



UNIVERSIDAD JUÁREZ DEL ESTADO DE
DURANGO

FACULTAD DE CIENCIAS EXACTAS

Maestría en Estadística Aplicada

Pronóstico del índice de compras al
por menor del comercio en México a
través de las tendencias de búsqueda
de Google

TESIS

Presenta

Omar David Luna González

Asesor

Dr. Mario Alberto García Meza

Victoria de Durango, Durango

Agosto 2021

Índice general

Resumen	1
Introduccion	2
1 Planteamiento del problema	4
1.1 Antecedentes	4
1.2 Objetivo General	6
1.2.1 Objetivos Particulares	6
1.3 Justificación	7
1.4 Alcances y limitaciones	7
2 Marco Teórico	8
2.1 Definiciones	8
3 Metodología	16
3.1 Tipología del estudio	16
3.2 Método de investigación	16
3.3 Hipótesis	16
3.4 Técnicas y procedimientos para recopilar información	17
3.4.1 Recopilación de los datos	17
3.5 Técnicas de análisis de información	17
4 Resultados	18
4.0.1 El modelo de pronóstico	18
4.0.2 Pronóstico del índice de compras al por menor	22
4.1 Análisis de los resultados	27

Resumen

El Instituto Nacional de Estadística y Geografía de México (INEGI) publica el índice de compras al por menor del comercio en México, junto con otras variables macroeconómicas importantes, como el índice de consumo general, índice de confianza del consumidor, entre otros, a través de su sección de Banco de Información Económica con un rezago de la publicación de la información de 3 meses. Este rezago se puede reducir usando información que está disponible para crear una estimación con series de tiempo y obtener la información al día. Para esto, se usa un modelo de series de tiempo de espacio de estados, que permite utilizar un gran conjunto de regresores para obtener una proyección del valor actual del índice de compras al por menor para el comercio en México. Este conjunto de regresores es conformado por las tendencias de búsquedas de Google.

Es importante resaltar que el uso de las consultas de búsqueda mejora significativamente la precisión de la proyección.

Introducción

El Instituto Nacional de Estadística y Geografía de México (INEGI) se encarga de publicar mensualmente el índice de compras al por menor para el comercio en México.

Este indicador es de suma importancia para los economistas, investigadores, gobiernos locales y algunas empresas para la toma de decisiones como saber, en qué momento una empresa debe invertir, dónde está situada en relación al ciclo económico y buscar otras alternativas en caso necesario.

La mayoría de los indicadores que se publican, a menudo se hacen con varios meses de atraso. En el caso de compras al por menor, la información más reciente está disponible con tres meses de rezago.

La existencia de nuevas fuentes de información significa que hay espacio para mejorar la oportunidad en que se publican los indicadores que pueden ser utilizados para la toma de decisiones de política macroeconómica.

Una posible fuente de datos que puede ser útil para emitir el índice de compras al por menor se puede encontrar en Google, los datos de consultas de búsqueda, disponibles en las herramientas Google Correlate.

Es razonable afirmar que una cantidad suficiente de artículos de consumo implica una actividad de búsqueda en Internet de tal manera que sea posible inferir la actividad de compras al por menor del presente o futuro a través de las búsquedas de Google.

El razonamiento es que algunas consultas importantes están correlacionadas con el índice de compras al por menor. Por ejemplo, las búsquedas de los productos sacos y celular de Liverpool que están correlacionadas con este indicador sugiere que las ventas de celulares por internet pudieran ser un indicador para el índice de compras.

Las ventas de celulares por internet, y ventas en general por internet son cada vez más presentes en la actividad económica de nuestro país, por ello es natural ver este tipo de búsquedas con alta correlación con respecto al índice de compras al por menor.

En esta investigación se trabajó un modelo de serie temporal de espacios de estado para pronosticar el índice de compras al por menor utilizando las tendencias de búsquedas de Google como variables regresoras. Las compras al por menor son un componente importante del índice de consumo, que a su vez es un importante elemento para el Producto Interno Bruto en el diario de la ciencia económica aplicada.

La difusión inmediata es cada vez más importante para la decisión política de los bancos centrales, tener la información al día de los indicadores económicos y su manera de pronós-

tico se convierte cada vez más en una necesidad, y se puede aprovechar toda la información que podemos obtener a través del uso del internet y las nuevas tecnologías, ya que vivimos en la era de los datos.

Capítulo 1

Planteamiento del problema

El INEGI recaba datos estadísticos en el área económica para generar índices que ayudan a comprender la situación económica en México, algunos de estos índices son: el Producto Interno Bruto (PIB), el Índice de Confianza del Consumidor (ICC), el Índice de Consumo (IC), entre otros. Es de gran beneficio para los expertos en el área económica, los dueños de las empresas, tener los datos al día, el problema es que se presentan con un rezago de tres meses, además de que en situaciones atípicas no se pueda recolectar, es importante no dejar de publicar la información de interés nacional.

1.1. Antecedentes

Vivimos en la era en que las computadoras están en medio de cualquier transacción, en la economía no es la excepción, al momento de ir al cajero a retirar efectivo, depositar, hacer una compra en tiendas físicas o en línea, se genera gran cantidad de información que se puede almacenar en bases de datos para su futuro análisis.

El inconveniente con estas bases de datos es su gran dimensión, por lo cual es necesario hacer reducción de los datos como método opcional para su correcto análisis (VARIAN, 2010). Una de las formas para representar la de reducción de datos es Google Trends, es un sistema que produce un índice de las principales búsquedas de los usuarios de Google. Otra opción es Google Correlate, es un sistema que genera un índice de búsquedas de Google correlacionado con una serie de tiempo ingresada por el usuario. Hay muchos usos para esta información, Choi y Varian (VARIAN, 2010) utilizaron estos datos para pronosticar indicadores macroeconómicos, este tipo de pronósticos ha sido de gran importancia para los bancos centrales.

La metodología de recolección de información de México está diseñada para coincidir

con las metodologías de organismos como las Naciones Unidas, EUROSTAT, la Organización de Cooperación Económica y Desarrollo (OCDE), y el Fondo Monetario Internacional (FMI) (INEGI, 2013). Esto significa que el PIB y sus componentes siguen un marco teórico estándar que ayuda a comparar una serie con otra, entonces un índice de compras al por menor puede compararse dentro de países con un grado de confianza razonable entre ambos, es natural tener escepticismo sobre la comparación de los índices, no obstante los índices y sus componentes son claramente un indicador confiable del estado de la economía.

La información de este índice es recopilada por el INEGI con datos de diferentes encuestas aplicadas a empresas en la industria manufacturera, servicios, y registros de la balanza comercial y el índice Nacional de Precios (INEGI, 2013). Estos datos se complementan con información de las asociaciones de productores, y estadísticas derivadas de instituciones como la compañía petrolera PEMEX, la Asociación Mexicana de la Industria Automotriz, y la oficina de estadísticas laborales de los Estados Unidos de América para las importaciones. Dado que las fuentes pueden diferir en los lanzamientos de tiempo, utilizan el método de Denton (DENTON, 1971) para ajustar los datos de en una publicación mensual. Desde que se hace el lanzamiento de los datos de INEGI, la información mensual tiene un retraso de tres meses, pero los datos de búsqueda de Google están disponibles con pocos días de diferencia, la creación de un pronóstico temprano del indicador puede resultar valiosa para los tomadores de decisiones. Las personas realizan consultas a diario sobre una amplia variedad de temas y el volumen de búsqueda varía en tiempo según la actividad económica y cambios de interés, por lo que se puede hacer inferencia sobre estas consultas que revelan más de lo que las encuestas y otros métodos para recopilar datos pueden decir. Los usuarios del buscador Google escriben una determinada consulta, por ejemplo "paquetería DHL", el volumen de consulta es registrado y se puede observar como una serie de tiempo de la popularidad de la consulta a lo largo del tiempo, la escala es de tal manera que el máximo de cualquier serie dada es 100, y el resultado puede ser restringido a un periodo determinado de tiempo para cualquiera de los países en donde se utiliza el servicio. Los datos pueden ser filtrados también por categorías de búsqueda (crédito, esparcimiento, instrumentos musicales, etcétera) para no confundir resultados similares, por ejemplo, la palabra "jaguar" se puede referir al animal o al auto, y por el servicio de Google (imágenes, noticias, videos de YouTube o el servicio de compras de Google), los resultados también se pueden filtrar a nivel estatal o por ciudad, cuando esté disponible. La fecha más temprana en que se obtuvieron los datos es a partir de enero de 2008 y puede buscarse en forma mensual, semanal, diaria, por hora, y por minuto las frecuencias. También es posible obtener las consultas con los índices de correlación más altos de series de tiempo utilizando la herramienta Google Correlate, que utiliza el hash asimétrico para encontrar los 100 más cercanos vecinos a través de millones de series de tiempo candidatas en menos de 200 milisegundos (VANDERKAM, 2013). Usa una correlación estándar de Pearson para medir qué tan cerca está una serie única de otra y aplica un algoritmo de k-medias para calcular la distancia aproximada entre vectores.

