	0,00	100,00	55,63	59,00	24,24
Vallan	0,00	100,00	40,60	43,00	15,51
Cualquier	16,00	100,00	72,49	75,00	17,29
Pasar	26,00	100,00	70,76	74,00	16,54
Bancopatagonia	0,00	100,00	52,79	58,00	24,22
Banco	28,00	100,00	65,39	69,00	17,22
Bancopatagonia e bank	0,00	100,00	22,18	8,00	28,99
Coopeplus bahía blanca	0,00	100,00	11,55	9,00	16,18
Coopeplus tarjeta	0,00	100,00	17,95	17,00	17,22
Cualquier coincidencia con la realidad es pura coincidencia	0,00	100,00	28,22	31,00	24,45
Cualquier coincidencia con la realidad	0,00	100,00	32,37	37,00	24,16
Como pasar crédito	0,00	100,00	28,73	31,00	21,03
Como pasar	8,00	100,00	54,46	56,00	23,24
Colectivos	21,00	100,00	38,05	36,00	11,25
Valla	0,00	100,00	61,33	70,00	23,01

Fuente: elaboración propia en base a datos de Google Inc.

De esta manera, el proceso completo consta de ingresar la serie del ICASFe en GC. Los cinco términos con patrones de búsqueda que más se asemejan a la serie son “coopeplus”, “vallan”, “cualquier”, “pasar” y “bancopatagonia”, estas salidas se miden en desvíos estándar alrededor del valor promedio de búsquedas del período enero 2004 a marzo 2017. Luego se introducen los cinco términos a GT por tres motivos. El primero de ellos es obtener las consultas relacionadas – en este caso: “Banco”, “Bancopatagonia e Bank” y todas las que completan la lista –; el segundo objetivo es la disponibilidad de datos desde marzo de 2017 (fecha desde la cual se discontinúan las publicaciones de GC) hasta la actualidad, y por último, este proceso permite lograr que todos los términos se encuentren medidos en índices de búsquedas, con un valor máximo de 100, correspondiente al momento en el que la cantidad de consultas del término en cuestión alcanza su valor más alto.

### 3.4. Modelos econométricos para predicción del ICASFe

Siendo el ICASFe la variable de interés predictivo, denotada por  $\widehat{ICASFe}_t$ , y dado que el modelo univariado no involucra otra variable, entonces el modelo ARIMA( $p,d,q$ ) puede escribirse en términos generales de la siguiente manera:

$$\widehat{ICASFe}_t = -(\Delta^d ICASFe_t - ICASFe_t) + \hat{\alpha}_0 + \sum_{i=1}^p \hat{\phi}_i \Delta^d ICASFe_{t-i} + \sum_{j=1}^q \hat{\theta}_j a_{t-j}. \quad (12)$$

En cambio, el modelo multivariado incluye el resto de las variables exógenas obtenidas mediante las herramientas de Google, por lo que el modelo econométrico es el siguiente:

$$\widehat{ICASFe}_t = \hat{\alpha} + \hat{\phi}t + \sum_{i=1}^p \hat{\gamma}_i ICASFe_{t-i} + \sum_{j=1}^q \hat{\beta}_j' \mathbf{x}_{t-j}, \quad (13)$$

donde  $\mathbf{x}_*$  es un vector  $15 \times 1$  que contiene los indicadores de búsqueda de las palabras de GC y GT en un momento determinado.

### 3.5. Implementación

El proceso de análisis del presente trabajo implica realizar las pruebas de estacionariedad de la serie ICASFe con el objeto de conocer el orden de integración para modelarla univariadamente. Estas pruebas se realizan para el período coincidente con la disponibilidad de las variables exógenas, i.e. enero 2004 – julio 2019.

Luego, tanto en el modelo univariado como en el multivariado se analiza la cantidad óptima de rezagos para la variable dependiente y las independientes, utilizando para ello los Criterios de Información de Akaike (AIC) y de Bayes (BIC). El método de estimación utilizado en ambos casos es el de máxima verosimilitud. En todos los casos se realiza un análisis de normalidad y auto-correlación de los residuos mediante prueba de Jarque Bera y Durbin-Watson, respectivamente.

Una vez obtenidos los parámetros de los modelos, se realiza una predicción de corto plazo para el período agosto a noviembre del 2019, los cuales son comparados con los valores reales del ICASFe analizando la precisión de las estimaciones logradas.

En el análisis econométrico se utilizan los programas *R Core Team* (2020) y *GRETL*<sup>7</sup>, ambos gratuitos y de acceso libre.

---

<sup>7</sup> Por sus siglas en inglés: *Gnu Regression, Econometric and Time-series Library*.

# Capítulo 4

## Resultados

### 4.1. Introducción

En este capítulo se exponen los resultados de las estimaciones de los modelos definidos en el apartado anterior. En primer lugar, se analizan de manera general las propiedades del ICASFe, se determina y selecciona el modelo univariado más adecuado y se exponen los parámetros correspondientes para la predicción. Luego se procede de manera similar con el multivariado, detallando las variables explicativas a incorporar, el modelo más adecuado y los estadísticos fundamentales del mismo. Posteriormente, se realizan las estimaciones correspondientes y se comparan los resultados obtenidos mediante ambos métodos (univariado y multivariado) con los verdaderos valores observados del ICASFe.

### 4.2. Modelo univariado para la predicción

#### 4.2.1. Análisis de las propiedades del ICASFe

En un primer paso, se predice el valor presente del ICASFe utilizando como variable explicativa el valor del índice en momentos anteriores a  $t$ , i.e. utilizando una estructura autorregresiva. Para ello, se evalúa la estacionariedad de la serie. En la tabla 5 se muestran los resultados de las pruebas de estacionariedad de Dickey-Fuller Aumentado para dos modelos, a saber, uno incorporando sólo una constante y, el segundo, una constante y

tendencia. Además, siguiendo a Schwert (1989) se considera un máximo de 14 retardos<sup>8</sup> eligiendo la cantidad de rezagos óptima según el criterio de AIC.

**Tabla 5.** Pruebas de raíces unitarias del ICASFe para diferentes modelos de la variable en nivel

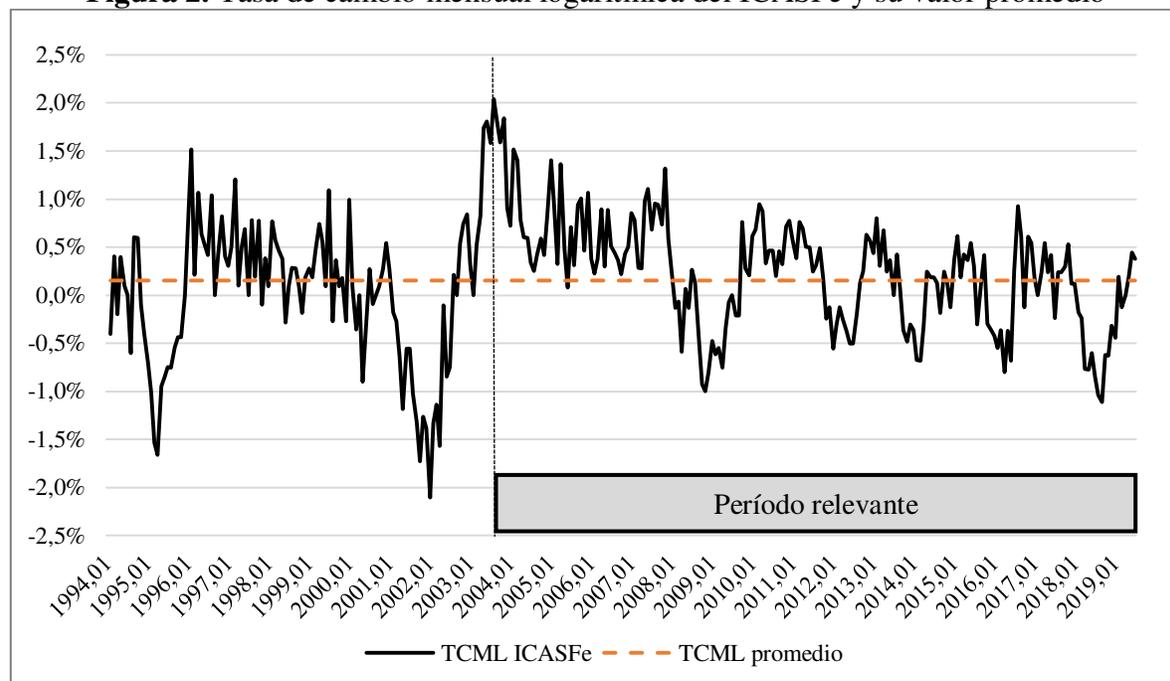
Modelo	Valor p del estadístico $\delta$
Con constante	0,160
Con constante y tendencia	0,261

Fuente: elaboración propia en base a datos de CES-BCSF

Se puede concluir que en ambos casos existe suficiente evidencia de la existencia de raíces unitarias, demostrando que la serie ICASFe en nivel no es estacionaria. El *valor p* del estadístico  $\delta$  en ambos casos es mayor a 0,05, por lo que la hipótesis nula de existencia de raíz unitaria no puede rechazarse al 95% de confianza.

Para saber el grado de integración de la serie, se realiza la prueba ADF a la serie de la primera diferencia, i.e. las tasas de cambio mensuales logarítmicas (TCML). Visualizando la TCML, la Figura 2 muestra la evolución desde febrero de 1994 hasta julio de 2019.

