

## NOWCASTING

En busca de un indicador adelantado del crecimiento del Producto Interno Bruto de Uruguay utilizando Google Trends.

Germán Hernández Magallanes

Programa de Maestría en Economía de la Facultad de Ciencias  
Económicas, Universidad de la República.

Montevideo - Uruguay

Noviembre de 2021

## TÍTULO DE LA TESIS

En busca de un indicador adelantado del crecimiento del Producto Interno Bruto de Uruguay utilizando Google Trends.

Germán Hernández Magallanes

Tesis de Maestría presentada al Programa de Maestría en Economía de la Facultad de Ciencias Económicas, Universidad de la República, como parte de los requisitos para la obtención del título de Magíster en Economía.

Director de tesis:

Profesor Titular Adrián Fernández

Codirector de tesis:

Director académico:

Profesor Titular Dr. Henry Willebald

Montevideo - Uruguay

Noviembre de 2021

# INTEGRANTES DEL TRIBUNAL DE DEFENSA DE TESIS

---

Profesor Titular Dr.

---

Profesor Titular Dr.

---

Dr. Nombre Apellido

Montevideo - Uruguay

Noviembre de 2021

*A mi tutor, por su paciencia y aportes claves para que el proyecto no naufragara antes de empezar y Federico Molina, Rafael Mosteiro, Augusto Pérez-Souto y Joaquín Álvarez. A R y mi familia.*

## Agradecimientos

*Agradezco a todos aquellos que, de una manera directa o indirecta aportaron a que este proyecto se pudiera concretar. Inicio por el Departamento de Economía y por el Centro de Posgrados de la Facultad de Ciencias Económicas y de Administración. Organizaciones detrás de este programa que tengo la dicha de completar con este trabajo final. Fueron estas organizaciones las que me asignaron a quien supo ser guía y orientador de mis primeras investigaciones, Profesor Henry Willebald. A mis amigos de la facultad, con quienes compartí la carrera de grado y con quienes conservo una amistad fructífera desde el intercambio intelectual. Sin ustedes, que fomentaron el inicio de esta tesis con intercambios, recomendaciones de temáticas y de fuentes, y que luego fueron faro cuando mis habilidades en R me dejaban al borde del naufragio, este proyecto no saldría a la luz. A mis demás amigos y a mi madre, por el tiempo que dejé de compartir con ustedes y el apoyo que me brindaron desde la motivación y lo afectivo en momentos donde la vida se transformó en un ejercicio de supervivencia, a causa de la situación pandémica global y todas sus consecuencias.*

*“Incluso las torres más altas empiezan en el suelo”*

Proverbio chino

*“The more parameters to estimate, the more room  
there is to go wrong”*

James D. Hamilton

## *Resumen*

*Esta tesis se plantea la doble meta de obtener un indicador adelantado del Producto Bruto Interno y de hacerlo testeando la utilidad de Google Trends como fuente de datos . Para ello se realiza el Nowcasting del Índice de Volumen Físico del PIB uruguayo mediante la estimación de Modelos de Factores Dinámicos utilizando el algoritmo de Maximización de Expectativas (EM), y experimentando con la estimación de modelos con variables tradicionales, provenientes de fuentes oficiales y de frecuencia superior a la trimestral, como con variables de la fuente en evaluación; haciendo uso extensivo de Modelos de Factores Dinámicos, acotando el set de variables a la ventana 2011-2020, siendo 2011 el año en el que se estabiliza el flujo de datos que conforma el panel de Tendencias de Google. 2011 es también el año en el cual el uruguayo mediano obtuvo una fuente de acceso diario a internet, de modo de considerar la posibilidad de que un mayor acceso a internet pudiera aumentar la influencia de las variables de Google sobre el resultado final. Esta tesis concluye que la inclusión de variables provenientes de Google no genera una ganancia considerable a la capacidad de Nowcasting de los modelos de factores dinámicos empleados, dado el subset de términos elegidos.*

### *Palabras clave*

*Nowcasting; Modelos de Factores Dinámicos; Jagged Edge; Tendencias de Google; Econometría de Series Temporales; Predicción de corto plazo del PIB en Uruguay.*

*Clasificación JEL: C13, C32, C53*

## *Abstract*

*The rationale of this Thesis is to obtain an appropriated model in order to leading GDP while evaluating the convenience of Google Trends Query Indexes as a data source. In order to reach the goal, there are presented several Dynamic Factor Models for Nowcasting the Uruguayan Real Gross Domestic Product Index, by the Expectation Maximisation Algorithm (EM) being building forecasting models both with traditionally used-for-GDP-short-term-forecasting variables, sourced in official authorities, and with higher frequency respect to quarterly GDP, likewise Google searches. An extensive use of Dynamic Factor Models is made, starting in January 2011 to December 2019, in an attempt to give real dimension to the weight of Google Trends searches, because it was on this date that the median Uruguayan get a source of daily access to Internet, considering that an extensive access could well have improved Google Trends queries' utility. Thesis Findings about the utility of Google Trend Query Indexes are promising but not conclusive in favour of a positive contribution in the sample.*

## *Keywords*

*Nowcasting; Dynamic Factor Models; Jagged Edge; Google Trends; Time Series Econometric; Short-Term forecasting of GDP.*

*JEL Codes: C13, C32, C53*



## Tabla de contenidos

Introducción _____	12
Fundamentos Teóricos _____	14
Antecedentes y Justificación .....	14
Antecedentes del uso de Tendencias de Google .....	17
Marco Teórico .....	22
Filtro de Kalman.....	25
Modelo de factores dinámicos.....	25
Hipótesis.....	26
Estrategia de Análisis _____	27
Modelo de Análisis.....	29
Estrategia empírica.....	29
Base de datos _____	30
Tipología de modelos.....	30
Principales herramientas econométricas utilizadas.....	33
Formalización del problema en términos dinámicos .....	33
Modelos a estimar presentados en su representación de Espacio-Estado estático	34
Software econométrico utilizado.....	36
Resultados Obtenidos _____	36
Conclusiones _____	40
Limitaciones y posibles extensiones .....	40
Referencias Bibliográficas _____	42
Anexo 1-Diccionario de variables utilizadas _____	45
Tradicionales .....	45
Google .....	45
Monetarias .....	1
Laborales .....	1