En esta investigación, tomamos el Índice de compras al por menor de México desde 2008

hasta marzo de 2017 y encontramos la mayor búsquedas correlacionadas para el período. Los Índices de correlación de Pearson dados por :

$$r(x, y) = \frac{cov(x, y)}{\sigma_x \sigma_y} = \frac{\sum_{j=1}^n [(y_j - \mu(y))(x_j - \mu(x))]}{\sqrt{\sum_{j=1}^n (y_j - \mu(y))^2} \sqrt{\sum_{j=1}^n (x_j - \mu(x))^2}} \quad (1.1)$$

donde x es un elemento del vector de regresores X . Este conjunto de regresores, está compuesto por varios términos relacionados con las compras al por menor, cabe destacar la proporción de búsquedas encontradas con las compras al por menor, considerando que las búsquedas no fueron restringidas por algún tipo de consulta. El conjunto de consultas encontradas relacionadas incluye, electronics, automóviles, artículos para el hogar, pasatiempos, comida, ropa y calzado. Es notable la cantidad de productos de consumo que se consideran como ventas al por menor como lo son: sacos, artículos para adornar, botas, moños, ropa de hombre y horarios de tiendas departamentales. Los usuarios del buscador de Google naturalmente buscan conocer los horarios de las tiendas departamentales para acudir a comprar, o incluso hacer compras en línea. La información recolectada a través del motor de búsqueda de Google es entonces usada como un predictor en un modelo de series de tiempo que se describe en la siguiente sección.

1.2. Objetivo General

Mejorar el pronóstico de los valores actuales del índice de compras al por menor en México, utilizando las tendencias de búsquedas de Google como variables regresoras en un modelo de espacio estado.

1.2.1. Objetivos Particulares

1. Construir un modelo de espacio estado sin regresores para pronosticar los valores actuales del índice de compras al por menor del comercio en México
2. Incluir al modelo anterior las tendencias de búsqueda de Google como variables regresoras
3. Comparar ambos modelos anteriores y decidir cual de los dos tiene una mejor estimación

1.3. Justificación

Debido a que los datos recabados por el INEGI se presentan con un rezago significativo para su estudio, los expertos en la ciencia económica buscan formas de obtener información prácticamente al día, el modelo de series de tiempo que se presenta en este estudio proporciona una estimación de los datos al día. El tener estos datos con observaciones diarias ayuda a prever a las empresas minoristas, a saber en que momento invertir, ahorrar, sobre todo en tiempos de incertidumbre económica.

1.4. Alcances y limitaciones

El propósito de esta investigación es hacer un modelo de pronóstico del índice de compras al por menor del comercio en México, gracias a esto, se tendrá a la mano un dato aproximado y al día de dicho indicador. Esto se logra a través de proyecciones al presente, o *nowcasting*. El *nowcasting* significa encontrar el valor que deberíamos encontrar de una variable, si pudiéramos recolectarla con una mayor frecuencia. En el caso de este trabajo, se proyecta una serie de tiempo de frecuencia trimestral a una frecuencia diaria. Sin embargo, la proyección al presente, requiere del contexto en el que los modelos se están generando. Esto significa que es necesario actualizar el modelo con la información presente para que este pueda ser válido en el contexto en el que se aplica. Un ejemplo de la importancia de este paso es el cambio en los hábitos de consumo producto de la pandemia COVID-19 (CASCO, 2020), donde las ventas al por menor sufrió cambios importantes.

También cabe señalar que no se pretende sustituir el trabajo que hace el INEGI para la recolección de información. El alcance de esta investigación es a nivel nacional, se puede replicar a nivel internacional o mundial adaptando el contexto de cada región.

Capítulo 2

Marco Teórico

2.1. Definiciones

1. Empresa comercial

Una empresa comercial es una sociedad mercantil que se dedica a la compra y venta de productos sin que estos sufran una transformación, los bienes comprados pueden ser materias primas, bienes terminados o semiterminados. Las empresas comerciales pueden ser minoristas o mayoristas (ASTUDILLO, 2012).

a) **Empresa comercial minorista**

Las empresas comerciales minoristas se dedican por lo tanto a las ventas al por menor, al menudeo, es decir, venta en pequeñas cantidades.

b) **Empresa comercial mayorista**

Las empresas mayoristas son aquellas que se dedican a la venta al por mayor, es decir, ventas en grandes cantidades.

2. Indicador económico

Es un dato estadístico que ayuda a comprender la situación económica a lo largo del tiempo, tanto en eventos pasados, presentes e incluso se pueden simular eventos futuros. La comparación de estos cambios es lo que le da valor agregado a estos indicadores, el no hacer esta comparación es recabar datos sin ninguna utilidad o relevancia. Los indicadores ayudan a comparar situaciones económicas entre países o incluso dentro de un mismo país. Este dato solo es importante en relación con algo, al escuchar este tipo de resultado por si solo no representa nada, a menos que se compare con meses anteriores para observar si aumentó o disminuyó y en que medida. Un ejemplo es el Índice de Confianza del Consumidor (ICC), en julio del 2010 fue de 87.4 (INEGI, 2010). Este número por sí solo no significa nada, sólo si lo comparamos con los datos de otros meses.

Los números índices se utilizan para expresar varios tipos de actividad económica como producción, precios, salarios, productividad e indicadores compuestos como coincidentes y adelantados. Es un método para resumir la información y poder interpretar rápidamente la dirección y tamaño del cambio de un periodo dado a otro (FRUMKIN, 2004).

a) Periodo base

Un número índice tiene un punto de partida, un año o el promedio de unos años consecutivos, se define como equivalente a 100 y todos los movimientos del indicador antes y después del periodo base se representan como diferencias porcentuales respecto a éste. Con base de 100, un índice de 95 quiere decir que el indicador para este periodo (mes, trimestre, año) es de 5 % menor al periodo base. La siguiente fórmula sirve para calcular el cambio porcentual entre dos periodos:

$$\frac{Periodo2 - Periodo1}{Periodo1} = \frac{Periodo2}{Periodo1} - 1 \quad (2.1)$$

Para calcular el cambio porcentual entre 95 y 114 para los periodos 1 y 2:

$$\frac{108}{90} - 1 = 1.2 - 1.0 = 0.2$$

En porcentaje se escribe como un 20 %, por lo tanto se puede decir que 95 es el índice del periodo 1 y 114 es el índice del periodo 2, el cambio porcentual entre los dos periodos es 20 % (ADAME, 2005).

La selección del periodo base se refiere al tiempo en que se realizó una encuesta para determinar los ponderadores del indicador. Cualquier valor del periodo que se tome como base será igual a 100 y de esta manera se podrá calcular la diferencia entre periodos. Esto permitirá hacer cambios de base que en realidad lo que hace es normalizar la serie, multiplicarla por un valor constante para que esto no afecte los cambios porcentuales entre periodos.

3. Índices reales

Cuando se estudia economía es necesario separar la función y efecto de los precios y de los valores reales. Comúnmente un indicador se presenta ya “deflactado”, es decir, dividido entre un índice de precios para reflejar únicamente las variaciones reales de algún tipo de actividad económica o lo que se conoce como “precios constantes”. Lo que se hace es dividir el índice “nominal” entre un índice de precios relevante. Esto implica dividir cada elemento del índice en cuestión por el nivel de precios que existe en el mismo periodo (ADAME, 2005).