**Figura 2.** Tasa de cambio mensual logarítmica del ICASFe y su valor promedio



Fuente: elaboración propia en base a datos de CES-BCSF

<sup>8</sup> La cantidad de retardos adecuada viene dada por la parte entera de calcular  $12(n/100)^{0.25}$ , donde n es el tamaño de la muestra, en este caso, 187 (datos mensuales desde enero de 2004 hasta julio de 2019).

Se observa que los valores oscilan en torno a un promedio de 0,2%, por lo que tienen una media constante. La amplitud de las variaciones es máxima entre los años 2002 y 2004 (de -2,0% a 2,0%), mientras que en el resto del período temporal se mantiene aproximadamente entre -1,0% y 1,0%.

En el período de interés (enero 2004 – julio 2019) la tasa de cambio del ICASFe reduce su variabilidad respecto al período anterior y no presenta una tendencia marcada. Los resultados de las pruebas ADF de la TCLM del ICASFe se exponen en la Tabla 6. En este caso, ambos modelos presentan un *valor p* menor a 0,05, lo que favorece el rechazo al 95% de confianza de la hipótesis nula de existencia de raíces unitarias. Existe evidencia suficiente para afirmar que la serie compuesta por las TCML del ICASFe en el período enero 2004 – julio 2019 es estacionaria, y que la serie en nivel es integrada de orden 1.

**Tabla 6.** Pruebas de raíces unitarias de la tasa de cambio logarítmica del ICASFe para diferentes modelos

<b>Modelo</b>	<b>Valor p del estadístico <math>\delta</math></b>
Con constante	0,000
Con constante y tendencia	0,002

Fuente: elaboración propia en base a datos de CES-BCSF

En el Anexo B se exponen los resultados de los contrastes ADF para ambas series como así también las funciones de autocorrelación y de autocorrelación parcial del ICASFe. Estas funciones permiten determinar la cantidad necesaria de rezagos a considerar en el modelo univariado.

#### **4.2.2. Selección del modelo univariado**

Dado que la serie de primeras diferencias del ICASFe es estacionaria, el modelo de serie de tiempo debe ser de orden  $d = 1$  en la parte integrada, i.e.  $ARIMA(p,1,q)$ , quedando por evaluar cuáles son los órdenes de autorregresión y medias móviles. Siguiendo el principio de parsimonia, analizando de un modo *ad hoc* y teniendo en cuenta las características de las funciones de autocorrelación mencionadas, se consideran hasta tres órdenes en cada caso, i.e.  $p = 1, 2, 3$  y  $q = 1, 2, 3$ , y una combinación de éstas. En la Tabla 7 se muestran los principales estadísticos considerados.

**Tabla 7.** Estadísticos de bondad de ajuste de diferentes modelos univariados

ARIMA( $p,1,q$ )		AIC	BIC	Raíz del Error Cuadrático Medio
$p$	$q$			
1	1	228,654	238,931	0,439
1	2	227,763	240,666	0,435
1	3	227,565	243,694	0,433
2	1	<b>225,705</b>	<b>238,608</b>	0,431
2	2	228,839	244,967	0,434
2	3	226,880	246,235	<b>0,429</b>
3	1	228,144	244,272	0,433
3	2	229,817	249,171	0,433
3	3	228,880	251,460	<b>0,429</b>

Nota: en negrita se muestran los valores mínimos. Fuente: elaboración propia en base a datos de CES-BCSF

Se puede ver que el modelo que posee órdenes de autorregresión de 2, integración de 1 y medias móviles de 1 tiene simultáneamente los valores del criterio de AIC y BIC mínimos, mientras que los modelos con autorregresivos de orden 2 y 3, con medias móviles de 3 minimizan la raíz del error cuadrático medio. Por tanto, el modelo ARIMA(2,1,1) es el que mejor ajuste presenta según dichos criterios. Cuando se analiza la adecuación del modelo y el cumplimiento de los supuestos, la prueba de normalidad de los residuos arroja un *p valor* de 0,515, lo que indica que no se puede rechazar la hipótesis de que los residuos obtenidos tengan una distribución normal. En cuanto a la no autocorrelación, ante la hipótesis nula de no autocorrelación el *valor p* es 0,036, y el contraste ARCH de no heteroscedasticidad proporciona un *p-valor* de 0,230. No existe evidencia ni de autocorrelación ni de heteroscedasticidad de los residuos al 1% de significancia, respectivamente; se puede considerar que el modelo seleccionado tiene el mejor ajuste y, además es aceptable dado que cumple con los supuestos. En el Anexo B se exponen de manera gráfica los residuos, su correlograma y distribución, que verifican estas condiciones.

En la Tabla 8 se muestran los coeficientes y desvíos estándar estimados de la aplicación del mismo a los datos del ICASFe entre enero de 2004 y julio de 2019:

**Tabla 8.** Coeficientes estimados del modelo multivariado seleccionado

Variable	Coefficiente	Desvío Estándar
Intercepto	0,315	0,182
ICASFe <sub>t-1</sub>	-0,104*	0,048
ICASFe <sub>t-2</sub>	0,767**	0,049
a <sub>t-1</sub>	1,000**	0,015

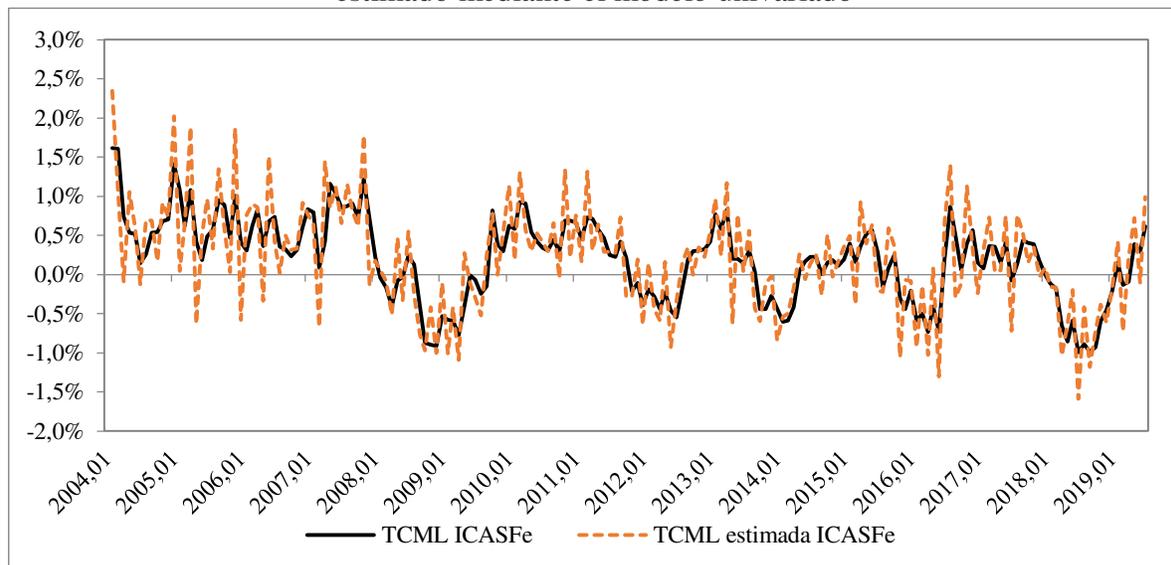
Nota: significativo al nivel \*5%, \*\*1%

Fuente: elaboración propia en base a datos de CES-BCSF

Se observa que el modelo ARIMA(2,1,1) posee un coeficiente positivo para el término autorregresivo de orden 2 y las medias móviles, mientras que el coeficiente es negativo en el caso del autorregresivo de orden 1. Todos los coeficientes son significativos al nivel 1%, excepto el intercepto.

La Figura 3 expone la tasa de cambio del ICASFe y su valor estimado a través del modelo ARIMA(2,1,1) para el período enero 2004 – julio 2019. Dicho espacio temporal es coincidente con la disponibilidad de datos provenientes de las herramientas de Google, permitiendo esto que ambos modelos sean comparables más adelante.