Anexo 2: Transformaciones disponibles a través de la función Bpanel de “Nowcasting” _____	1
Anexo 3: Scripts _____	2
Anexo 4: Salidas del análisis de robustez _____	2

# Introducción

El objetivo de esta tesis es construir un indicador adelantado de la actividad económica uruguaya que tenga capacidad predictiva para el presente y el muy corto plazo, estrategia empírica conocida como Nowcasting, testeando adicionalmente la capacidad de adelantamiento de los términos de búsqueda cuyas series de frecuencia se encuentran disponibles gratuitamente en el portal de Tendencias de Google (Google Trends).

La predicción en tiempo real se realiza mediante el uso de diversas variables de frecuencia mayor a la trimestral, cuya publicación ocurre de forma desfasada y adelantada a la publicación de cuentas nacionales, incorporando como insumos tanto variables tradicionalmente asociadas a indicadores adelantados de la actividad económica, como variables provenientes de la base de datos de búsquedas de Google (Google Trends).

La relevancia del problema abordado se fundamenta en la necesidad de contar con información económica confiable y oportuna, de modo de brindarle tanto a los hacedores de política como a los agentes económicos en general, información relevante para la toma de decisiones resolviendo, por lo menos parcialmente, el problema resultante del retardo de aproximadamente 85 días entre la generación de datos oficiales de cuentas nacionales trimestrales, y su efectiva publicación. (BCU, 2020)

La elección del Nowcasting como herramienta para la resolución de la mencionada brecha de información se fundamenta en la existencia de pocos antecedentes para el caso nacional, entre los que destacan “Un indicador de la evolución del PIB uruguayo en tiempo real” (*Rodriguez, 2014*) y “Modelos Puentes para proyectar el PIB en el corto plazo. Enfoque sectorial” (*Brum & Rodríguez, 2016*); a su vez la utilización de Google Trends como fuente de datos fue ensayada previamente como forma de captación de incertidumbre en un contexto de indicadores de economía política (*Lanzilotta, Mordecki, & Umpiérrez, 2018*), siendo novedoso el intento por incluirlos en un ejercicio de predicción macroeconómico para el contexto uruguayo.

A pesar de que el Nowcasting es relativamente novedoso en el área de estudio de la economía, siendo tradicional como herramienta de predicción meteorológica, su uso actual está muy extendido a nivel práctico, ya que numerosos bancos centrales, tanto de América Latina como del mundo desarrollado, lo utilizan para reducir el conocimiento imperfecto de la realidad del estado actual de la economía, hecho que se refleja en la publicación de numerosos documentos de trabajo que forman parte de la literatura revisada.

Su potencial como problema teórico y empírico complejo está basado en los diversos retos intermedios que plantea: la elección de las variables a incluir, la conveniencia de las bases de datos a utilizar, la reducción de dimensiones del problema, la superación del problema de no coincidencia temporal de las publicaciones de las series, así como de la agregación de series de frecuencias diferentes, y la estimación oportuna, de modo de lograr el fin de obtener predicciones en tiempo real. Estos desafíos intermedios, que serán detallados a lo largo de este trabajo, discurren entre problemas de carácter teórico económico, econométrico y tecnológico-informacional.

La hipótesis central considera la existencia de un modelo que utiliza como insumos fundamentales variables de cierta tradición para la predicción del PIB uruguayo, junto con índices de búsqueda de Google, y que es capaz de proporcionar estimaciones de mayor frecuencia y mejor ajuste que las experiencias previas de Nowcasting y que otras técnicas de estimación como el modelado ARIMA. Alineado a esto, se buscará responder la siguiente pregunta: ¿Es Google Trends una fuente generadora de datos que colabora para mejorar la precisión de las estimaciones de muy corto plazo del crecimiento económico uruguayo?

Con el objetivo de verificar o refutar la hipótesis se lleva a cabo la estimación de una serie de modelos de Nowcasting: un Modelo de Factores Dinámicos que incluya variables clásicas de adelantamiento, y otro con factores generados a partir de un set de información compuesto por variables típicas más aquellas provenientes de Google Trends.

Adicionalmente, se realizará un análisis de validación del indicador compuesto que consistirá en el análisis de las predicciones tanto dentro como fuera de la muestra, procediendo también a una comparación respecto a otros tipos de modelos alternativos, como un modelo VAR sin reducción de dimensiones, o un modelo ARIMA sobre el PIB trimestral, siendo ambos benchmark únicamente útiles para analizar el error de predicción en frecuencia trimestral. Para referenciar la capacidad de ajuste del modelo mensual se deberá recurrir a la validación fuera de la muestra y a la comparación con un Documento de Trabajo del Banco Central del Uruguay, sobre un indicador del PIB en tiempo real. Véase Rodríguez (2014).

La metodología se enmarca en la aplicación de técnicas avanzadas de series temporales, técnicas que permitirán superar una serie de metas intermedias entre las que se incluyen las siguientes: la elección de un modelo de Nowcasting que haga un uso eficiente de los datos, que sea capaz de integrar variables con frecuencias diferentes, y publicadas en diferentes momentos del tiempo, y de abordar la existencia de datos faltantes (variables latentes) resultante del desfase en la publicación de las variables cargadas a los factores. También es relevante explicitar la importancia que juega la elección

de la fuente de datos, tanto por sus características de disponibilidad como por su riqueza a la hora de brindar datos de muy alta frecuencia.