- 4. Fuentes de los indicadores** Hoy en día en México se puede obtener información de suma importancia a través de los datos abiertos, proporcionados por diversas instituciones del país. Los datos abiertos son aquellos que pueden ser utilizados, reutilizados

y redistribuidos libremente por cualquier persona. Los datos pueden ser de primero, segundo y tercer nivel, según su importancia. De primer nivel son aquellos que pueden impactar los mercados financieros, bolsa de valores o el mercado de dinero. Las estadísticas de primer nivel son presentadas de manera regular, se avisa la fecha en que serán divulgadas, por ejemplo los censos económicos que realiza el INEGI, primero, presentan el resultado de datos oportunos, aproximadamente seis meses antes de los resultados definitivos. También existen instituciones privadas que presentan información mediante indicadores de alta importancia que no tienen aún la formalidad necesaria como la Asociación Mexicana de la Industria Automotriz (AMIA) y la Asociación Nacional de Tiendas de Autoservicio y Departamentales (ANTAD). Entre el INEGI y el Banco de México se producen la mayoría de los indicadores, además de que se puede consultar la información en internet, estas instituciones ofrecen suscripciones para recibir los boletines de prensa en el momento en que se reportan, también el personal de estas instituciones se encuentra elaborando aplicaciones para la consulta de datos de manera interactiva como el DENU que muestra la relación geográfica de todos los establecimientos económicos, y esto es de gran utilidad para sus usuarios, ya que muestra un panorama general de la distribución de los negocios, con esta herramienta se puede observar que tipo de negocio y en que cantidad se ubican en ciertas zonas de interés, ya sea para un emprendedor o para un consumidor (HEATH, 2012).

a) INEGI

El Instituto Nacional de Estadística y Geografía (INEGI), es el encargado de regular, normar y coordinar el Sistema Nacional de Información Estadística y Geográfica (SNIEG). El INEGI fue creado en el año de 1983, en el mes de enero, como una institución oficial encargada de realizar los censos nacionales y recopilar información estadística. En el año del 2008 entró en vigor una ley que establece al INEGI como un órgano autónomo y como una dependencia de gobierno, en la llamada Ley del Sistema Nacional de Información Estadística y Geográfica (LSNIEG), en esta ley se delega al INEGI la función de responsabilidad de toda la información de interés nacional, es por ello que esta institución genera la gran mayoría de los indicadores económicos de mayor relevancia. A pesar de que el INEGI se fundó en 1983, desde 1882 se genera información estadística económica básica a través de la dirección general de estadística quien posteriormente se convierte en INEGI.

En la página de internet del INEGI se encuentra el calendario de divulgación de 16 familias de indicadores económicos que en su mayoría se publican mensualmente. El público puede darse de alta y recibir los comunicados de prensa por correo electrónico, esto mediante el Banco de Información Económica (BIE) del INEGI. Se puede tener acceso a los comunicados pulsando en el enlace correspondiente, se muestran dos archivos, uno en formato word y el otro es el tabulado a manera de resumen de los datos en excel, cada comunicado tiene la explicación metodológica del indicador en cuestión y se presenta a las 8 de la mañana del día calendarizado. En el comunicado de coyuntura se presenta la tendencia del indicador a lo largo del último año hasta el último dato,

incluye una gráfica de las series desestacionalizadas y tendencia-ciclo. El BIE contiene más información histórica del indicador, aproximadamente 160 mil series históricas agrupadas en 16 grandes temas. El tema Indicadores económicos de coyuntura.^{es} el más importante, contiene todas las series de primer nivel. El INEGI está trabajando en publicar la información con un mes más de oportunidad, existe un indicador que se publica de manera retardada con respecto a Estados Unidos, es llamado producción industrial, en ambos países se publican en la misma fecha, sin embargo el dato de México hace referencia al mes anterior al que se reporta en Estados Unidos. Este hecho no es bueno, ya que la producción industrial es una de las variables más importantes para comprender la coyuntura de la economía mexicana, por la importancia que tiene el sector como motor de crecimiento. Cabe mencionar que los datos en México son más precisos y en Estados Unidos son preliminares, por lo tanto, se sacrifica oportunidad por precisión. Los índices de difusión son los indicadores que primero se comunican por su relativa facilidad de procesamiento y de reportar. Por lo regular estos indicadores se dan a conocer en los primeros días hábiles del mes siguiente. Los más oportunos en cuanto a los datos duros son los de la balanza comercial y de empleo, estos se dan a conocer al público aproximadamente tres semanas después de haberse concluido el mes. En julio del 2011 el INEGI, empezó a fungir como responsable del cálculo y difusión de los índices nacionales de precios. A pesar de que en la página del Banco de México se puede consultar la información de precios, el INEGI publica de manera más puntual todos los índices. El índice nacional de precios al consumidor (INPC) se considera el más importante, se publica cada quince días, 8 o 9 días posterior a la conclusión del periodo (INEGI, 2009a).

b) BANXICO

El banco responsable de las políticas monetaria y financiera es el Banco de México (BANXICO). Es un órgano que fue fundado en septiembre de 1925 y alcanzó su autonomía en 1994, esta institución elaboraba índices de precios al consumidor y productor, hasta que cedió estas funciones al INEGI en el 2011, también produce información sobre tipos de cambio, actividad financiera, agregados monetarios y tasas de interés. Una vez al mes levanta una encuesta de opinión y expectativas entre los especialistas en economía del sector privado. En conjunto con el INEGI trabaja en la elaboración de los estadísticos de comercio exterior, aunque el responsable es el INEGI y hacen los reportes de manera trimestral de la balanza de pagos y las mensuales de las remesas. Tiene su encuesta propia de opinión donde contempla preguntas que el INEGI no aplica, también en su página de internet tiene un apartado de estadísticas muy importantes que son relevantes para la política monetaria. Elabora indicadores de interés analítico que no son divulgadas al público, también tiene su calendario de difusión en donde únicamente pone fechas para las estadísticas más relevantes para la institución y de decisión de política monetaria. Hoy en día su política de comunicación ha mejorado respecto a otros años, pero no es del todo transparente (BANCO DE MÉXICO, 2011).

c) SHCP

La encargada de manejar la información pertinente acerca de las finanzas públicas y la deuda pública es la Secretaría de Hacienda y Crédito Público (SHCP). También cuenta con su página de internet donde se pueden consultar estadísticas oportunas de finanzas públicas y otros tipos de informes semanales, mensuales y trimestrales sobre la situación económica, las finanzas públicas y deuda pública. Los datos más importantes que proporciona la SHCP son los grandes agregados de las finanzas públicas, por ejemplo, el balance financiero total, el balance primario y la deuda pública (PODER EJECUTIVO, 2007).

d) NBER

Una institución de Estados Unidos que realiza investigación económica es la National Bureau of Economic Research (NBER) quién en su página de internet ofrece un servicio sin costo de notificación de los indicadores económicos de la economía estadounidense. Se puede escoger qué indicadores son de interés y recibir una notificación con anticipación de la hora y lugar donde se puede ubicar la información (BARTELSMAN, 1996).

5. Los ciclos económicos

A los periodos de expansión y retracción de la actividad económica de un país a lo largo del tiempo se les llama ciclos económicos, los cuáles son estudiados por la macroeconomía.

La toma de las decisiones que las personas realizan en cuanto a sus inversiones, negocios y empleo, son en función del lugar en que se encuentra el ciclo económico del país en el que vive, por ello la importancia de estudiar el comportamiento de los indicadores. Si existe un crecimiento económico aumenta la demanda de empleos, mientras que el crecimiento de la población genera una oferta laboral, de aquí la importancia de que la economía crezca lo suficiente para tener una tasa baja de desempleo, aún así, si el crecimiento es muy alto, la demanda puede exceder la oferta laboral y esto ocasiona presiones inflacionarias. La inflación provoca una merma en el poder adquisitivo, distorsiona el sistema de precios y hace más difícil la asignación eficiente de recursos económicos.

Los altibajos que experimentan la actividad económica a través del tiempo es lo que se conoce como ciclos económicos, las variaciones pueden ocasionarse por ajustes en la política interna o en los países, por una guerra, desequilibrios en ciertos mercados, variaciones en los precios de las materias primas.

Cada ciclo económico tiene una duración y amplitud distinta, por ello, no pueden existir dos ciclos económicos iguales. Es importante identificar de manera oportuna las diferentes tareas de un ciclo económico ya que afecta de distintas formas las inversiones, negocios y empleos. Una de las principales razones de estudiar y analizar los ciclos económicos es para entender en que parte del ciclo estamos y hacia donde vamos. El término ciclo económico (business cycle) se refiere a las variaciones en

actividad o producción económica que abarca desde unos cuantos trimestres hasta varios años. Un ciclo consta de expansiones simultáneas, seguidas de recesiones generales simultáneas, seguidas de recesiones generales similares, contracciones y recuperaciones que se unen en una etapa nueva de expansión en un siguiente ciclo. Un ciclo varía de mas de un año a periodos de diez o doce años y no son divisibles en ciclos mas cortos con características similares. Existen cuatro puntos fundamentales en un ciclo económico.

- a) Los ciclos económicos son fluctuaciones en la actividad económica agregada.
- b) Un ciclo consiste en expansiones simultáneas en la mayoría de las actividades económicas, seguida de recesiones regeneralizadas de forma similar.

Un ciclo económico se enfoca en el ecomportamiento generalizado de la economía no solo en un sector en particular, ni en un negocio específico. Hay dos etapas principales en un ciclo económico, la de disminución (momento de resección) y la de crecimiento. El punto maximo de un ciclo es llamado (peak) ya que representa el nivel mas alto y el piso o valle (through) que es el mínimo comunmente el máximo, se usa como punto de partida para uno nuevo y para el final del anterior, esto significa que la primera fase de un ciclo es la resección y termina en el punto mas bajo. Cuando la economía empieza a crecer nuevamente se dice que está en su etapa de recuperación y comienza al concluir la resección, termina al regresar al punto máximo anterior, entonces comienza la etapa de expansión. Esta fase por lo general es la más prolongada, así pasa con la mayoría de los países desarrollados. El comportamiento del ciclo económico en general se conforma con las tres etapas básicas mencionadas, recesión, recuperación y expansión (MITCHELL, 1941).