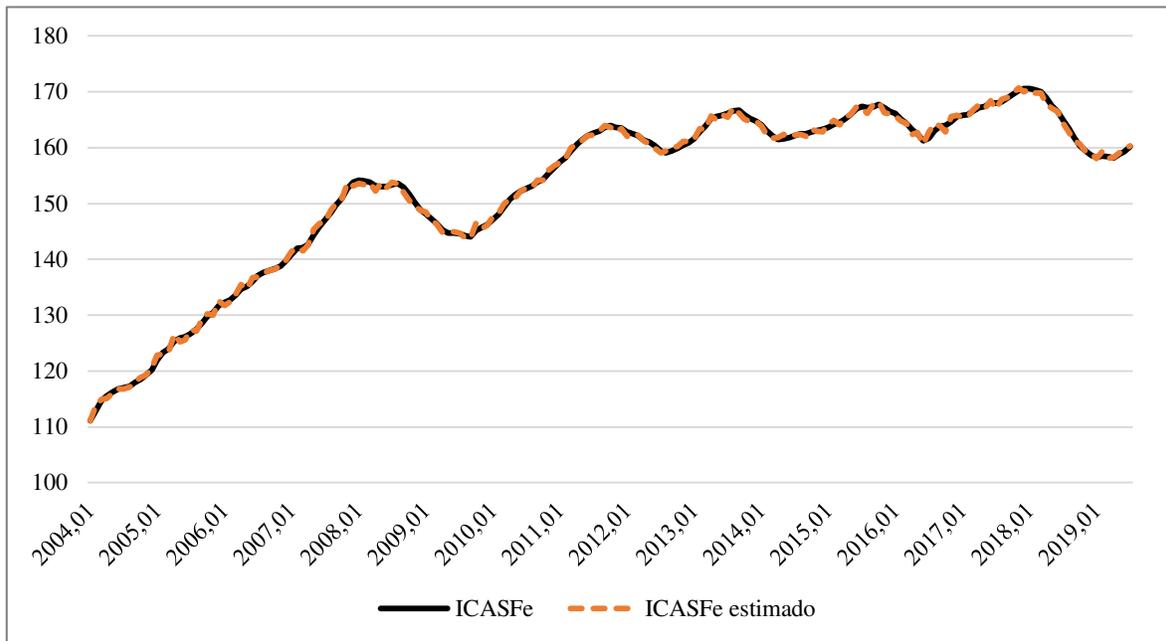
**Figura 3.** Evolución de la tasa de cambio mensual logarítmica del ICASFe y su valor estimado mediante el modelo univariado



Fuente: elaboración propia en base a datos de CES-BCSF

Se puede ver que la predicción dentro del conjunto utilizado para la estimación de los parámetros coincide con la evolución de los valores reales de la tasa de cambio del ICASFe en el período en cuestión (Figura 3).

**Figura 4.** ICASFe y su valor estimado mediante modelo univariado seleccionado



Fuente: elaboración propia en base a datos de CES-BCSF

Con el objetivo de ver el ajuste en la variable en nivel, en la Figura 4 se muestra la predicción del ICASFe en nivel obtenida mediante el modelo univariado. El grado de coincidencia con el verdadero valor del índice para el período bajo análisis es considerable bueno. Sin embargo, no logran apreciarse las desigualdades debido a la amplitud de la escala.

### 4.3. Selección del modelo multivariado para la predicción

Al igual que en el caso univariado, la selección de las variables a incorporar ya sean autorregresivas o exógenas se realiza en función del modelo que presenta mejores criterios de información, considerando hasta dos órdenes para las partes autorregresivas, primando la parsimonia del modelo. Se analizan ocho alternativas que comprenden dos modelos con 15 variables explicativas (5 de GC más 10 de GT), dos modelos con 5 variables (obtenidas de GC), y cuatro modelos con 10 variables (uno con 5 de GC y 5 de GT, el otro con las 10 de GT).

En la Tabla 9 se presentan los estadísticos de las ocho alternativas consideradas. Todas las medidas de bondad de ajuste coinciden en que la mejor alternativa es aquella que incluye

las 5 variables obtenidas de *Google Correlate* e incorpora dos rezagos para las mismas y dos para el ICASFe. Considerando los criterios AIC y BIC entre el modelo univariado y el multivariado, el modelo ARDL(2,2) es la mejor de las alternativas multivariadas y, su valor es levemente superior (234,001 versus 225,705) al del caso univariado.

**Tabla 9.** Principales estadísticos de ajustes de los modelos multivariados

Nº variables	Origen	ARDL	R <sup>2</sup>	R <sup>2</sup> ajustado	Error estándar residual	AIC	BIC
5	GC	(1,1)	0,998	0,998	0,704	410,596	452,531
5	GC	(2,2)	<b>0,999</b>	<b>0,999</b>	<b>0,433</b>	<b>234,001</b>	<b>295,193</b>
10	GT	(1,1)	0,998	0,998	0,731	433,932	508,124
10	GT	(2,2)	<b>0,999</b>	<b>0,999</b>	0,440	253,170	362,662
10	GC (5) GT (5)	(1,1)	0,998	0,998	0,714	425,224	499,416
10	GC (5) GT (5)	(2,2)	<b>0,999</b>	<b>0,999</b>	0,440	252,785	362,277
15	GC (5) GT (10)	(1,1)	0,998	0,998	0,726	439,530	545,980
15	GC (5) GT (10)	(2,2)	<b>0,999</b>	<b>0,999</b>	0,434	258,337	416,134

Nota: en negrita se muestran los valores mínimos. Fuente: elaboración propia en base a datos de CES-BCSF y Google Inc.

En lo que respecta a las pruebas sobre la adecuación del modelo multivariado a los supuestos de partida, i.e., los contrastes de normalidad, correlación serial y autocorrelación, se encuentran en la Tabla 10.

**Tabla 10.** Estadísticos de adecuación del modelo multivariado seleccionado a los supuestos

Contraste	Hipótesis nula	Valor p	Conclusión sobre H <sub>0</sub>
Shapiro-Wilk	H <sub>0</sub> : Normalidad de los residuos	0,063	No se rechaza al 99%
Jarque-Bera	H <sub>0</sub> : Normalidad de los residuos	0,537	No se rechaza al 99%
Breusch-Godfrey	H <sub>0</sub> : No correlación serial	0,764	No se rechaza al 99%
Durbin-Watson	H <sub>0</sub> : No autocorrelación de residuos	0,449	No se rechaza al 99%
Ljung-Box	H <sub>0</sub> : No autocorrelación de residuos	0,812	No se rechaza al 99%

Fuente: elaboración propia en base a datos de CES-BCSF y Google Inc.

El modelo seleccionado genera residuos sin problemas de falta de normalidad, ya que tanto en la prueba Jarque-Bera como en el caso de Shapiro-Wilk no se puede rechazar la hipótesis nula de normalidad. De las pruebas de autocorrelación, las pruebas de Durbin-

Watson, Ljung-Box y Breusch-Godfrey se indican que no hay problemas de autocorrelación de los residuos. A su vez, para mayor detalle, los residuos del modelo multivariado, su correlograma y su distribución se encuentran expuestos de manera gráfica en el Anexo B.

En la Tabla 11 se muestran los parámetros estimados del modelo seleccionado, y su significancia. De las variables incorporadas en el modelo sólo el ICASFe resulta significativa al 99% de confianza, con ambos rezagos. Si bien el mejor modelo multivariado es el que incorpora como variables exógenas las de GC, ninguna resulta significativa incluso al nivel del 10%. Sin embargo, con motivo de comparar la capacidad predictiva de este con la del modelo univariado se decide mantener todos los términos.

**Tabla 11.** Coeficientes estimados del modelo multivariado

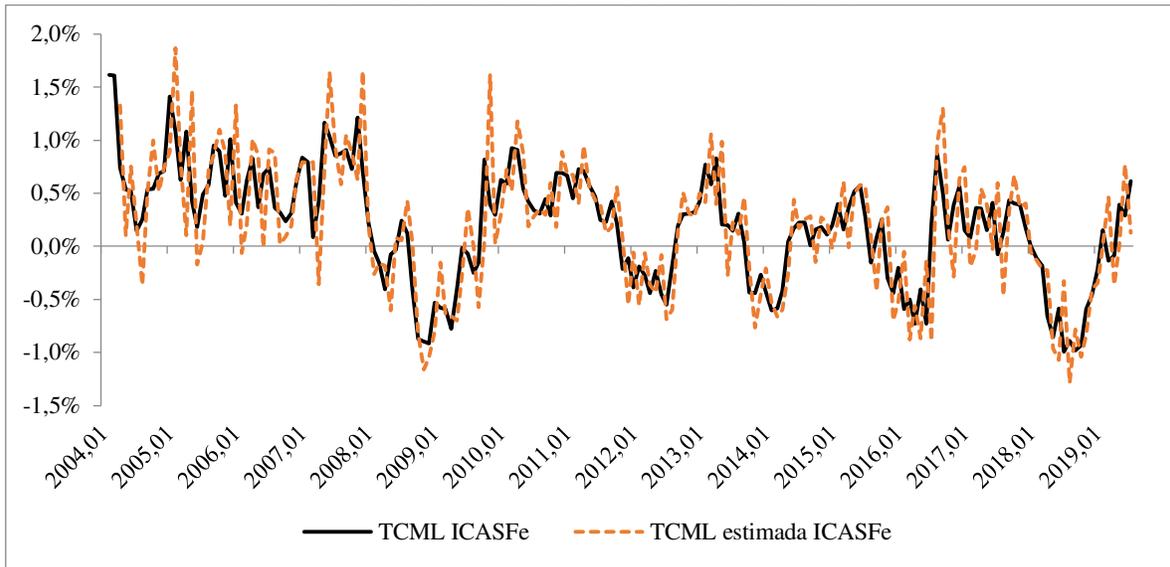
<b>Variable</b>	<b>Coefficiente</b>	<b>Desvío Estándar</b>
Intercepto	1,549*	0,697
Coopeplus <sub>t</sub>	0,003	0,002
Coopeplus <sub>t-1</sub>	0,001	0,002
Coopeplus <sub>t-2</sub>	-0,003	0,002
Vallan <sub>t</sub>	0,004	0,002
Vallan <sub>t-1</sub>	0,001	0,002
Vallan <sub>t-2</sub>	0,000	0,002
Cualquier <sub>t</sub>	0,002	0,004
Cualquier <sub>t-1</sub>	0,003	0,004
Cualquier <sub>t-2</sub>	0,000	0,004
Pasar <sub>t</sub>	0,004	0,009
Pasar <sub>t-1</sub>	0,012	0,010
Pasar <sub>t-2</sub>	-0,003	0,009
Bancopatagonia <sub>t</sub>	-0,004	0,005
Bancopatagonia <sub>t-1</sub>	-0,004	0,005
Bancopatagonia <sub>t-2</sub>	0,003	0,005
ICASFe <sub>t-1</sub>	1,767**	0,047
ICASFe <sub>t-2</sub>	-0,785**	0,046

Nota: significativo al nivel \*5%, \*\*1%

Fuente: elaboración propia en base a datos de CES-BCSF y Google Inc.