## Fundamentos Teóricos

En las siguientes secciones se llevará a cabo una descripción sobre los antecedentes más influyentes en las diferentes líneas de investigación abarcadas por este trabajo, prosiguiendo con un resumen técnico de las metodologías empleadas, para finalizar con el planteo de las hipótesis.

### Antecedentes y Justificación

La idea de que el desempeño de la economía puede ser adelantado explotando la relación inobservable de un set de variables no es nueva. Tal y como se referencia en los trabajos de Galaso y Rodríguez López (2014) y de Rodríguez (2014), la literatura alrededor de la construcción de indicadores líderes es basta y data de los años 30, siendo los trabajos de Burns y Mitchell (1946), investigadores del National Bureau of Economic Research de Estados Unidos, pioneros en proponer la utilización de indicadores compuestos para aproximar el comportamiento del ciclo real de negocios. Sin embargo, más allá del carácter de pioneros de dichos autores, no se encontraron exentos de ser acusados como hacedores de medición desprovista de teoría, debido al carácter centrado puramente en la predicción de sus trabajos (Galaso & Rodríguez López, 2014).

El Nowcasting surge como refinamiento de la práctica predictiva en la Meteorología en la década de 1970. En ciencias meteorológicas se define Nowcasting como el enfoque mediante el cual se hace posible la predicción en tiempo real de las condiciones climáticas, desde la descripción presente hasta las siguientes dos horas, siendo tecnológicamente viable únicamente desde la confluencia del radar, la tecnología satelital y la computación para el análisis de grandes cantidades de datos (Browning & Collier, 1989).

En Economía podemos rastrear antecedentes a la denominación de la línea de investigación “Nowcasting”, palabra que habría sido adoptada por autores italianos a principios de la década del 2000, siendo el antecedente más antiguo encontrado en mi revisión, el paper titulado “Forecasting industrial production and the early detection of turning points” (Bruno & Lupi, 2004). Antecedentes anteriores deben ser rastreados por la vía de la similitud de problemas abordados, en este caso mediante las referencias a los conceptos de predicción en tiempo real y por la vía de la estimación de modelos de factores dinámicos al servicio de la construcción de indicadores de adelantamiento. No obstante, la combinación de ambos, esto es, la aplicación de modelos de factores dinámicos en la

consecución de predicciones de muy corto plazo, sienta sus bases en el trabajo de Giannone, Reichlin y Small (2008), trabajo en el que se estiman modelos de factores dinámicos, recurriéndose a una estrategia de estimación de dos pasos (“Two Stages”), donde en un primer paso se filtran las variables a utilizar mediante un Análisis de Componentes Principales, para luego reestimar sobre la base de aquellas con mayor aporte, haciendo uso del Filtro de Kalman, el cual es particularmente útil para lidiar con el problema de que no existe coincidencia en el día en el que se publican las diferentes observaciones de cada variable explicativa. El Filtro se utiliza para computar recursivamente el valor esperado de los factores comunes, como será expuesto en el siguiente capítulo de manera formal. Se trata de una versión paramétrica perteneciente a la familia de modelos de factores comunes, versión que puede ser utilizada para medir de manera explícita la precisión de los factores comunes.

Bañbura, Gianonne, Reichlin (2010) realizan una sofisticación al anterior modelo, por la vía de incorporar a la estimación una metodología capaz de lidiar con datos faltantes, por la vía del Algoritmo de Maximización de Expectativas (Expectation Maximisation en idioma original, abreviado como EM). Detalles sobre el funcionamiento del algoritmo serán abordados en el marco teórico, pero es posible resumirlo, tal cual lo hace uno de sus primeros autores, como una aproximación a la computación iterativa de estimaciones máximo-verosímiles cuando las observaciones incluyen datos faltantes o existe incompletitud en la base, la cual consiste en dos pasos: el primero, conocido como etapa de expectativas, siendo el segundo el asociado a la maximización (Dempster, Laird, & Rubin, 1977). De forma más clara, se trata de un algoritmo que actúa de forma iterativa maximizando una función de verosimilitud, de forma tal que es capaz de aprovechar la información disponible para cubrir datos faltantes. Es particularmente resaltable su eficiencia en la capacidad de extraer información de las variables integrantes del set de datos, sin la necesidad de desechar las incompletas, en el afán de completar los datos faltantes, según explican de Valk, de Mattos y Ferreira (2019).

La naturaleza del problema que da pie al uso del Nowcasting, lo hace fácil de encuadrar en un contexto de construcción de indicadores de adelantamiento, lo que en la literatura es conocido como indicadores líderes. De hecho, el ejercicio de construcción de un modelo que sea capaz de mostrar en tiempo real la situación de una economía es equivalente a la construcción de un indicador líder especializado en manejo de datos de alta frecuencia y capacidad predictiva instantánea, esto es, con el poder de brindar al usuario una proyección realista de lo que serán los datos del PIB, fijando ideas, de cierre del cuarto trimestre del 2020, con el cuarto trimestre recién cerrado. Esto brinda una ventaja enorme, por ejemplo, ante modelos ARIMA que dependan exclusivamente de los datos del PIB, en tanto una proyección hecha con los datos del Q3 estaría disponible unos 15 días antes del cierre de

Q4; mientras tanto, para esas fechas, todas las variables mensuales y de frecuencia superior ya habrían sido publicadas al menos en dos oportunidades. Esta relación lleva a considerar como antecedentes los intentos por construir índices líderes de menor refinamiento que el que perseguirá este estudio. Cabe recordar al lector que las primeras publicaciones del PIB de Estados Unidos datan del primer trimestre de 1947, con lo que, en sentido estricto, eso acota el horizonte temporal en el estudio de indicadores líderes cotejables con el PIB hasta ese momento. En ese sentido, los grandes autores de referencia son Stock y Watson (1988) quienes publicaron un artículo seminal conocido como “Un modelo de probabilidad de Indicadores Económicos Coincidentes”, donde llevan a cabo la presentación de un modelo de factores dinámicos, al que dan a conocer como modelo de series temporales explícito, donde se intenta adelantar “la situación general de la economía con un único indicador” con datos comprendidos entre 1959 y 1987. Son esos mismos autores los que citan a Burns y Mitchell (1938), como artífices de los primeros intentos por adelantar el desempeño de la economía mediante indicadores de resumen, al tiempo que se plantean como benchmark la comparación de su trabajo con el trabajo de los pioneros, que esa época aún se encontraba en uso como referencia por parte de la secretaría de comercio de Estados Unidos.