6. Series de tiempo

Una serie de tiempo es un conjunto de datos estadísticos que se recopilan y se organizan a través del tiempo, este conjunto de observaciones X_t se registran en un tiempo t , de manera diaria, semanal, mensual, anual o en cualquier otro intervalo. Un ejemplo de series de tiempo son los precios de cierre de acciones en la bolsa, la medición de presión sanguínea en el tiempo, resonancia magnética para estudiar reacciones del cerebro a estímulos.

Proveer una descripción compacta de la información, realizar estudios de simulación, hacer predicciones, éstos son los objetivos principales de las series de tiempo (MONTGOMERY, 2015).

a) Series de Tiempo de espacio de estado

La propiedad de Markov es la base para los modelos de espacio de estado, implica la dependencia del futuro de un proceso a partir de su pasado, dado el estado actual del sistema, en este sistema el estado del momento actual contiene la información pasada que se requiere para predecir el futuro comportamiento del proceso. En este tipo de series temporales, un vector X_t es quien modela el estado del sistema en el momento t . Los elementos de este vector no son precisamente observados, por lo que, un modelo de espacio de estado consiste

en dos ecuaciones: una ecuación de medición que describe como se producen las series de tiempo a partir del vector de estado, por otro lado, la ecuación de estado que describe la forma en que el vector de estado evoluciona a través del tiempo. Estas dos ecuaciones se describen de la siguiente forma:

$$Y_t = h_t X_t + \varepsilon_t \quad (2.2)$$

$$X_t = A X_{t-1} + G_t \quad (2.3)$$

Donde:

La ecuación 2.2 se conoce como ecuación de observación.

La ecuación 2.3 se conoce como ecuación de estado.

Y_t Es el vector que modela el espacio o la observación del sistema.

X_t Es el vector que modela el estado del sistema.

En la ecuación de observación h_t es un conocido vector de constantes y ε_t es el error de observación.

Si la serie de tiempo es multivariada entonces y_t y ϵ_t se convierten en vectores, y el vector h_t se convierte en una matriz H . En la ecuación de estado A y G son matrices conocidas y a_t es el proceso de ruido.

La ecuación de estado se asemeja a un modelo AR(1) multivariado excepto que representa las variables de estado en lugar de una serie de tiempo observada, y tiene una matriz G extra (DURBIN, 2012).

7. Acerca del Internet

En la actualidad las personas se refugian en las redes sociales donde buscan otras personas con quienes compartir ideas y emociones, buscar perfiles afines y tener muchos amigos.

Al existir en las redes sociales una amplia diversidad de personas con distintos pensamientos, existe mayor probabilidad de que personas se encuentren con opiniones opuestas.

Facebook muestra conexiones sociales débiles, una persona tiene como contacto al amigo del amigo, el primo del amigo, el tío incómodo, el enemigo de la infancia, entre otro tipo de contactos. Son personas con las que no pueden compartir un helado, una cerveza, una carne asada o ver una película juntos. El internet une personas con diferentes puntos de vista políticos, sociales o religiosos.

El 5 de septiembre del 2006 los ingenieros de Facebook introdujeron una actualización importante en su página de inicio. Las primeras versiones de facebook solo permitían a los usuarios hacer click en los perfiles de sus amigos para saber qué estaban haciendo. En ese momento el sitio web tenía 9.4 millones de usuarios, pero después los ingenieros proporcionaron a los usuarios actualizaciones sobre las actividades de todos sus amigos.

Google encontró el secreto de cuan poderosa puede ser la información de lo que las personas publican en las redes sociales y escriben en las páginas de búsqueda. Hay

marcas que en la actualidad están utilizando algoritmos de inteligencia artificial para vincular las preferencias de los usuarios y lanzar publicidad digital promoviendo sus productos o servicios (STEPHENS-DAVIDOWITZ, 2017).

Capítulo 3

Metodología

3.1. Tipología del estudio

En esta investigación es de tipo documental ya que se recopiló información de distintas fuentes bibliograficas y documentales, también es exploratoria ya que se formula el problema del rezago de la publicación de los indicadores, a su vez es descriptiva porque analiza el comportamiento del índice de compras al por menor y sus detalles como fenómeno económico.

3.2. Método de investigación

El método de esta investigación es de análisis, ya que se estudia el problema de manera seccionada, descomponiendo el objeto de estudio de forma contextual, a través de un indicador; además se utiliza el método hipotético, puesto que se pretende demostrar una hipótesis.

3.3. Hipótesis

H_0 : El error medio del modelo de espacio estado sin regresores es igual al error medio del modelo con regresores.

H_1 : El error medio del modelo de espacio estado sin regresores es mayor que el error medio del modelo con regresores.

3.4. Técnicas y procedimientos para recopilar información

La recopilación de la información se llevó a cabo mediante la consulta bibliográfica y revisión documental como libros, revistas científicas y páginas web. La información del índice de compras al por menor en México es recopilada por el INEGI con datos de diferentes encuestas aplicadas a empresas en la industria manufacturera, servicios, y registros de la balanza comercial y el índice Nacional de Precios (INEGI, 2013). A través de Google Correlate se obtuvo información acerca de las tendencias de búsquedas de Google correlacionadas con el índice de compras al por menor en México.

3.4.1. Recopilación de los datos

En esta sección se describirán los datos utilizados por el modelo, la variable de interés es el índice de compras al por menor, el cual se describe más adelante. El índice de compras al por menor puede ser muy descriptivo y útil para la economía, además puede ser razonable pensar que el índice puede ser pronosticado por la información obtenida de los datos de búsqueda de Google, que se describe más adelante en esta sección junto con su interpretación.

3.5. Técnicas de análisis de información

Para probar que existe diferencia significativa de errores entre el modelo de espacio estado sin regresores y con regresores, se siguen los siguientes pasos:

1. Obtener bases de datos
2. Se construye un modelo de predicción sin componentes regresores
3. Se utilizan las tendencias de búsqueda de Google correlacionadas con el índice de compras para anexarlos al modelo de predicción
4. Se hace una prueba para comparar los resultados con y sin regresores, lo anterior, para decidir cuál de los dos modelos hace una mejor estimación.

Capítulo 4

Resultados

4.0.1. El modelo de pronóstico

En esta sección se utiliza un modelo de serie temporal de espacio de estado que incluye una tendencia, estacionalidad y un conjunto de regresores. Usamos la estructura y la metodología descrita en Scott and Varian (2014) de series de tiempo bayesianas estructurales, así como el paquete R desarrollado por los autores.

Un modelo de serie de tiempo es dividido en dos ecuaciones, la ecuación de observación y la ecuación de transición de estado. En este tipo de modelo, un estado latente es observado indirectamente a través de otra variable, por lo tanto en:

$$y_t = Z_t^T \alpha_t + \varepsilon_t; \quad \varepsilon_t \sim N(0, \sigma_\varepsilon^2) \quad (4.1)$$

La variable y_t es la observación de el estado latente α_t . La observación es alterada por un conjunto de parámetros estructurales contenidos en Z_t y un ruido de observación, capturado por la distribución normal ε_t , donde cada observación es independiente de el resto de de las series de tiempo. El estado latente α_t evoluciona con el tiempo de acuerdo a:

$$\alpha_t = T_t \alpha_{t-1} + R_t \eta_t; \quad \eta_t \sim N(0, \sigma_\eta^2) \quad (4.2)$$

donde : T_t y R_t son también parámetros estructurales y η_t es independiente para ε_t y para sí mismo también. No es poco común que Z_t , T_t y R_t , contienen en su mayoría estratégicamente colocados ceros y unos, los cuales son indicadores de la relevancia de la variable de estado al modelo. El modelo más simple es un modelo a nivel local, donde $\alpha_t = \mu_t$, Z_t , T_t , R_t son todos una instancia del escalar 1.

$$y_t = \mu_t + \varepsilon_t; \quad \varepsilon_t \sim N(0, \sigma_\varepsilon^2) \quad (4.3)$$

$$\mu_{t+1} = \mu_t + \eta_t; \quad \eta_t \sim N(0, \sigma_\eta^2) \quad (4.4)$$

Note que si $\sigma_n = 0$ entonces μ es una constante y la ecuación de observación muestra diferentes instancias ruidosas del mismo estado. Esto significa que el mejor estimador para cualquier observación particular de y_t es la media de las observaciones previas. Entonces un modelo en el que σ_ε implica que el mejor estimador de y_t es y_{t-1} , ya que es una caminata aleatoria. También cabe señalar que el estado actual depende indirectamente del conjunto completo de observaciones, aunque solo depende del estado anterior (i.e., it is Markov). Una de las principales razones para usar modelos de series de tiempo de espacio de estado es su flexibilidad y transparencia (Box and Jenkins, 2015). La forma de los parámetros puede ser modificado de forma modular para modelar la tendencia, estacionalidad, regresiones y otros efectos que pueden ser necesarios con flexibilidad. Este modelo es considerablemente transparente y que no se basa en las diferencias, resagos y medias móviles. En cambio, la regresión y la estacionalidad se incluye directamente en el modelo y se pueden inspeccionar visualmente. Además, cualquier modelo ARIMA puede ser expresado en modelos de espacio de estado.