En la Figura 5 se muestra la tasa de cambio mensual logarítmica del ICASFe y el valor estimado de la misma mediante el modelo multivariado seleccionado.

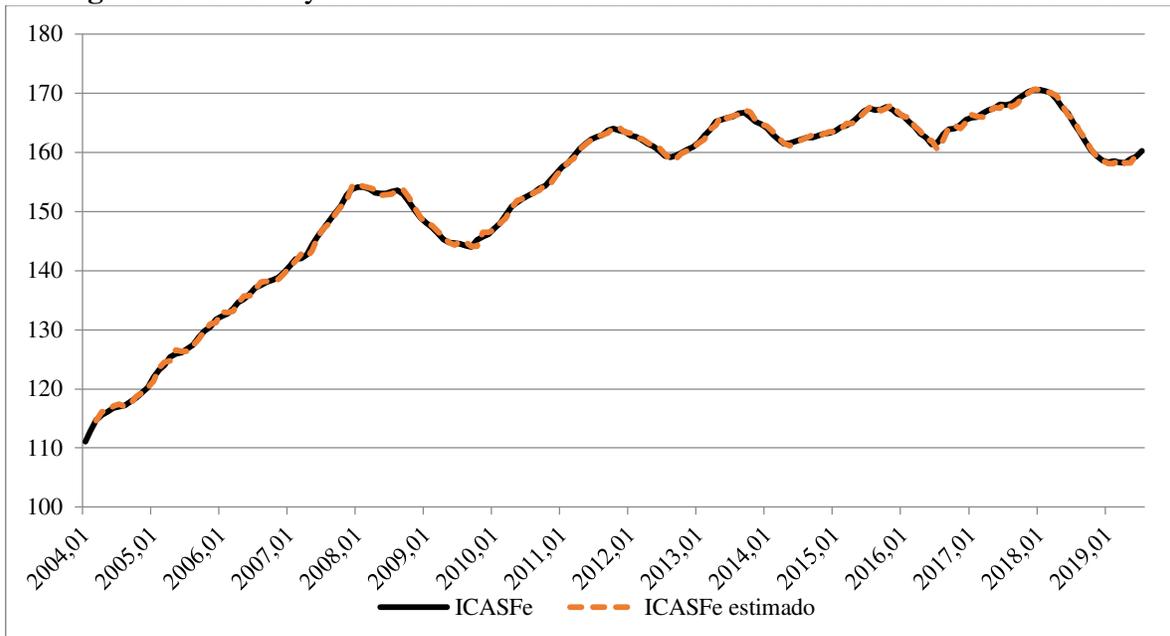
**Figura 5.** Evolución de la tasa de cambio mensual logarítmica del ICASFe y su valor estimado mediante el modelo multivariado



Fuente: elaboración propia en base a datos de CES-BCSF y Google Inc.

La evolución de la la predicción coincide, en general con la verdadera variación del índice. Del mismo modo, el ajuste entre la predicción del ICASFe en niveles y el verdadero valor del índice se expone en la figura 6, mostrando el valor de interés.

**Figura 6.** ICASFe y su valor estimado mediante modelo multivariado seleccionado



Fuente: elaboración propia en base a datos de CES-BCSF y Google Inc.

#### 4.4. Predicción a corto plazo del ICASFE con modelos seleccionados

Para cumplir con el objetivo de analizar cuál de los dos modelos produce mejores predicciones a corto plazo por fuera de la muestra de entrenamiento, se realizan predicciones para los meses de agosto a noviembre del 2019 utilizando para ellos los parámetros obtenidos anteriormente. Así, en la Tabla 12 se muestran los valores verdaderos del ICASFe, sus predicciones y los errores de estimación de ambos modelos para los meses de agosto a noviembre de 2019. Se observan que hay dos meses en los que los modelos estiman valores con diferencias (septiembre y noviembre) y, en ambos casos, el ARIMA sobreestima el verdadero valor del ICASFe, mientras que el ARDL lo subestima. El modelo multivariado muestra errores de predicción menores para septiembre, mientras que el univariado presenta errores menores en agosto, octubre y noviembre. Si bien se está prediciendo a corto plazo en ambos modelos, se podría decir que el modelo multivariado se desempeña adecuadamente en las predicciones inmediatas, mientras que el univariado va acumulando mayor error a medida que más lejos en el futuro se intenta predecir.

**Tabla 12.** Predicciones del ICASFe y error de estimación de ambos modelos

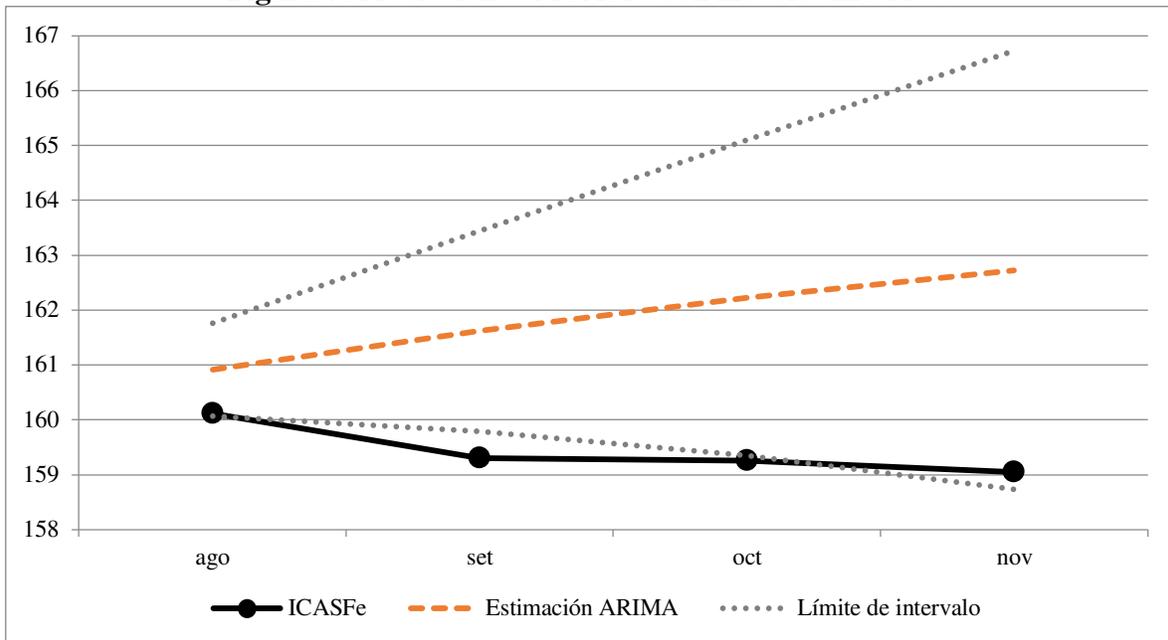
Período	Valor Real del ICASFe	Estimación de		Error estimado de	
		ARIMA (2,1,1)	ARDL (2,2)	ARIMA (2,1,1)	ARDL (2,2)
2019,08	160,211	160,916	161,024	-0,705	-0,813
2019,09	159,400	161,619	157,749	-2,219	1,651
2019,10	159,358	162,224	163,002	-2,866	-3,644
2019,11	159,146	162,727	153,397	-3,581	5,749

Nota: los valores en negrita corresponden al mínimo error de estimación de ambos modelos.

Fuente: elaboración propia en base a datos de CES-BCSF y Google Inc.

Por otra parte, el error medio de estimación del modelo ARDL es de 2,964 mientras que en el caso del ARIMA es 2,343, es decir, levemente inferior al del modelo multivariado. Esto indica que la internalización de información proveniente de las herramientas de Google no logra una mejora en la predicción fuera de la muestra respecto a la univariada.