Es de entender que a medida que avanza el desarrollo del campo de la economía de series temporales, nuevas herramientas y sofisticaciones quedan a disposición de quienes se interesen en predicción. Recordemos que el trabajo de Wold es de 1938 y la metodología de Box y Jenkins de 1976, habiéndose construido una gran cantidad de modelos orientados a la predicción macroeconómica durante la década de 1960, de gran magnitud en el empleo de cientos de ecuaciones y variables, seguido de un estado de desilusión ante los resultados obtenidos en cuanto a performance, lo que dio lugar a los modelos univariados ARMA (Hamilton, 1994). La metodología de Box y Jenkins constituye un antecedente a este trabajo final, habida cuenta de que, con el fin de evitar caer en relaciones de índole espuria, la metodología de Giannone et al. requiere de la transformación de las series que las componen en estacionarias, previa aplicación del algoritmo que se crea conveniente (“Dos etapas”, “Dos etapas con agregación” o “Maximización de expectativas”). No obstante, los nuevos avances en el campo de series temporales multivariadas, y de las técnicas de reducción de dimensiones, de las cuales hacen uso frecuente los autores de modelos con capacidad de adelantamiento, junto con los avances en computación, son las que hacen posible la existencia de modelos de mayor precisión y oportunidad.

## Antecedentes del uso de Tendencias de Google

El mundo de los buscadores de internet basados en índices de búsquedas es relativamente reciente y se remonta a la aparición del algoritmo de Google en 1998. Dicha empresa requiere de la clasificación de toda la información capaz de ser indexada por su algoritmo con el fin de emparejarla con el término que el usuario ingrese en el buscador, con lo que la idea de crear series de tiempo basadas en la cantidad absoluta de búsquedas de cada término ingresado es tan antigua como Google mismo. Sin embargo, no es hasta 2004 que dichas series son publicadas, constituyendo el primer escalón para lo que sería la extensión de un gran panel de datos multinivel, donde es posible, al día de hoy, tomar series de tiempo condicionadas al país, región de interés e incluso ciudad, obteniendo una serie conocida como índice de frecuencia de búsqueda asociada a un término en particular, sensible a los tildes y las faltas de ortografía en general, excepto se trate de mayúsculas y minúsculas, en cuyo caso la búsqueda se guarda en el mismo término.

Existen diversos antecedentes a nivel internacional del uso de datos de Google en ejercicios de Nowcasting, como podemos visualizar en la Tabla 1, presentada a continuación. Es importante destacar que la variable económica a predecir varía según el trabajo, siendo posible utilizar la información de búsquedas de Google para la predicción de cortísimo plazo de variables tan distintas como la tasa de desempleo o el valor de las acciones en el mercado bursátil.

Teniendo en cuenta la fuente considerada, no es sorprendente que los pioneros en el estudio de la utilidad de los términos de Google estén vinculados a esa empresa. En lo que constituye uno de los primeros antecedentes, Choi y Varían (2011) se encargan de “mostrar cómo usar los datos de motor de búsqueda para predecir indicadores económicos en el corto plazo”. Al igual que en el caso del Nowcasting, los autores descubren que los términos de búsqueda ya se encontraban en uso en otros campos, citando el ejemplo de la epidemiología.

El uso de Google Trends cuenta con un auge en el campo de la Macroeconomía. Si bien, en los últimos tres años se pueden encontrar antecedentes de investigación estudiantil, tesis de maestrías en universidades europeas, tal y como es el caso del trabajo de Heikkinen (2019), cuya meta es testear la capacidad de Google Trends para mejorar los modelos de predicción del PIB alemán y finlandés, o el de Niesert, Oorschot, Veldhuisen, Brons y Lange (2020) que se centra en buscar mejoras para las predicciones de los índices de precios al consumo, desempleo o la confianza del consumidor para Reino Unido, Estados Unidos, Canada, Alemania y Japón; la idea de integrar estos índices de



búsquedas tiene su origen en Ettredge et al. (2005), quien intentó predecir el comportamiento de la variable desempleo, tal y como consignan Choi y Varían (2011).

En el ámbito de investigación nacional destaca la utilización de Índices de búsqueda de Google por parte de Lanzilotta, Mordecki y Umpiérrez (2018), quienes en su trabajo titulado “Political Economic uncertainty in a small & open economy: the case of Uruguay” construyen dos indicadores de volatilidad e incertidumbre para la economía uruguaya, realizando un análisis de componentes principales con el fin de seleccionar las variables a incluir en cada índice, citando el trabajo de Bontempi et al. (2016) , quienes anteceden al trabajo de las autoras por la vía de crear un indicador de incertidumbre basada en búsquedas de internet. En línea con la argumentación de Varian, Bontempi fundamenta su trabajo en la idea de que los usuarios de internet manifiestan explícitamente su incertidumbre buscando determinadas palabras con determinada frecuencia (Lanzilotta, Mordecki, & Umpiérrez, 2018).