La figura 4.1 muestra la tendencia y estacionalidad del índice de compras al por menor de México desde enero del 2008 al 2017. Esta información puede ser obtenida de un modelo de espacio de estado sin componentes de regresión. Un modelo general con tendencia μ_t , un patrón de estacionalidad τ_t y componentes de regresión $\beta^T x_t$ puede ser declarado como:

$$y_t = \mu_t + \tau_t + \beta^T x_t + \varepsilon_t \quad (4.5)$$

$$\mu_t = \mu_{t-1} + \delta_{t-1} + u_t \quad (4.6)$$

$$\delta_t = \delta_{t-1} + v_t \quad (4.7)$$

$$\tau_t = - \sum_{s=1}^{s-1} \tau_{t-s} + w_t \quad (4.8)$$

En este modelo, el término de error de la variable de estado es un vector $\eta = (u_t, v_t, w_t)$ que contiene elementos de un ruido Gaussiano. La tendencia es similar al modelo de nivel

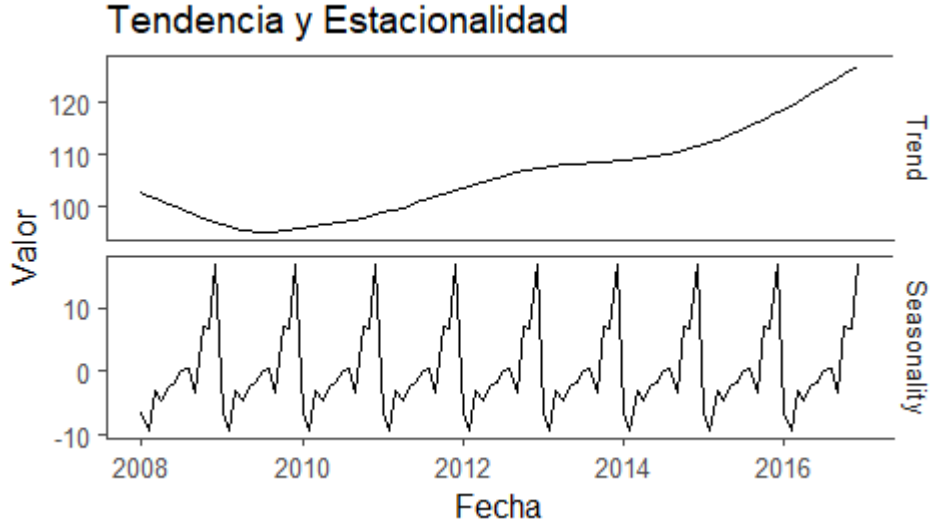


Figura 4.1: Fuente: elaboración propia.

local anterior, además de un término δ_t agregado como una pendiente de la tendencia lineal local. Esta adición refleja la creencia de que la serie de tiempo se está desviando en una dirección particular, de modo que la predicción de la siguiente observación debe reflejar un aumento o disminución continua en la dirección de la tendencia corriente. Las pendientes regularmente multiplican alguna variable, pero en este caso el coeficiente de δ_t es solo $\Delta_t = 1$. La variable de estado de (3) se convierte en una matriz diagonal con σ_u^2 , σ_v^2 y σ_w^2 como sus elementos y $H_t = \sigma^2 \varepsilon$.

El componente estacional puede entenderse como una regresión con S variables estacionales. En Scott and Varian (2014), el análisis se hace con datos semanales donde los autores establecen $S = 52$. Dado que este documento trabaja con datos mensuales se establece $S = 12$. El vector de componentes de regresión x se realiza a partir de los datos de búsqueda de Google.

Una vez que el modelo está en forma de espacio de estado, podemos usar el filtro de Kalman (Kalman, 1960) que conduce a la suavidad de Kalman y el aumento de datos bayesiano (cf. Hoeting 1999) para las predicciones. El filtro de Kalman es un procedimiento recursivo para calcular el estimador óptimo del vector de estado, dada la información disponible.

Sea $y_{1:t}$ el conjunto de observaciones de la serie de tiempo hasta el tiempo t . Dado el supuesto de que el estado inicial es $\alpha_t \sim N(a_0, P_0)$ con a_0 y P_0 conocidos, es posible calcular la distribución predictiva $p(\alpha_{t+1}|y_{1:t})$ (Harvey, 1990). Esto se hace combinando $p(\alpha_{t+1}|y_{1:t})$ y $p(\alpha_t|y_{1:n})$, donde n es la longitud de la serie de tiempo completa, puede ser probado que el resultado es equivalente a una regresión lineal (Bode, 1950). Se usó el método de Durbin y Koopman (2001) para simular ruido aleatorio con las covarianzas extraídas de $p(\alpha_{1:n}|y_{1:n})$. El conjunto de consultas de búsqueda con alta correlación con el índice de compras al por menor se utiliza en la ecuación de observación como $\beta^T x$ en la ecuación (3). Aunque parezca que es un gran conjunto de regresores, se espera un alto grado de esparcimiento, es decir, la mayoría de los predictores tienen un coeficiente cero. Una descripción de veintidós algoritmos diferentes para la selección de las variables se realizan en Castle (2009), donde

identifican los enfoques como criterios de información, selección de una cartera"del mejor subconjunto de modelos, algoritmos generales a específicos, enfoques de regresión paso a paso y modelo de promediación Bayesiano.

La técnica de "punta y losa" para seleccionar las variables de regresión del conjunto disponible fue propuesta por primera vez por George y McCulloch (1997), Madigan y Raftery (1994). Sea γ una variable ficticia del mismo tamaño, como β es cero cuando una variables no es incluida en la regresión y uno cuando está incluida. Por lo tanto β_γ representa el conjunto de regresores incluidos en la regresión, es decir, el subconjunto para el cual $\gamma_i = 1$. La varianza residual del modelo de regresión es σ^2 . Una punta y losa antes de la distribución conjunta de $(\beta, \gamma, \sigma^{-2})$ puede ser calculado por:

$$(\beta, \gamma, \sigma^{-2}) = p(\beta_\gamma | \gamma, \sigma^{-2}) p(\sigma^{-2} | \gamma) p(\gamma) \quad (4.9)$$

Se supone que cada variable tiene distribución Bernoulli, por lo que la anterior es el producto

$$\gamma \sim \prod_i \pi_i^{\gamma_i} (1 - \pi_i)^{1-\gamma_i} \quad (4.10)$$

Si no hay información detallada, es una práctica común establecer π_i en el mismo número para todos los regresores. En el caso de un conocimiento previo de los regresores esperados, es posible establecer k fuera de los K promedios distintos de cero por un ajuste $\pi = k/K$. En esta aplicación, algunas consultas de búsqueda se establecen en cero y el resto recibe un valor previo de $\pi_i = 0.1$. Este es el componente "punta" del método de selección. La losa es una parte anterior de los valores con coeficientes distintos de cero. Sea b una suposición previa de los coeficientes de los regresores y Ω^{-1} una matriz de precisión anterior. Una losa conjugada anterior es dada por

$$\beta_\gamma | \gamma, \sigma_\varepsilon^2 \sim N(b_\gamma, \sigma_\varepsilon^2 (\Omega_\gamma^{-1})^{-1}), \frac{1}{\gamma, \sigma_\varepsilon^2} \sim \Gamma\left(\frac{v}{2}, \frac{ss}{2}\right) \quad (4.11)$$

donde: Ω_γ^{-1} es la matriz con renglones y columnas de Ω^{-1} para el cual $\gamma_i = 1$. Donde $\Gamma(r, s)$ denota la distribución Gamma con media r/s y varianza r/s^2 . Podemos escoger los parámetros previos para hacer la losa informativa y condicional a γ . Es conveniente establecer $b = 0$ con excepción del término intercepto. El valor v se considera el tamaño de la muestra anterior y ss la suma de cuadrados previa. Ambos valores pueden ser establecidos preguntando el valor esperado R^2 de la regresión. Entonces, establecemos $ss/v = (1 - R^2)s^2y$, donde s^2y es la desviación estándar marginal de la respuesta.