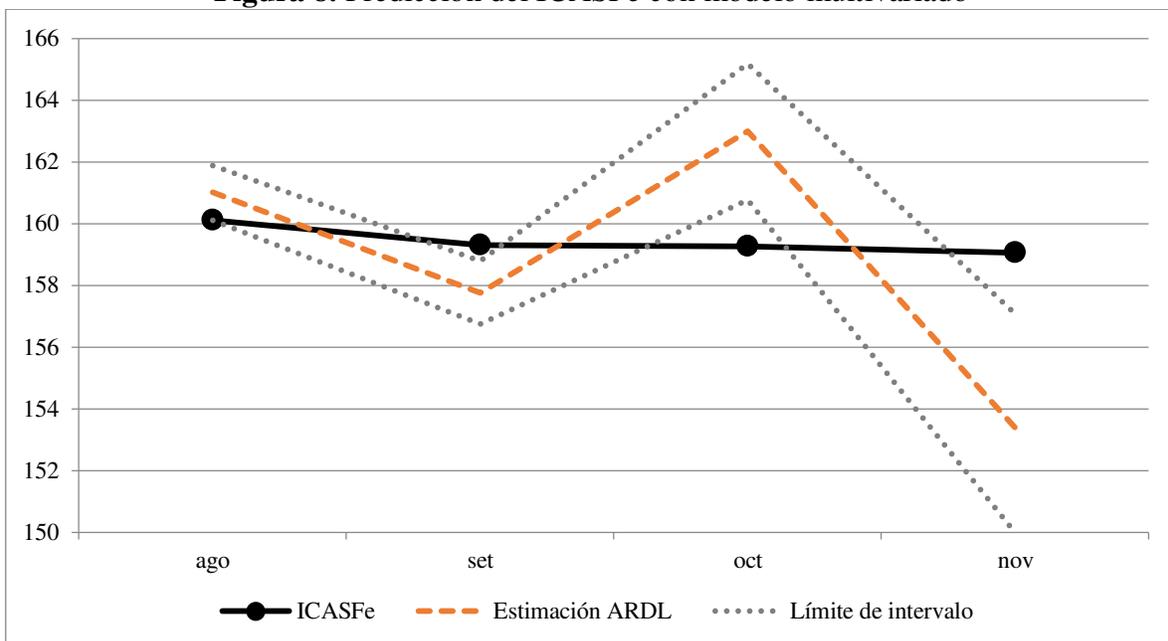
**Figura 7.** Predicción del ICASFe con modelo univariado



Fuente: elaboración propia en base a datos de CES-BCSF

En la Figura 7 se muestran las predicciones del modelo univariado con sus respectivos intervalos de predicción. Tal como se menciona, en los cuatro meses el modelo sobreestima al valor real del ICASFe. El intervalo de confianza de la estimación incluye el verdadero valor de la variable en los meses de agosto y noviembre, pero no en septiembre y octubre.

**Figura 8.** Predicción del ICASFe con modelo multivariado



Fuente: elaboración propia en base a datos de CES-BCSF y Google Inc.

Por otro lado, en la Figura 8 se muestra el valor del ICASFe y su predicción mediante el modelo ARDL(2,2), además de los intervalos de confianza para las predicciones. Los intervalos estimados no contienen al valor real de ninguno de los meses de interés predictivo de corto plazo.

Por todo lo desarrollado anteriormente, se puede concluir que en este caso el modelo multivariado no contribuye al logro de una mejor estimación en relación a la obtenida mediante el método univariado, aunque parece predecir mejor en el futuro inmediato, es decir, a uno o dos meses. Los criterios de selección de modelo (univariados o multivariados) no dan lugar a una decisión concluyente y los resultados del modelo ARIMA son levemente mejores a los del ARDL, ya que con menos requerimiento de información producen resultados muy similares y más estables.

# Capítulo 5

## Discusión y Conclusiones

### 5.1. Discusión

En el análisis de los estudios comparados relacionados, se puede mencionar que en términos generales, persiguen objetivos centrados en la predicción mediante la utilización de variables exógenas cuya información proviene de las herramientas de Google. Casi en su mayoría, logran encontrar que los resultados obtenidos en la modelación con dichas herramientas son más precisos y mejoran en las predicciones con relación a modelos univariados. Cuando se compara con los resultados del presente estudio, se observa que no se ha podido arribar a iguales conclusiones, afirmando la mejora unívoca en predicción de la variable de interés; pero tampoco resultan despreciables en el muy corto plazo las predicciones logradas. Contrariamente a lo evidenciado en la bibliografía, no se evidencian una superioridad significativa de los modelos multivariados utilizando información exógena, pero tampoco sus resultados son totalmente descartables, ya que en la caso particular, se observa que es útil para el muy corto plazo y sigue mejor la evolución de la serie.

Aun cuando los resultados del presente trabajo no han sido del todo favorables, se puede destacar que se realiza una contribución importante en comparación con la bibliografía consultada respecto de las herramientas de google utilizadas. En este sentido, la mayoría de

los antecedentes se han abocado de manera intensiva y exclusiva al *Google Trends* como herramienta, sin considerar otros buscadores, entre las que se puede mencionar *Google Advance* y *Correlate*. En este sentido, el presente trabajo avanza sobre estas herramientas alternativas utilizando la última, con el objetivo de analizando la posibilidad de incorporar dicha herramientas como fuente de datos útil tanto con propósitos de predicción como modelación de fenómenos económicos, sobre todo en épocas donde la disponibilidad de internet y facilidad de acceso a grandes bases de datos es tema de actualidad.

Las diferencias encontradas con los antecedentes en cuanto a la imposibilidad de encontrar relaciones claras y significativas entre indicadores de búsqueda y la variable de interés llevan a tratar de encontrar los motivos de dichos resultados. Se cree uno de los principales podría radicar en la inadecuada selección de variables predictivas o exógenas utilizadas en el presente estudio. Casi todos los trabajos encontrados, tanto internacionales (e.g. Artola y Galán (2012), Bortoli y Combes (2015), Park et al. (2016), entre otros) y nacionales (Blanco, 2014), parten de la premisa de una existencia causal entre las búsquedas de determinadas palabras en Internet y la toma de decisiones que afecta de manera directa o indirecta a la economía real, es decir, se sustentan en un modelo económico para realizar las búsquedas y, por tanto, seleccionar las variables exógenas.

Sin embargo, como se menciona en la metodología, en el presente estudio se opta por utilizar indicadores de búsqueda de los términos con mayor coeficiente de correlación respecto a la evolución de la variable a predecir; no se recurre a un criterio basado en el significado de los términos, ni tampoco las búsquedas relacionadas con los indicadores que alimentan al ICASFe, tal como lo realizan en el resto de los trabajos analizados en el estado del arte. En la relación del ICASFe y las búsquedas no se presume *a priori* la existencia de relaciones causales, sino que provienen de un proceso introspectivo basado en nivel de correlación. Esta selección *ad hoc* de las variables podría ser el principal motivo de la falta de significancia de las mismas en la modelación. Sin embargo, cabe mencionar que la justificación de haber realizado tal procedimiento de selección radica en que siendo la variable de interés una construcción regional, no se tenía *a priori* intuición sobre los términos de búsqueda que podrían estar relacionados con la misma.

Pero además de la forma en que se seleccionan las variables exógenas, otro de los puntos clave es el diseño del modelo donde se relacionan éstas con la variable de interés. En los trabajos de referencia analizados, diversos autores han comparado predicciones obtenidas mediante modelos tanto univariadas como multivariadas, al igual que se propone en el presente estudio. Sin embargo, los modelos multivariados mayoritariamente utilizados en la bibliografía se caracterizan por ser de tipo endógenos, es decir, sin asumir exogeneidad entre las búsquedas y la variable de interés predictivo o, si la asumen, realizan pruebas de exogeneidad que permiten concluir dicho sentido de la relación. En cambio, aquí no se realizan este tipo de análisis con lo cual queda abierta la posibilidad que la dirección de la relación sea la inversa a la especificada, y ello podría ser otra fuente de diferencia con la bibliografía analizada.

## **5.2. Conclusiones**

Disponer de información oportuna tiene una gran importancia en el ámbito económico, ya que permite disminuir la incertidumbre a la que se enfrenta cualquier agente a la hora de tomar decisiones respecto a la utilización y administración de recursos escasos. Muchos organismos de tipo públicos y privados realizan publicaciones con el objetivo de poner a disposición datos e información para ello. Pero, sin embargo, la información necesaria para ello y las publicaciones suelen estar rezagada en el tiempo, es decir, se publican datos que corresponden al pasado, muchas veces un pasado bastante lejano. En este contexto, el avance tecnológico ha permitido avanzar en la disposición de un mayor caudal de información hacia los usuarios lográndose mejoras en la oportunidad y la posibilidad de reducir la imprevisibilidad sobre acontecimientos futuros.

De los antecedentes mencionados a lo largo de este trabajo, se pudo observar que refuerzan la idea de que las herramientas informáticas contribuyen a hacer frente a los problemas de falta de certidumbre, poniendo a disposición mecanismos robustos para estimar cuestiones del futuro. Es así que, luego de analizar diferentes estudios que intentan aprovechar la nueva tecnología y *big data*, se observa que los resultados en la mayoría de ellos concluyen sobre la recomendación de utilizar información proveniente de motores de búsqueda, dado que esta permite lograr estimaciones económicas precisas. Destacan el inherente potencial que poseen dichas herramientas informáticas como fuentes de información, partiendo de la

hipótesis inicial de que las personas consultan cada vez en mayor medida estas herramientas y que las mismas pueden reflejar ciertos patrones sociales y económicos. Algunos ejemplos de áreas bajo estudio son las variables referidas al sector turístico, consumo minorista, consumo de bienes durables, entre otros.

El ICASFe se presenta como un indicador publicado por un organismo que informa sobre la evolución del nivel de actividad de la provincia de Santa Fe y permite que los agentes puedan contar con información local para poder reducir la incertidumbre en el proceso de toma de decisiones. Es un índice que sintetiza información proveniente de más de una serie económica que posee la ventaja de ser de fácil de interpretación para quienes lo consultan, ya sean agentes públicos o privados. Sin embargo, la publicación de dicho indicador tiene un rezago de dos a tres meses debido a que, en su elaboración se utilizan datos de varias series económicas que están disponibles con retrasos. Por tanto, contar con una predicción y estimación de su valor lo más actual en el tiempo y próximo a su verdadero valor contribuye positivamente al proceso de formación de expectativas a partir de información confiable, y da lugar a la consecuente mejora en términos de eficiencia en el proceso de toma de decisiones.