*Tabla 1: Antecedentes del uso de Google Trends como fuente de información para la estimación de modelos económicos.*

Estudios	País	Variable Económica	Resultados principales
Choi y Varian (2009a, 2009b y 2012)	Estados Unidos, Hong Kong y Australia	Varias variables económicas	Google Trends provee información útil sobre solicitudes de desempleo, confianza del consumidor, ventas de viviendas y viajes.
D'amuri y Marcucci (2009)	Estados Unidos	Tasa de desempleo	Modelos Google tienen las predicciones más adecuadas
Suhoy (2009)	Israel	Tasa de desempleo	Búsquedas de Google proveen información adicional sobre desempleo
Askitas y Zimmermann	Alemania	Tasa de desempleo	Búsquedas de Google pueden explicar tasas de desempleo
Tuhkuri (2014)	Finlandia	Tasa de desempleo	Modelo basado en Google Trends generan las predicciones más acertadas
Anttonen (2018)	Zona Euro	Tasa de desempleo	La eficiencia inicial de los BVAR no es mejorada por búsquedas de Google
Huang y Della Penna (2009)	Estados Unidos	Confianza del Consumidor	Búsquedas de Google tienen correlación con la confianza del consumidor
Vosen y Schmidt (2011 y 2012)	Estados Unidos y Alemania	Consumo	Google Trends es capaz de explicar consumo privado
Kholodilin, Podstawski y Siliverstovs (2010)	Estados Unidos	Consumo	Google Trends produce resultados similares a los modelos de confianza del consumidor
Wu y Brynjolfsson (2015)	Estados Unidos	Mercado inmobiliario	Se encontró vinculación entre búsquedas de Google, ventas futuras inmobiliarias y precios
McLaren y Shanbhogue (2011)	Reino Unido	Mercado inmobiliario	Google Trends proveen predicciones más acertadas
Veldhuizen, Vogt y Voogt (2016)	Países Bajos	Mercado inmobiliario	El término de búsqueda "Mortgages" muestra una correlación significativa con transacciones inmobiliarias
Artola y Martínez-Galán (2012)	España	Turismo	Google Trends es útil en la generación de información sobre visitantes futuros
Preis, Reith y Stanley (2010)	Estados Unidos	Volumen del mercado de activos	Encuentra una significativa correlación entre el volumen de comercio de activos de S&P 500 y datos de Google Trends
Bank, Larch y Peter (2011)	Alemania	Liquidez y activos	Se encontró de datos de Google Trends respecto al volumen de ventas de valores y liquidez.
Perlin, Caldeira, Santos y Pontuschka (2017)	Australia, Canada, Reino Unido y Estados Unidos	Mercados financieros	Google Trends proporciona información adicional sobre mercados bursátiles
Koop y Onorante (2013)	Estados Unidos	9 variables macroeconómicas diferentes	La utilización de datos de Google Trends mejora la exactitud de la predicción
Donadelli (2015)	Estados Unidos	Incertidumbre política	Datos de Google generan información similar a otros indicadores de incertidumbre
Gotz y Knetsch (2019)	Alemania	Producto Bruto Interno	Datos de encuestas tienen mejor performance que los de Google Trends
Ferrara y Simoni (2019)	Zona Euro	Producto Bruto Interno	Los modelos más exactos <b>no</b> incluyen datos de Google Trends

Fuente: Heikkinen, Joni (2019)

Los modelos de factores dinámicos han sido utilizados en Macroeconomía y Finanzas desde su introducción por Geweke (1977) Sargent y Sims (1977), como una manera de capturar las correlaciones entre las variables (Forni, Hallin, Lippi, & Reichlin, 2000).

Como se mencionó previamente, el primer trabajo en materia de indicadores líderes de Stock y Watson se trató de la presentación de un modelo de factores dinámicos, donde se hace uso del Filtro de Kalman, de modo de que es este filtro el que se encarga de construir, de forma recursiva, la estimación con mínimo error cuadrado medio de un vector de variables de estado inobservables. (Stock & Warson, 1988)

Más cercano en el tiempo, los ya citados Forni, Hallin, Lippi y Reichlin (2000) incursionan en el planteo de un Modelo de factores dinámicos generalizados. Se trata de un artículo de marcado perfil econométrico teórico, con el planteo de distribuciones en el límite, y testeos teóricos de similar alcance. El ejercicio de aplicación planteado consta de la construcción de un indicador de coincidencia para el espacio monetario del euro construido en base a la media ponderada de los factores comunes de los respectivos PIB de los países dentro de la unión monetaria, con base en un set de indicadores macroeconómicos abarcativo, convenientemente reducido mediante la metodología planteada.

Si revisamos a nivel bancocentralista, Varían y Scott (2014) destacan que el conjunto de datos de Google Trends es utilizado por la Reserva Federal de San Luis (FRED, Estados Unidos) en sus ejercicios de Nowcasting mediante métodos bayesianos, así como también se puede encontrar un antecedente para el caso chileno en Carrière-Swallow y Labbè (2010) , trabajo en el que se realiza un testeo sobre la capacidad de mejora de la eficiencia del Nowcasting de venta de autos mediante la construcción de un indicador adelantado que busca medir la intención de comprar un vehículo, construido en base a datos de búsquedas de Google, hallándose mejoras en la capacidad predictiva del modelo tanto dentro como fuera de la muestra. Un grupo de antecedentes señalados por Heikkinen, previamente citados también por Choi y Varían (2011), enfatizan en la utilización de datos de Google para la predicción de una variada gama de variables económicas como el desempleo, destinos de vacaciones o demanda automotriz. En otras experiencias destacan Guzmán (2011), quien examina la capacidad predictiva de los datos de Google sobre la Inflación; el trabajo de Baker y Fradkin (2011), quienes utilizan la base para testear cómo las búsquedas de trabajo se ven afectadas por extensiones del subsidio por desempleo; y el trabajo de Shimshoni et al. (2009), que describe la

capacidad predictiva de la base de búsquedas de Google, señalando que una buena cantidad de términos son fácilmente predecibles usando métodos de descomposición estacional simples (Choi & Varian, 2011).