Sea \mathbf{X} la matriz de diseño compuesta por los predictores, de modo que la probabilidad de un modelo de regresión común tenga una matriz de información de Fisher $\mathbf{X}^T \mathbf{X} / \sigma_\varepsilon^2$. Esto

significa que si tomamos $\Omega^{-1} = k\mathbf{X}^T\mathbf{X}/n$, estamos colocando un peso a la media anterior b equivalente de k observaciones (Zellner 1986). En esta aplicación usamos los ajustes por defecto $k = 1$, $R^2 = 0.5$, $v = 0.01$, y $\pi_k = 0.5$. Para calcular la distribución posterior de los parámetros del modelo, usamos simulaciones de montecarlo de cadenas de Markov. El principal objetivo es calcular el valor de los parámetros $\beta, \sigma_\varepsilon^2$ y θ que comprenden todos los demás parámetros del modelo necesarios. Sea ϕ la colección de parámetros que denota $(\beta, \sigma_\varepsilon^2, \theta)$. Entonces, la distribución posterior de los datos completos está dada por:

$$p(\phi, \alpha_{1:n}|y_{1:n}) = p(\phi)p(\alpha_0) \prod_{t=1}^n p(y_t|\alpha_t, \phi)p(\alpha_t|\alpha_{t-1}, \phi) \quad (4.12)$$

Es posible extraer una muestra de la distribución posterior de los parámetros posteriores utilizando un algoritmo de muestreo de Gibbs, alternando entre los extraídos de $p(\alpha_{1:n}|\phi, y_{1:n})$ y $p(\phi|\alpha_{1:n}, y_{1:n})$. Esto producirá una secuencia de pares $(\phi, \alpha_{1:n})_0, (\phi, \alpha_{1:n})_1$ como cadena de Markov para un número de iteraciones dado. Es habitual quemar una cierta cantidad de las primeras iteraciones para garantizar que la distribución resultante de ϕ refleje con mayor precisión el valor esperado posterior de $p(\phi, \alpha_{1:n}|y_{1:n})$.

Durbin y Koopman (2001) describen una técnica de suavizado rápido para calcular E con filtros Kalman. Por lo tanto, utilizamos este algoritmo para simular el estado latente de $(\alpha_{1:n}|y_{1:n}, \theta, \beta, \sigma_\varepsilon^2)$. Luego procedemos a simular $\theta \sim p(\theta|y_{1:n}, \alpha_{1:n}, \beta, \sigma_\varepsilon^2)$ y usamos esta información para simular los parámetros β y σ_ε^2 de una cadena de Markov con distribución estacionaria $p(\beta, \sigma_\varepsilon^2|y_{1:n}, \alpha_{1:n}, \theta)$. Las distribuciones de los parámetros son el principal resultado del trabajo y se utilizan para la predicción del índice de compras al por menor. En la sección 4, hacemos una predicción con $\beta = 0$ y con el conjunto completo de consultas de búsqueda como predictores, estimando los parámetros. El resultado del algoritmo MCMC es, naturalmente, un conjunto de números aleatorios distribuidos cuya media tomaremos como los valores de los parámetros.

4.0.2. Pronóstico del índice de compras al por menor

En esta sección se describe la aplicación del modelo de pronóstico para obtener el índice de compras a través de las consultas de búsqueda de Google. Se utilizó Google correlate para obtener los 100 términos más correlacionados.

El conjunto de datos se divide en conjunto de entrenamiento desde enero del 2008 hasta diciembre del 2016, y un conjunto de prueba desde enero del 2017 hasta noviembre del 2018. Luego, se establecen las especificaciones del modelo, con una estacionalidad de $S = 12$. La Figura 3.2 muestra un pronóstico con un modelo de Serie de Tiempo de espacio de estado sin regresores. Del modelo como se indicó anteriormente, se puede ver que el pronóstico devuelve una distribución, en lugar de un número, y se ve cómo la precisión del modelo disminuye a medida que el momento del pronóstico está más lejos del último punto de datos disponible.

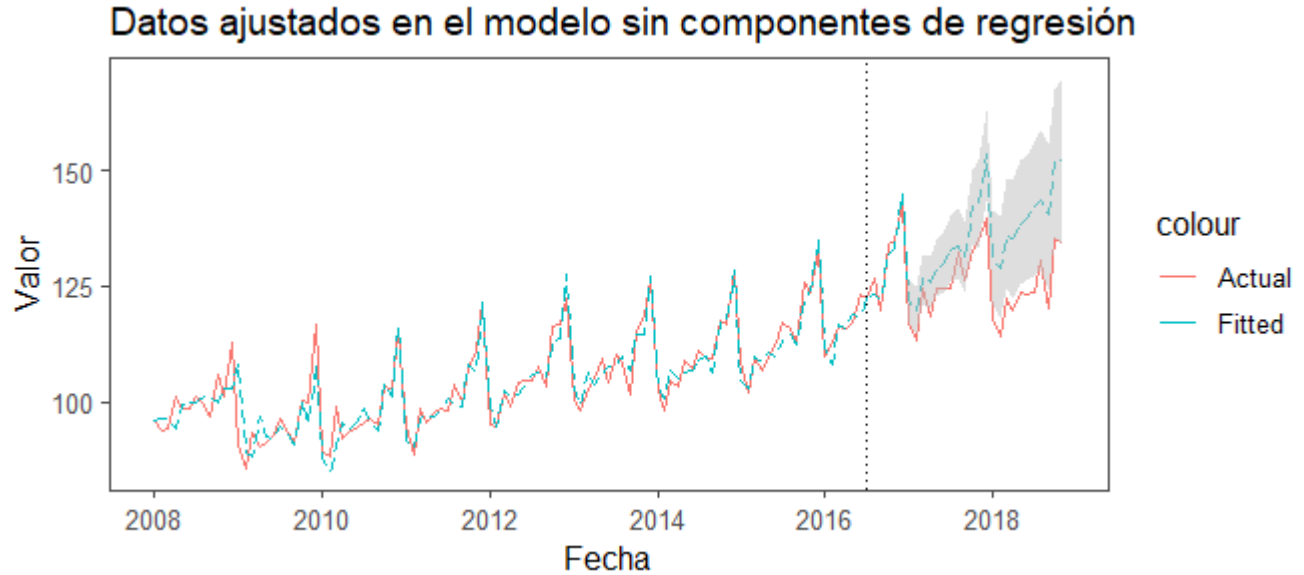


Figura 4.2: Fuente: elaboración propia.

La principal contribución de la investigación actual es mostrar que este modelo se puede mejorar al incluir componentes de regresión al modelo en el que los datos se toman de consultas de búsqueda relacionadas. Se puede pensar que los valores del índice de compras al por menor se pueden pronosticar con mayor precisión antes del lanzamiento oficial de las estadísticas. El modelo con regresores se ejecutó para 5000 iteraciones, utilizando el paquete R de Scott y Varian (2011). Quemamos las primeras 200 iteraciones y calculamos los parámetros posteriores del resto. Establecemos un tamaño esperado de $k = 20$, esto es, esperamos ver en promedio 20 variables con probabilidad positiva de inclusión en la regresión. La Figura 4.3 muestra la media del coeficiente de los regresores interpretado como una probabilidad de inclusión. Incluimos solo los términos de búsqueda con un valor de coeficiente absoluto por encima de 0.05. Como se puede observar, las consultas de búsqueda más correlacionadas positivamente son "tienda.el". La búsqueda de tiendas en línea en estos tiempos representa la nueva manera de comprar, que puede sugerir un porcentaje de las compras al por menor en México. Otros artículos de consumo que también se incluyen en el conjunto son teléfonos celulares, horario Sams, automóviles, sacos y ropa hombre. Muy pocos de los términos de búsqueda que mostraron una alta correlación con el índice tienen una relación poco clara.