La importancia económica del indicador, su retraso en la disponibilidad y la posibilidad de utilizar herramientas gratuitas de los buscadores para predicción fueron los motivos que impulsaron a la presente tesina. Con el objetivo general de evaluar el potencial predictivo de la información de Google para estimar de manera oportuna el valor del índice de actividad provincial se procedió, en primer lugar, a determinar las palabras claves que poseían un patrón de búsquedas con mayor correlación respecto al ICASFe en el período enero 2004 a julio 2019, obteniéndose medidas de búsqueda o indicadores que informan sobre la actividad en Google correlacionada a la evolución del índice. Se diseñaron dos modelos para la predicción, uno univariado y otro multivariado, cuyos parámetros fueron estimados por máxima verosimilitud. El primero, un modelo autorregresivo y el segundo, una combinación de un proceso autorregresivo y variables exógenas obtenidos de las herramientas de búsqueda. Estimados los parámetros se realizaron predicciones con ambos modelos para un período no incluido en la muestra, a saber, de agosto a noviembre del 2019 y se compararon con los valores reales de la variable de interés.

Contrariamente a lo esperado en la hipótesis inicial, los resultados obtenidos no permiten afirmar de manera concluyente que las herramientas de Google sean una alternativa que mejora la posibilidad de obtener un estimador confiable del ICASFe respecto a los modelos univariados usualmente utilizados, ya que no resultaron significativos los índices de búsqueda tenidos en cuenta. Se ha encontrado que el error medio de predicción del modelo univariado es menor al del multivariado, sin embargo las diferencias no son significativas. Lo que se puede destacar es que las predicciones del modelo multivariado son más estables, en el sentido que no acumula errores. El modelo univariado genera predicciones que siempre sobrestiman el verdadero valor del índice mientras que el multivariado oscila alrededor del verdadero valor, estando más cerca en las predicciones inmediatamente posteriores. Esto lo hace más útil para predicciones a muy corto plazo.

Una limitación a destacar es la falta de evidencia empírica suficiente para recomendar la utilización de modelos multivariados que internalizan información de Google. Esto va en contraposición con los trabajos presentados en los antecedentes cuyos resultados son concluyentes respecto a dicha recomendación. Tal como se menciona en la discusión, se cree que las razones de ello están en diversos factores, entre ellos, la consideración de los términos a incluir como variables explicativas y la forma en la que se internaliza y modela dicha información.

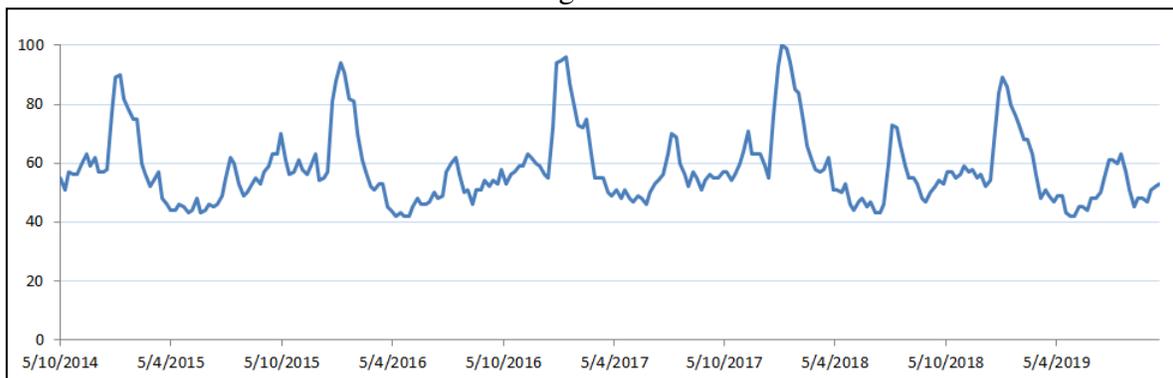
Por lo tanto, de las limitaciones encontradas en el presente estudio surgen las propuestas a tener en cuenta como futuras líneas de trabajo. Entre ellas se menciona la posibilidad de evaluar otras herramientas disponibles gratuitamente, como *Google Analytics*, y si éstas pueden brindar información útil para la modelación. De igual manera, es necesario mejorar y analizar otros caminos para selección de las variables a considerar, como seleccionar dentro de la lista de términos con mayor correlación aquellos que posean un significado económico comprobable; o utilizar todos los indicadores disponibles y realizar reducción de dimensiones a partir, por ejemplo de análisis de componentes principales. A su vez, también es necesario analizar otros modelos predictivos, a saber, aquellos endógenos como vectores autorregresivos, modelos de vectores de corrección de errores o cointegración, entre otros.

# Anexo A

## Ejemplo de Búsqueda en las Herramientas de Google

En el presente anexo se presenta el procedimiento realizado por cada una de las palabras que resultaron tener una correlación alta con la variable de interés predictivo. Así, por ejemplo, tomando la palabra “hotel” y colocándola en *Google Trends*, se puede observar la evolución de las búsquedas de este término a modo de índice. En la Figura A1 se muestra la evolución de las búsquedas de la palabra “hotel” en Argentina. Dicha evolución tiene una periodicidad mensual y corresponde al período desde septiembre de 2014 a septiembre de 2019. Los datos de búsqueda corresponden al total nacional para Argentina.

**Figura A1.** Evolución de las búsquedas semanales en Google de la palabra “hotel” en Argentina

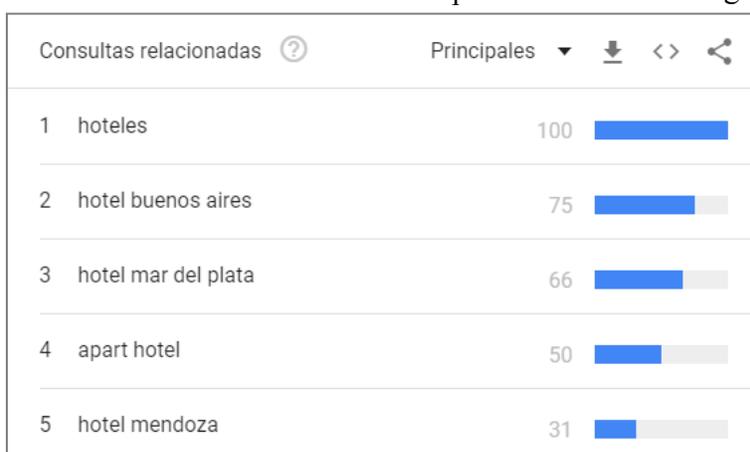


Fuente: elaboración propia en base a datos de *Google Trends*

Se puede ver que el valor máximo del índice se ubica a comienzos del año 2018, y que existe un comportamiento a simple vista estacional.

Pero además del índice de búsqueda, también te informa sobre “otras consultas” que realiza el usuario que buscó “hotel”. En la Figura A2 se muestra el cuadro de “consultas relacionadas”, donde muestra búsquedas que históricamente (en el período de referencia) se relacionaron con la palabra clave, o bien, aquellas que poseen una tendencia de búsqueda en aumento con la palabra “hotel”.

**Figura A2.** Consultas relacionadas a la palabra “hotel” en Argentina



Fuente: elaboración propia en base a datos de *Google Trends*.

En este caso particular del ejemplo, seleccionando las consultas principales, se obtuvieron 25 términos relacionados, siendo los cinco primeros los que se presentan en la Figura A2.

Este procedimiento se realiza para aquellas palabras claves que aparecen asociadas a ICASFe.

# Anexo B

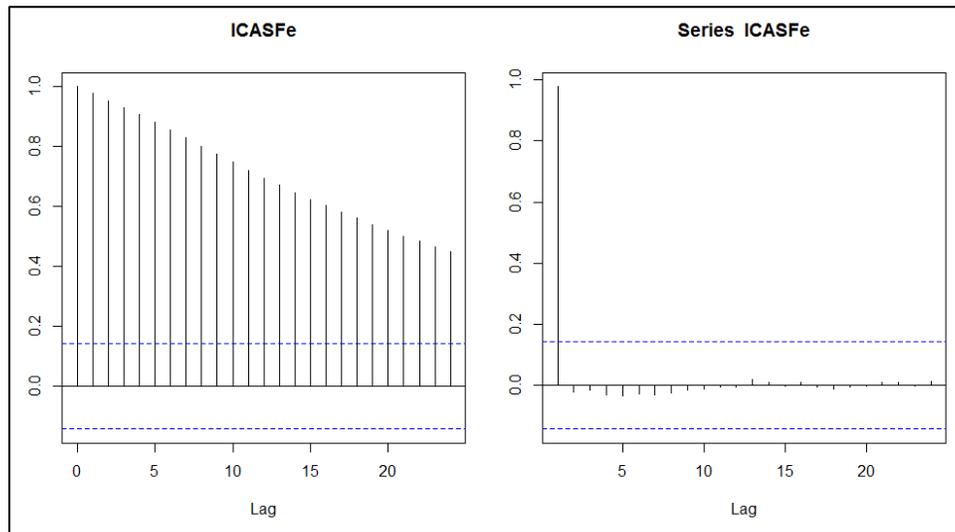
## Contrastes y Estimaciones de los modelos

En este apartado se muestran las diferentes salidas de los programas utilizados tanto en las pruebas de estacionariedad, como en las estimaciones y los análisis de los supuestos de los modelos seleccionados.