El presente trabajo toma como precursor el trabajo seminal de la técnica de Nowcasting, aportado por Gianonne, Reichlin y Small (2008), donde se lleva a cabo una demostración de la utilidad de variables de mayor frecuencia que el Producto Bruto Interno para mantener a la autoridad monetaria bien informada sobre cambios en la actividad económica. Con dicho fin, los autores hacen uso de modelos puentes, integrados por unas pocas variables mensuales con capacidad predictiva del crecimiento del PIB, explotando la colinealidad entre las mismas, de modo de reducir la dimensión del panel a unos pocos factores comunes que capten el grueso de las interacciones entre las series, y provean un modelo que se comporte adecuadamente en cuanto a la predicción. Un segundo paper de referencia del ámbito europeo, es Bańbura, Gianonne y Reichlin (2010), profundizando en el estudio de los paneles desbalanceados; esto es, haciendo un intento por conciliar la capacidad predictiva de variables de alta frecuencia respecto al PIB, maximizando la información aportada por las mismas, y resolviendo el problema de la heterogeneidad, ya no solo de las frecuencias, sino también de la desincronización en las fechas de publicación de las diferentes series. Este panel desbalanceado tiene una forma particular que la literatura define como “Panel de bordes irregulares” o Jagged Edge (a veces también referido como Ragged Edge, literalmente *“de bordes precarios o desiguales”*). Otro aporte trascendental de este artículo yace en la técnica de estimación del Modelo de factores dinámicos exactos, que supera la propuesta inicial de Componentes principales, para adentrarse en la aplicación del algoritmo de maximización de expectativas. Sobre ambos se profundiza en posteriores capítulos de esta tesis.

A partir de todos los antecedentes planteados surge la hipótesis a chequear sobre la conveniencia de la inclusión de un bloque de variables provenientes de Google Trends, innovando para el caso uruguayo. A su vez, la combinación de los diversos antecedentes brinda señales sobre los términos que se deberían elegir de cara a conformar el set de variables de Google.

## Marco Teórico

Nowcasting se define como la predicción del tiempo presente, del pasado o del futuro muy próximo en tiempo real. (Bańbura M. , Giannone, Modugno, & Reichlin, 2013) Se trata de una contracción para las palabras Now [ahora] y Forecasting [Predicción], cuyo origen se remonta a la predicción instantánea del clima en el campo de la Meteorología, siendo reciente en el ámbito de la Economía. (Giannone, Reichlin, & Small, 2008)

En términos matemáticos, un problema de Nowcasting de  $y_t^Q$  se define como la proyección ortogonal de la variable observable (variable a la que se realiza el Nowcasting, en este caso el Índice de Volumen Físico del PIB uruguayo, observado trimestralmente, de ahí el índice Q) en el set de información disponible  $\Omega_v$  (donde v es por vintage, o "generación"), a través de regresores conocidos como factores comunes, los cuales son estimados de un set de datos de frecuencia mensual. Debido a que el objetivo del modelo es explotar la mayor cantidad de datos disponibles, se presenta la necesidad de conciliar una especificación parsimoniosa, pero al mismo tiempo retener el poder predictivo del grupo mayor. Este problema, conocido en la literatura como Reducción de dimensiones, será abordado por la vía del resumen de información a través de una cantidad inferior de factores comunes, tal y como llevan a cabo Giannone, Reichlin y Small (2008).

Al respecto, es de orden presentar, en términos de Stock y Watson (2011), la definición básica de Modelos de Factores Dinámicos, modelización que permite la reducción de dimensiones y que será usada en este trabajo.

Partiendo del siguiente sistema de ecuaciones, procederé a transcribir la formulación matemática aportada por los referidos autores.

$$X_t = \lambda(L)f_t + e_t \quad (1)$$

$$f_t = \Psi(L)f_{t-1} + \eta_t \quad (2)$$

Siendo  $X_t$  la variable objetivo a modelar,  $f_t$  el vector [qx1] de factores dinámicos utilizados para modelar la variable objetivo,  $\lambda(L)$  se define como la matriz [Nxq] de carga o ponderación de los factores dinámicos. En la segunda ecuación se puede observar la modelación de los factores dinámicos como un proceso autorregresivo VAR (1) más un término de error  $\eta_t$  [qx1] . El vector  $e_t$

$[N \times 1]$  representa las perturbaciones idiosincráticas del modelo, pudiendo incluir errores de medida de las variables y características específicas que afectan a cada serie individual y responde a perturbaciones no modeladas por el componente común de las series,  $\lambda(L)f_t$ .

Algunos supuestos son necesarios en orden de asegurar un buen comportamiento del modelo, como la necesidad de que los procesos descritos en ambas ecuaciones sean estacionarios, que los términos  $\eta_t$  y  $e_t$  sigan una Distribución Gaussiana de media cero y se encuentren incorrelacionados para todo  $(t,k)$ ,  $E(e_t \eta_{t-k}) = 0$ .

La ecuación 3 hace explícita la capacidad del modelo planteado de adelantar el comportamiento de la variable de interés, a partir de una combinación lineal de factores más una combinación lineal de retardos.

$$\begin{aligned}
 E[X_{it+1} | X_t, f_t, X_{t-1}, f_{t-1}, \dots] &= E[\lambda_i(L)f_{t+1} + e_{it+1} | X_t, f_t, X_{t-1}, f_{t-1}, \dots] \quad (3) \\
 &= E[\lambda_i(L)f_{t+1} | X_t, f_t, X_{t-1}, f_{t-1}, \dots] + E[e_{it+1} | X_t, f_t, X_{t-1}, f_{t-1}, \dots] \\
 &= E[\lambda_i(L)f_{t+1} | f_t, f_{t-1}, \dots] + E[e_{it+1} | e_{it}, e_{it-1}, \dots] \\
 &= \alpha(L)f_t + \delta(L)X_{it}
 \end{aligned}$$

Los modelos que serán estimados a lo largo de esta tesis seguirán el formato de la derivación de la ecuación (3), emulando el modelo unifactorial de Engle y Watson (1981), la memoria del VAR empleado, o la cantidad de bloques utilizados para la estimación de la variable objetivo.