La mayoría de las marcas buscadas de artículos de consumo son marcas populares en el país. El conjunto completo de los factores predictivos incluyen empresas de crédito de consumo (es decir, empresas cuyo modelo de negocio considera vender el bien a el consumidor en pagos diferidos), bancos, grandes almacenes, alimentos, muebles y zapatos. Los términos más correlacionados son así porque se han correlacionado persistentemente a lo largo del ciclo económico. Será posible encontrar marcas actuales que sean más informativas del consumo actual, pero con menos poder predictivo a lo largo del tiempo. Como se mencionó anteriormente, una de las principales ventajas de los modelos de espacio de

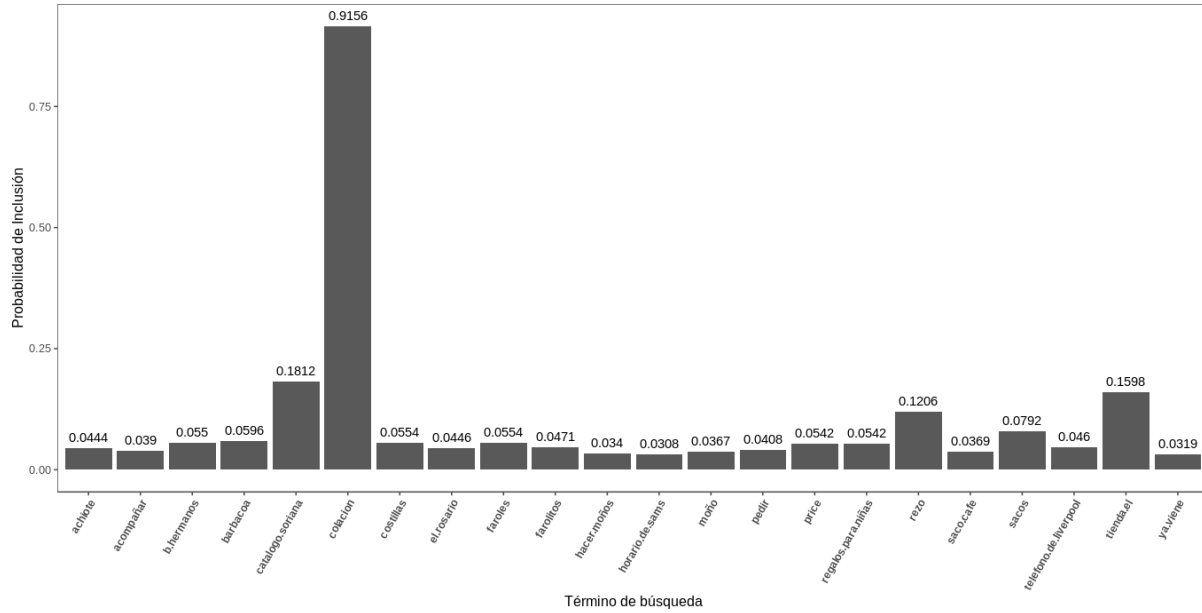


Figura 4.3: Fuente: elaboración propia con datos del INEGI.

estado frente a otras metodologías es su transparencia. En la Figura 4 podemos ver los componentes de tendencia, estacionalidad y regresión del modelo. También este gráfico muestra el valor promedio de los componentes en las iteraciones de cada mes durante el período. Los elementos de la Figura 4 se muestran hasta diciembre de 2016 y se combinan en el modelo para crear un pronóstico de enero de 2017 a noviembre de 2018. El pronóstico resultante y su distribución se muestran en la Figura 5. Nota que la distribución del pronóstico muestra un límite superior e inferior en 97.5 % y 2.5 %, respectivamente, más cerca de la media que la del modelo sin regresores. En la Figura 6, el pronóstico absoluto acumulativo de un paso son comparados los errores en ambos modelos, donde la línea de puntos muestra los errores en el modelo sin la inclusión de consultas de búsqueda. Es posible observar que el tamaño de los errores disminuye considerablemente en el modelo con regresores, mostrados en la línea negra. La inclusión de los términos de búsqueda para la difusión actual del nivel actual del índice de compras al por menor muestra una mejora significativa.