### **B1. Funciones de autocorrelación y de autocorrelación parcial del ICASFe**

La Figura B1 muestra la función de autocorrelación simple y de autocorrelación parcial del ICASFe para el período 2004-2019. Se observa que la autocorrelación simple es significativa en todos los retardos expuestos, mientras que la autocorrelación parcial lo es en el primer retardo. Esto indica que el efecto del ICASFe en  $t - 2$  sobre la serie en el momento  $t$  se traduce siempre a través del valor en  $t - 1$ .

**Figura B1.** Función de autocorrelación y función de autocorrelación parcial del ICASFe



Fuente: elaboración propia en base a datos del CES-BCSF

## B2. Pruebas de raíces unitarias

En la Figura B2 se exponen los resultados del contraste de Dickey-Fuller Aumentado aplicado a la serie del ICASFe en nivel para el análisis de estacionariedad de la misma. Se muestra los resultados de la prueba para dos modelos, con y sin tendencia. Para cada modelo, la cantidad de rezagos es seleccionada a partir del AIC, para lo cual en el primer caso es un rezago y en el segundo, tres. A partir de la observación del *valor p* asintótico, ambas modelaciones muestran que existe raíz unitaria en la serie ICASFe a nivel, lo que implica que no es estacionaria en nivel.

**Figura B2.** Resultados del Aumentado de Dickey-Fuller a la serie del ICASFe en nivel

```
Contraste aumentado de Dickey-Fuller para ICASFe
Contrastar hacia abajo desde 14 retardos, con el criterio AIC
Tamaño muestral 185
la hipótesis nula de raíz unitaria es: [a = 1]

contraste con constante
incluyendo un retardo de (1-L)ICASFe
modelo: (1-L)y = b0 + (a-1)*y(-1) + ... + e
valor estimado de (a - 1): -0,00540588
estadístico de contraste: tau c(1) = -2,33768
Valor p asintótico 0,1601
Coef. de autocorrelación de primer orden de e: 0,014

con constante y tendencia
incluyendo 3 retardos de (1-L)ICASFe
modelo: (1-L)y = b0 + b1*t + (a-1)*y(-1) + ... + e
valor estimado de (a - 1): -0,0116682
estadístico de contraste: tau ct(1) = -2,64367
Valor p asintótico 0,2607
Coef. de autocorrelación de primer orden de e: 0,002
diferencias retardadas: F(3, 177) = 105,219 [0,0000]
```

Fuente: elaboración propia en base a datos del CES-BCSF

Si la serie en nivel no es estacionaria, debe encontrarse el orden de integración, y para ello, el primer paso es analizarla en primera diferencia. Así, en la Figura B3 se muestran los resultados de la aplicación del mismo contraste a la serie compuesta por la TCML del ICASFe, que es el ICASFe en primeras diferencias. Ambos modelos con AIC más bajo no incluyen retardos, independientemente de si tiene o no tendencia.

**Figura B3.** Resultados del contraste Aumentado de Dickey-Fuller a la tasa de cambio de la serie del ICASFe

```
Contraste aumentado de Dickey-Fuller para d_ICASFe
Contrastar hacia abajo desde 14 retardos, con el criterio AIC
Tamaño muestral 185
la hipótesis nula de raíz unitaria es: [a = 1]

contraste con constante
incluyendo 0 retardos de (1-L)d_ICASFe
modelo: (1-L)y = b0 + (a-1)*y(-1) + e
valor estimado de (a - 1): -0,180958
estadístico de contraste: tau c(1) = -4,37244
valor p 0,0004567
Coef. de autocorrelación de primer orden de e: -0,001

con constante y tendencia
incluyendo 0 retardos de (1-L)d_ICASFe
modelo: (1-L)y = b0 + b1*t + (a-1)*y(-1) + e
valor estimado de (a - 1): -0,207824
estadístico de contraste: tau ct(1) = -4,54322
valor p 0,00169
Coef. de autocorrelación de primer orden de e: 0,014
```

Fuente: elaboración propia en base a datos del CES-BCSF

En este caso, se puede observar que la serie en diferencia es estacionaria, lo que implica que el ICASFe es I(1). Se puede ver que los valores p permiten rechazar de la hipótesis nula de presencia de raíz unitaria al nivel 1%.

### B3. Estimación del modelo univariado

Luego de analizar de manera *ad hoc* diferentes diseños del modelo univariado, el AIC permite seleccionar el modelo con mejor ajuste, el cual es el ARIMA(2,1,1). Los resultados de dicha estimación se muestran en la Figura B4 correspondiente a la salida del programa Gretl.

**Figura B4.** Resultados del modelo univariado seleccionado.

Evaluaciones de la función: 111					
Evaluaciones del gradiente: 35					
Modelo 2: ARIMA, usando las observaciones 2004:02-2019:07 (T = 186)					
Estimado usando AS 197 (MV exacta)					
Variable dependiente: (1-L) ICASFe					
Desviaciones típicas basadas en el Hessiano					
	Coefficiente	Desv. típica	z	valor p	
-----	-----	-----	-----	-----	-----
const	0,315113	0,182057	1,731	0,0835	*
phi_1	-0,104040	0,0482764	-2,155	0,0312	**
phi_2	0,767001	0,0485185	15,81	2,72e-056	***
theta_1	1,00000	0,0147984	67,57	0,0000	***
Media de la vble. dep.	0,261646	D.T. de la vble. dep.	0,774501		
Media de innovaciones	-0,009409	D.T. innovaciones	0,427519		
Log-verosimilitud	-107,4973	Criterio de Akaike	224,9946		
Criterio de Schwarz	241,1234	Crit. de Hannan-Quinn	231,5306		

Fuente: elaboración propia en base a datos del CES-BCSF

### B4. Estimación del modelo multivariado y predicciones

Del mismo modo que con el modelo univariado, en la Figura B5 se muestra la salida correspondiente a la estimación del modelo multivariado seleccionado. Sin embargo, el programa utilizado en esta oportunidad es el R por medio del paquete 'dLagM' (Haydar, 2020).

**Figura B5.** Resultados del modelo multivariado seleccionado

```

ICASFe ~ coopeplus + cualquier + pasar + vallan + bancopatagonia
Coefficients:
      Estimate Std. Error t value Pr(>|t|)
(Intercept)  1.5492731  0.6974691  2.221  0.0277 *
coopeplus.t  0.0027639  0.0022068  1.252  0.2122
coopeplus.1  0.0014609  0.0021268  0.687  0.4931
coopeplus.2 -0.0031781  0.0022133 -1.436  0.1529
cualquier.t  0.0021845  0.0043294  0.505  0.6145
cualquier.1  0.0025089  0.0043051  0.583  0.5608
cualquier.2 -0.0002830  0.0041582 -0.068  0.9458
pasar.t      0.0041742  0.0092702  0.450  0.6531
pasar.1     0.0118527  0.0099251  1.194  0.2341
pasar.2    -0.0032266  0.0091180 -0.354  0.7239
vallan.t    0.0037486  0.0024902  1.505  0.1341
vallan.1    0.0007754  0.0024830  0.312  0.7552
vallan.2    0.0000657  0.0023509  0.028  0.9777
bancopatagonia.t -0.0041850  0.0050020 -0.837  0.4040
bancopatagonia.1 -0.0039855  0.0054979 -0.725  0.4695
bancopatagonia.2  0.0029129  0.0047708  0.611  0.5423
ICASFe.1    1.7672749  0.0470554  37.557 <2e-16 ***
ICASFe.2   -0.7848654  0.0461975 -16.989 <2e-16 ***
---
Signif. codes:  0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

Residual standard error: 0.4317 on 167 degrees of freedom
Multiple R-squared:  0.9992,    Adjusted R-squared:  0.9991
F-statistic: 1.232e+04 on 17 and 167 DF,  p-value: < 2.2e-16
    
```

Fuente: elaboración propia en base a datos del CES-BCSF y Google Inc.

Una vez obtenido el modelo predictivo se realizan las estimaciones de los valores fuera de la muestra. Los resultados de las predicciones del modelo multivariado, como así también sus intervalos de confianza para la predicción al 95% se exponen en la Figura B6.