La información sobre la que se estiman los factores comunes tiene su origen en el panel  $\Omega_v$ , el cual posee la particularidad de estar integrado por variables de diversas frecuencias y que tienen la característica de no publicarse en el mismo momento dentro del periodo de tres meses, que se corresponde con el ciclo de la frecuencia de la variable objetivo del Nowcasting, motivo por el cual se lo califica como un panel desbalanceado. Esto, a pesar de dificultar el problema de agregación (en comparación con el estado de naturaleza ideal donde todas las variables observables son publicadas con frecuencia mensual y perfecta sincronía), constituye una ventaja en cuanto permite ir ajustando la proyección, incorporando actualizaciones al modelo con la publicación de cada variable, elemento que refuerza al Nowcasting en su capacidad predictiva dinámica, y le otorga flexibilidad y habilidad para incorporar información de muy reciente publicación, rápidamente, mejorando la predicción para el periodo de referencia mensual y, por extensión, trimestral.

Al respecto de este mecanismo de actualización constante, la dinámica matemática detrás de este proceso es explicada en detalle a continuación, siguiendo a Bańbura, Giannone, Modugno y Reichlin (2013) :

Sean  $\Omega_v$  y  $\Omega_{v+1}$  dos colecciones consecutivas de información que difieren por dos motivos:  $\Omega_{v+1}$  contiene variables que no habían estado disponibles en  $\Omega_v$  y, además, alguna de las variables contenidas en  $\Omega_v$  ha sufrido algún tipo de actualización o revisión. En notación de conjuntos, las condiciones impuestas implican que  $\Omega_{v+1}$  es una expansión de  $\Omega_v$ , o lo que es equivalente,  $\Omega_v \subset \Omega_{v+1}$ . Recordemos que el subíndice hace referencia a un *vintage* o realización específica del set de información, y no a su fecha de publicación. Es de recibo hacer dicha aclaración, habida cuenta dentro de cada set conviven observaciones de variables publicadas en diferentes momentos del tiempo y con frecuencias diferentes, pudiendo ser anuales, trimestrales, mensuales, o incluso diarias. Además, se debe aclarar que la observación de cada variable incluida en el set adolece de retraso entre su medición y su publicación efectiva. Todas esas puntualizaciones hacen necesaria la existencia de una notación lo suficientemente flexible que no naufrague en inconsistencias.

La siguiente ecuación describe el proceso de actualización:

$$\mathbb{E}[y_{t,1}^{k_1} | \Omega_{v+1}] = \mathbb{E}[y_{t,1}^{k_1} | \Omega_v] + \mathbb{E}[y_{t,1}^{k_1} | A_{v+1}] \quad (4)$$

Nueva Predicción      Vieja Predicción      revisión

Donde  $A_{v+1} = (a_{v+1,1} \cdots a_{v+1,J_{v+1}})'$ ,  $a_{v+1,j} = y_{t,j,n_j} - \mathbb{E}[y_{t,j,n_j} | \Omega_v]$ ,  $j = 1, \dots, J_{v+1}$ .  $A_{v+1}$  representa la parte de las actualizaciones de variables que son ortogonales a la información ya contenida en  $\Omega_v$ . Se trata de la parte que el modelo no estaba considerando. En terminología de Bańbura et al., , (2013),  $A_{v+1}$  son las novedades.

El problema, como fuera introducido al inicio de este trabajo, gira entorno a la construcción de un indicador capaz de adelantar el comportamiento de la variable PIB. El Nowcasting, como problema de frontera en el análisis de series temporales, exige enfrentar una variedad de problemas intermedios, a saber: enfrentar la elección de variables a incluir, lidiar con un problema de agregación temporal generado a partir de las diferencias de frecuencia entre las series y su no coincidencia en momento de publicación, y la elección del modelo de Nowcasting.

## Filtro de Kalman

En este trabajo utilizaré el filtro de Kalman como herramienta para obtener una estimación de la proyección de los regresores basada en su esperanza condicional al set de información. Siguiendo la definición de Hamilton (1994), el filtro de Kalman es un algoritmo que actualiza secuencialmente la proyección lineal de un sistema. Entre otros beneficios, este algoritmo provee una manera de estimar vectores autorregresivos que varían en el tiempo, lo que nos será útil en el ejercicio de Nowcasting, ya que modelizaremos los factores comunes como un VAR.

Se debe adicionar que el Filtro de Kalman es un procedimiento recursivo para el cómputo óptimo estimado del vector de estado, que recordemos está expresado en la ecuación (2), en el momento  $t$ , que se nutre de la información disponible hasta el momento de en el que se ejecuta la predicción.

En notación de Bańbura et al (2013):

$$X_{t|\Omega_v} = \mathbb{E}_\theta[X_t|\Omega_v], P_{t|\Omega_v} = \mathbb{E}_\theta[(X_t - \mathbb{E}_\theta[X_t|\Omega_v])(X_t - \mathbb{E}_\theta[X_t|\Omega_v])'] \quad (5)$$

Nótese que se trata de una proyección lineal de la esperanza de la variable de estado condicional en el set de información.

Bańbura et al (2013) resaltan la capacidad del Filtro de Kalman de obtener proyecciones tanto de las variables observadas como de las de estado, así como también de su capacidad para adaptarse a los problemas básicos que enfrenta el Nowcasting, como la posibilidad de incluir variables de diferente frecuencia y no coincidencia de fechas de publicación. Los referidos autores destacan que la formalización propuesta por ellos es un avance que supera la simplicidad de las viejas ecuaciones puente, dotando a la metodología de una rigurosa base en modelos donde es posible evaluar la forma en la que las distintas variables son importantes en señalar cambios en la actividad económica.