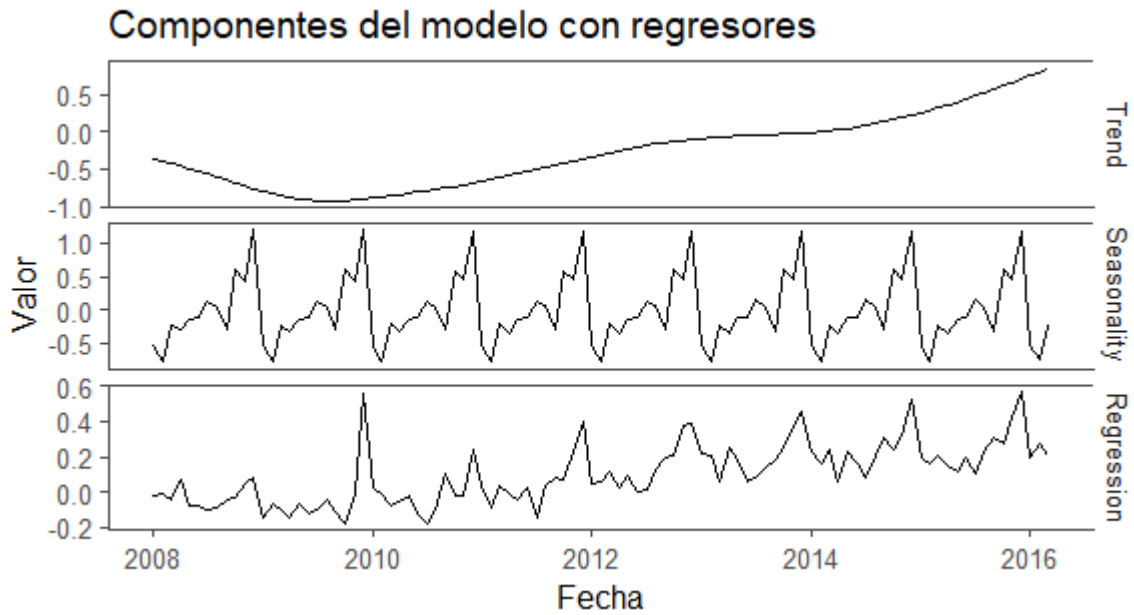


Figura 4.4: Fuente: elaboración propia

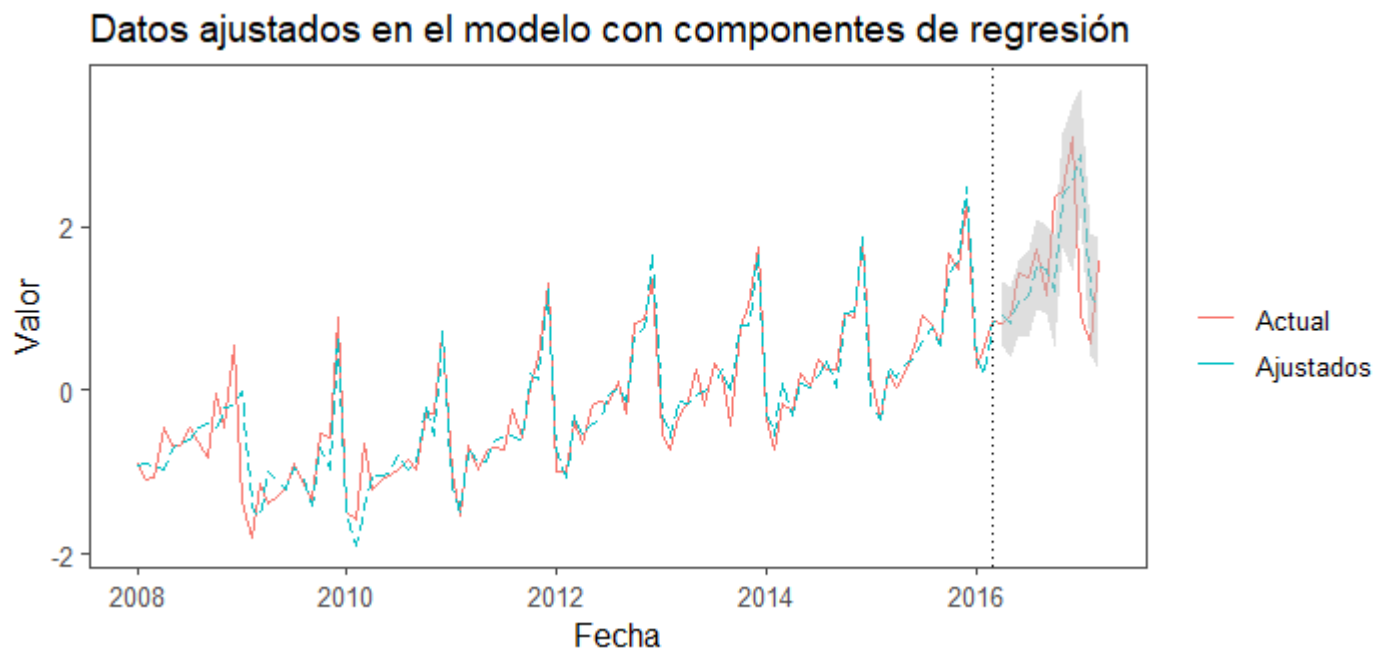


Figura 4.5: Fuente: elaboración propia

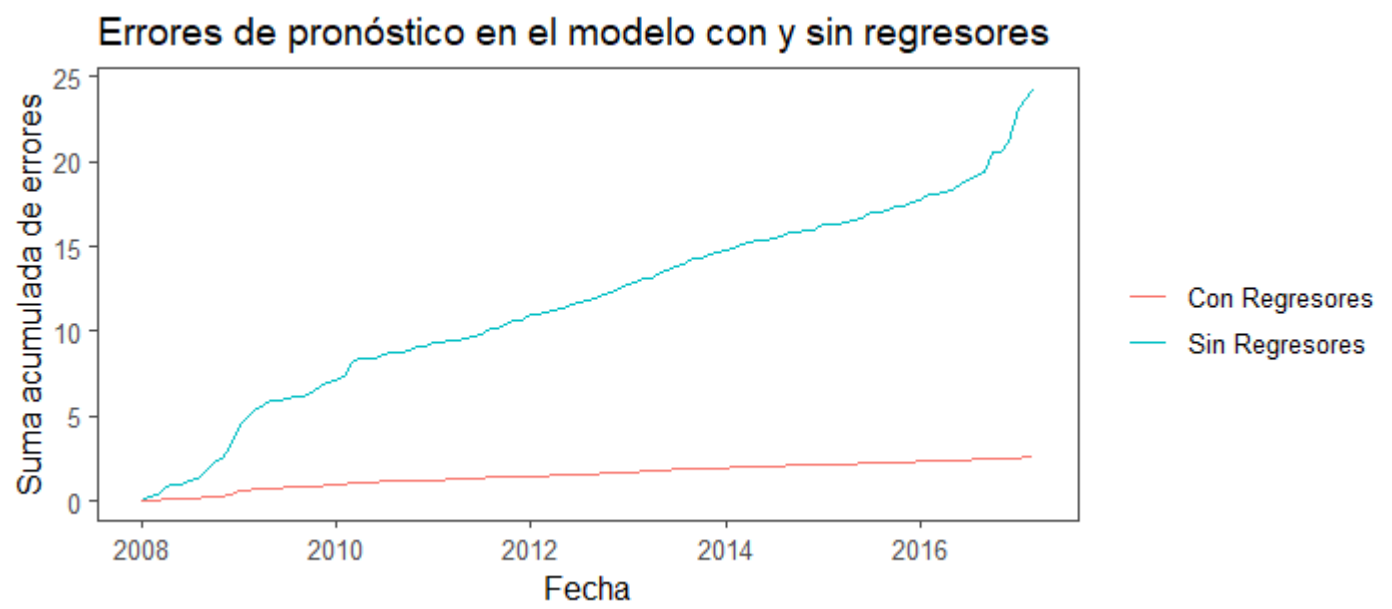


Figura 4.6: Fuente: elaboración propia

4.1. Análisis de los resultados

En esta investigación se desarrolló un modelo para pronosticar el índice de compras al por menor del comercio en México. La información de los datos de Google se usa, por lo tanto, para hacer un modelo de regresión, modelando el comportamiento de las series de tiempo como un modelo de espacio de estado, donde la información de búsqueda de Google se observa y se usa para inferir el índice de compras al por menor que sigue un proceso de Markov Gausiano. Este tipo de modelo permitió incluir la tendencia y la estacionalidad como parte esencial del comportamiento de las series de tiempo. Además, se utilizó una regresión de Spike y Slab, para seleccionar del conjunto de consultas de búsqueda, aquellas con mayor probabilidad de inclusión. Esto se puede hacer estableciendo una prioridad en el modelo sobre el tamaño del conjunto de regresores y los coeficientes que mostraron. Luego, los parámetros se calculan utilizando un algoritmo que genera una distribución de resultados, la media de la salida se utiliza como parámetro.

Se observó de manera gráfica que el modelo de series de tiempo que incluye las tendencias de búsqueda de Google como variables regresoras presenta diferencia significativa contra el modelo que no incluye los regresores, además de que en promedio se observa que es menor. Se realizó una comparación de medianas mediante la prueba U Mann - Whitney ya que los errores no se distribuyen de manera normal, se obtuvo un valor $p = 2^{-16}$ suficientemente menor que el nivel de significancia escogido de $\alpha = 0.05$ y por lo tanto, se encontró evidencia estadística significativa para rechazar la hipótesis nula y concluir que el error medio del modelo de series de tiempo de espacio de estados sin regresores es mayor que el error medio del modelo con regresores, este resultado nos indica que el incluir las tendencias de búsquedas como variables regresoras al modelo es benéfico para reducir el error de predicción y por lo tanto considerarlo como un mejor modelo de pronóstico para el índice de compras al por menor de comercio en México.

La inclusión de consultas de búsqueda en el modelo mostró una reducción significativa en el error de pronóstico. Esto podría resultar en una ventaja para pronosticar el valor presente del índice de compras al por menor cuyo registro oficial generalmente se publica con un retraso de tres meses. Dicha mejora en la disponibilidad de información podría resultar útil para la toma de decisiones de empresas y gobiernos. El conjunto de consultas de búsqueda utilizadas abarca los años anteriores, durante y después de la gran recesión. Además, el ascenso y la caída de diferentes marcas de productos están cubiertos en la información disponible. Esto significa que algunos de los términos correlacionados en el período cubierto pueden dejar de ser significativos en contextos posteriores y el modelo tendría que ser reajustado para tener en cuenta los nuevos artículos y hábitos de consumo. El modelo propuesto permite incluir en el conjunto de predictores otras variables diferentes de las consultas de búsqueda para hacer mejores predicciones. Ya que este tipo de modelos no se basa en diferencias, rezagos, ni medias móviles, se puede decir que es más confiable que un ARIMA por ejemplo, sin embargo se encontraron algunas desventajas: Este tipo de modelos se tornan menos intuitivos, no son tan fáciles de interpretar, por otro lado, Google Correlate ya no existe, es necesario construir manualmente un algoritmo para obtener los datos. El conjunto de regresores se tiene que estar actualizando cada vez que se construya

un modelo.

Se concluyó que el modelo con regresores arroja mejores resultados, sin embargo al seleccionar de manera aleatoria el conjunto de entrenamiento y de prueba, se cae en errores de estimación que posteriormente se tendrá que mejorar con diferentes técnicas de validación, como lo son: Validación simple, validación cruzada, bootstrapping o K fold Cross Validation, este tipo de técnicas requiere alta capacidad computacional.

Referencias

- Achuthan, L., y Banerji, A. (2004). *Beating the business cycle: How to predict and profit from turning points in the economy*. Currency.
- Adame, C. J. C., Lara, A. A. G., y Miguel, R. A. (2005). *Introducción a los indicadores económicos y sociales de México*. UNAM.
- Astudillo Moya, M., y Paniagua Ballinas, J. F. (2012). *Fundamentos de economía*. UNAM, Instituto de Investigaciones Económicas: Probooks.
- Bañbura, M., Giannone, D., Modugno, M., y Reichlin, L. (2013). Now-casting and the real-time data flow. En *Handbook of economic forecasting* (Vol. 2, pp. 195–237). Elsevier.
- Bartelsman, E. J., y Gray, W. B. (1996). The nber manufacturing productivity database. *NBER working paper*(t0205).
- Box, G. E., Jenkins, G. M., Reinsel, G. C., y Ljung, G. M. (2008). *Time series analysis: forecasting and control* John Wiley & sons. Hoboken, NJ.
- Casco, A. R. (2020). Efectos de la pandemia de covid-19 en el comportamiento del consumidor. *Innovare: Revista de ciencia y tecnología*, 9(2), 98–105.
- Choi, H., y Varian, H. (2012). Predicting the present with google trends. *Economic record*, 88, 2–9.
- de México, B. (2008). Medidas instrumentadas por el gobierno federal y el banco de México para preservar la estabilidad financiera. *Informe de inflación julio–septiembre 2008*.
- Durbin, J., y Koopman, S. J. (2012). *Time series analysis by state space methods*. Oxford university press.
- EJECUTIVO, P. (2007). Secretaría de hacienda y crédito público.
- García-Meza, M. A., y Venegas-Martínez, F. (2018). A bayesian structural time series approach to forecast Mexico’s consumer index. *Journal of Applied Economic Sciences*, 13(3).
- Heath, J. (2012). *Lo que indican los indicadores. cómo utilizar la información estadística para entender la realidad económica de México*. inegi. México.
- Mitchell, W. C. (1941). *Business cycles and their causes*. University of California Press.
- Montgomery, D. C., Jennings, C. L., y Kulahci, M. (2015). *Introduction to time series analysis and forecasting*. John Wiley & Sons.
- Poyser, O. (2019). Exploring the dynamics of bitcoin’s price: a bayesian structural time series approach. *Eurasian Economic Review*, 9(1), 29–60.
- Scott, S. L., y Varian, H. R. (2014). Predicting the present with bayesian structural time series. *International Journal of Mathematical Modelling and Numerical Optimisa-*

tion, 5(1-2), 4–23.

Stephens-Davidowitz, S. (2017). Everybody lies: how google search reveals our darkest secrets. *The Guardian*, 9.

Stephens-Davidowitz, S., y Varian, H. (2014). A hands-on guide to google data. *further details on the construction can be found on the Google Trends page*.