**Figura B6.** Predicciones del ICASFe con el modelo multivariado.

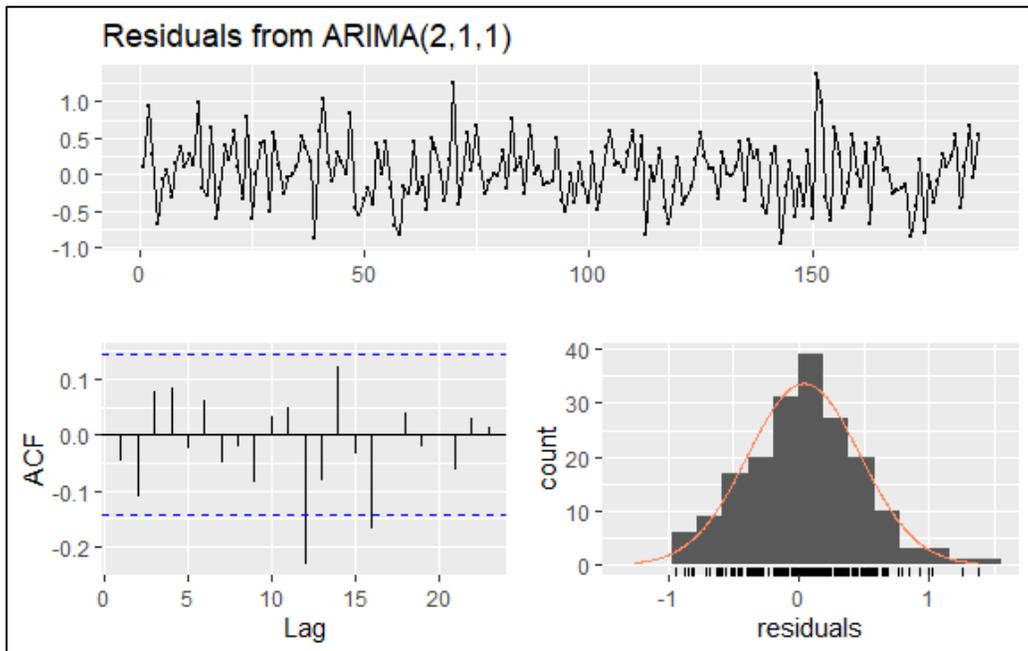
\$forecasts			
	95% LB	Forecast	95% UB
1	160.1949	161.0236	161.8779
2	156.6443	157.7488	158.7828
3	160.8773	163.0016	165.3208
4	149.5617	153.3970	156.6761

Fuente: elaboración propia en base a datos del CES-BCSF y Google Inc.

## B5. Análisis de los residuos

La figura B7 muestra los residuos del modelo ARIMA(2,1,1), su correlograma y distribución. Como se puede observar, no se evidencia correlación significativa antes del orden 12, y la distribución es aproximadamente simétrica y acampanada, dando sustento a los resultados obtenidos a través de las pruebas correspondientes.

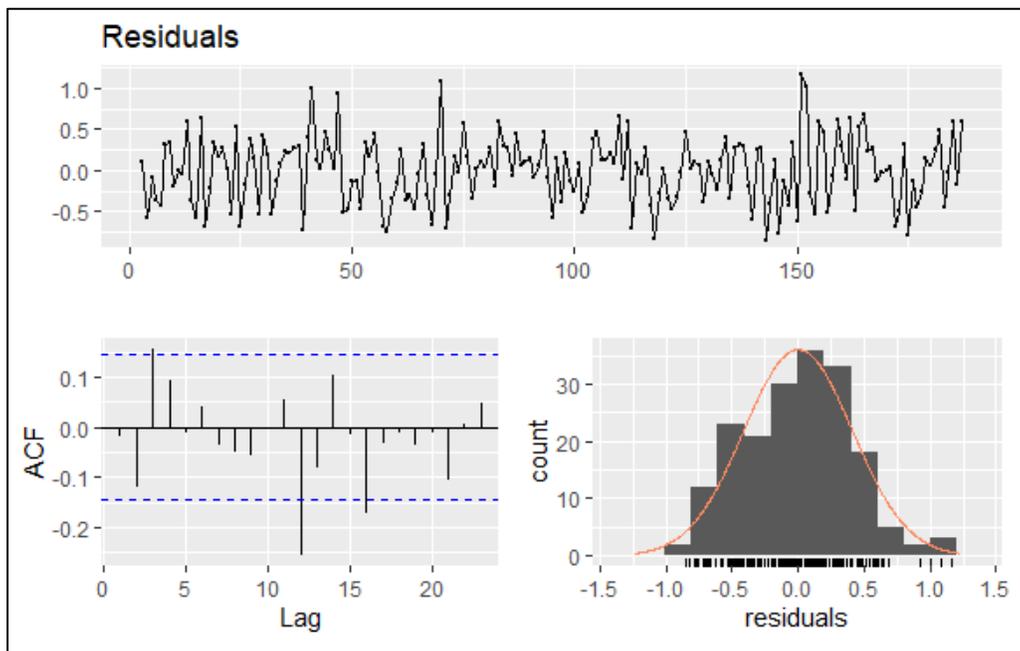
**Figura B7.** Correlograma y distribución de los residuos del modelo univariado



Fuente: elaboración propia en base a datos del CES-BCSF

De manera similar al caso del modelo univariado, se estima la evolución, correlación y distribución de los residuos del modelo multivariado ARDL(2,2). Las representaciones gráficas se muestran en la figura B8.

**Figura B8.** Correlograma y distribución de los residuos del modelo multivariado



Fuente: elaboración propia en base a datos del CES-BCSF y Google Inc.

La distribución de los residuos del modelo multivariado es aproximadamente simétrica y acampanada. Por su parte, la autocorrelación no es significativa antes del retardo 12.

# Bibliografía

- Akerlof, G. A. (1970). The market for "lemons": quality uncertainty and the market mechanism. *The Quarterly Journal of Economics*, Vol. 84, No. 3, 488-500.
- Artola, C., & Galán, E. (2012). *Tracking the future on the web: construction of leading indicators using Internet searches*. Madrid, España: Banco de España.
- Askitas, N., & Zimmermann, K. F. (2009). *Google econometrics and unemployment forecasting*. Bonn, Germany: IZA.
- Bartlett, M. S. (1946). On the theoretical specification of sampling properties of autocorrelated time series. *Journal of the Royal Statistical Society, serie B*, vol. 27, 27-41.
- Blanco, E. (2014). *Herramientas de big data: ¿podemos aprovechar Google Trends para pronosticar algunas variables macro relevantes?* Asociación Argentina de Economía Política (AAEP).
- Bortoli, C., & Combes, S. (2015). *Contribution from Google Trends for forecasting the short-term economic outlook in France: limited avenues*.
- Box, G. P., & Pierce, D. A. (1970). Distribution of residual autocorrelations in autoregressive integrated moving. *Journal of the American Statistical Association*, vol. 65, 1509-1526.
- Carrière-Swallow, Y., & Labbé, F. (2011). *Nowcasting with Google Trends in an emerging market*. Santiago, Chile: Journal of Forecasting.
- Choi, H., & Varian, H. (2009). *Predicting the present with Google Trends*. Google Inc.
- Choi, S., Jun, S.-P., & Yoo, H. 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.

- D'Jorge, M. L., Cohan, P. P., Lazzaroni, M., Cherri, A., Camusso, F., & Zanini, L. (2018). *14 Sub-indicadores considerados por el Índice Compuesto Coincidente de Actividad Económica de la Provincia de Santa Fe (ICASFe)*. Santa Fe, Argentina: Centro de Estudios y Servicios - Bolsa de Comercio de Santa Fe.
- Dergiades, T., Mavragani, E., & Pan, B. (2018). Google Trends and tourists' 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*, *vol. 74*, 427-431.
- Fuller, W. A. (1976). *Introduction to statistical time series*. Nueva York: John Wiley & Sons.
- GRET. Gnu Regression, Econometrics and Time-series Library. Software package for econometric analysis, written in the C programming language. Latest stable version 2018b (August 12, 2018)
- Gujarati, D. N., & Porter, D. C. (2009). *Econometría - 5ta Edición*. México: McGraw Hill.
- Haucap, J., & Heimeshoff, U. (2013). *Google, Facebook, Amazon, eBay: Is the Internet driving competition or market monopoly?* Düsseldorf, Germany: Düsseldorf Institute for Competition Economics (DICE).
- Haydar, D. (2020). dLagM: time series regression models with distributed lag models. R package, version 1.0.21.
- Heimer, R. Z., Stehulak, T., & Killiner, D. (2015). *Assessing consumer confidence with*. Cleveland, USA: Federal Reserve Bank of Cleveland.
- Mohebbi, M., Vanderkam, D., Kodysh, J., Schonberger, R., Choi, H., & Kumar, S. (2011). *Google Correlate whitepaper*.
- Morán, J. (2016). *Google Trends: una nueva herramienta para la predicción económica. Aplicaciones para complementar el análisis econométrico tradicional*. Buenos Aires, Argentina: Universidad de San Andrés.
- 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.
- R Core Team (2020). R: A language and environment for statistical computing. R Foundation for Statistical Computing. Viena, Austria. URL: <https://www.r-project.org>.
- Santana Jimenez, L. J. (2017). *Can Google Trends data help to improve the nowcasting and short-term forecasting of the arrivals of tourists to the Dominican Republic?*

- Schmidt, T., & Vosen, S. (2009). *Forecasting private consumption: survey-based indicators vs. Google Trends*. Bochum, Germany: Ruhr-Universität Bochum (RUB).
- Schwert, G. W. (1989). Tests for unit roots: A Monte Carlo investigation. *Journal of Business and Economic Statistics* 7(2): 5-17.
- Spence, M. (1974). *Market signaling*. Cambridge MA: Harvard University Press.
- Stiglitz, J. E., & Weiss, A. (1981). Credit rationing in markets with imperfect information. *The American economic review*, 71(3), 393-410.
- Telecomunicaciones, U. I. (2018). *UIT: Comprometida para conectar el mundo*. Retrieved from <https://www.itu.int>
- Varian, H. (2010). *Microeconomía Intermedia. Un Enfoque Actual. 8va Edición*. España: Antoni Bosch Editor.
- Vilaseca, J., Torrent, J., & Lladós, J. (2001). De la economía de la información a la economía del conocimiento: algunas consideraciones conceptuales y distintivas. *Revista Tendencias. Vol. II. Nº 2*, 45-63.