## Modelo de factores dinámicos

Dada la naturaleza del problema de Nowcasting, donde el set de información considerado incluye una multiplicidad de series temporales, debemos recurrir a un método que logre extraer los componentes principales que luego sean útiles para realizar la estimación, y a la vez logre sortear el desafío de la dimensionalidad, lo cual se consigue reduciendo la cantidad de predictores a estimar. En ese sentido, el Modelo de Factores Dinámicos (DMF por su expresión en inglés) constituye el marco teórico de



reducción de información de un gran set de variables a uno de menor cantidad de series temporales, tal y como explican Stock y Watson (2020).

Siguiendo la definición de “Introducción a la Econometría” (Stock & Watson, 2020), se define Modelo de Factores dinámicos como una representación de  $N$  variables de series temporales, donde cada una es expresada como la suma de un número reducido ( $r$ ) de factores comunes inobservables, más  $q$  shocks idiosincráticos incorrelacionados con los factores y con los shocks idiosincráticos de las demás variables. En la práctica,  $r$  y  $q$  serán elegidos por criterios de información. Una segunda forma de definir DMF es considerando un modelo de este estilo como una forma de capturar correlaciones dinámicas y de sección cruzada entre las múltiples series de una manera parsimoniosa. (Fiorentini, Galesi, & Sentana, 2018)

Un modelo de factores dinámicos es capaz de modelar los comovimientos de un grupo grande de series temporales, utilizando para ello la relación entre un pequeño grupo de variables inobservables, los factores comunes. Existen alternativas a esta elección, como la elegida por Scott y Varían (2013) (2014). En dichos casos se opta por modelos de tipo estructurales bayesianos complementados por métodos de computación de estimación como el de Monte Carlo, tópicos que exceden los alcances de este trabajo final. La ventaja del modelo que finalmente se ha decidido estimar consiste en una mayor simplicidad teórica, sumada a la hipótesis de que el algoritmo de Scott y Varían sería de extrema utilidad si contáramos con los datos diarios de búsquedas para Uruguay, que al momento de elaboración de esta tesis no son de publicación libre.

Tal y como realizan Bańbura, Giannone y Reichlin (2010), en este trabajo recurriré a una estrategia de dos pasos consistente en la estimación de un Modelo de Factores comunes y luego en la utilización de los Factores comunes en la estimación de la ecuación (3) y (5), las cuales difieren únicamente en el hecho de que (3) está definida en notación matricial, en tanto (5) hace referencia a una de las componentes de (3).

## Hipótesis

La hipótesis central considera la existencia de un modelo que utiliza como insumos fundamentales índices de búsqueda mediante Google, y que es capaz de proporcionar estimaciones que superen en precisión y frecuencia de ejercicios anteriores de Nowcasting nacional, y al resto de modelos

obtenidos en este ejercicio. Alineado a esto, se buscará responder la siguiente pregunta: ¿Es Google Trends una fuente generadora de datos que colabora con mejorar la precisión de las estimaciones de muy corto plazo del crecimiento económico uruguayo?

## Estrategia de Análisis

En este trabajo seguiré la estrategia de Bańbura, Giannone y Reichlin (2010), en el sentido de que, al igual que lo hacen los autores, traduciré el problema de la no coincidencia de fechas de publicación y la heterogeneidad de frecuencias de las series como un problema de variable latente no observada. En ese sentido, por ejemplo, a las variables trimestrales se las considera como variables mensuales con 2 observaciones perdidas. En este ejercicio la variable de frecuencia baja por excelencia es la transformación que hace estacionaria al PIB luego de un modelado mediante el uso de la función “Auto.Arima” correspondiente al paquete Forecast de R. En este caso se trata de la diferencia trimestral de la diferencia anual del IVF previamente estandarizado.

El set de datos es cargado a R, estandarizado y modelado también mediante la función “Auto.Arima”, y se obtiene para cada variable la información de qué transformaciones son requeridas para volver a cada una de ellas estacionarias. El programa de mi autoría se nutre de dicha información y la transforma en un vector de transformaciones, asignando a cada variable un código alineado al que se proporciona en la tabla de transformaciones en el Anexo 1. Será a partir de esta codificación que la función BPanel, correspondiente al paquete Nowcasting de R, se encargará de transformar a cada variable en estacionaria, balanceando el panel.

Posteriormente se aplica la función Nowcast, del paquete Nowcasting al igual que la anterior, especificando que el algoritmo a emplear será el de Maximización de Expectativas. Para que dicha función actúe es necesario establecer una matriz de bloques, que en el caso de este trabajo coincide con la fuente de cada bloque empleado en el modelo en cuestión, aunque existen antecedentes como el que presentan Ferreira et al (2019), donde cada factor se particiona en tres grupos: global, real y nominal. Los bloques que conforman este trabajo empírico son los que se detallan a continuación:

- 1- Bloque de variables tradicionales, de diversas fuentes.
- 2- Bloque de Variables de Google Trends
- 3- Bloque de variables monetarias, fuente Banco Central del Uruguay

- 4- Bloque de variables de desempeño del mercado laboral, fuente Instituto Nacional de Estadística.

Se recurrirá al empleo de Filtro de Kalman para la mejora dinámica de las predicciones a medida que nueva información es publicada. Finalmente, siguiendo a Gianonne et al. (2008) se procederá a estimar un modelo de factores dinámicos, cuyo empleo permite atacar dos desafíos al mismo tiempo: reducir la cantidad de estimadores a estimar, y superar el problema de selección de variables mediante el empleo de los factores comunes. La representación de Factores Dinámicos dialoga muy bien con la introducción del filtro de Kalman y permite incorporar con facilidad los cambios dinámicos en el set de información.

El testeo el grado de aporte individual de las variables provenientes de Google Trends se realizará mediante un ejercicio de robustez de tipo leave-one-out, donde se procederá a estimar cada modelo dejando por fuera una variable de Google a la vez. El testeo general de la utilidad del bloque de variables de términos de Google se realizará mediante la estimación de varios modelos que no incorporen dicho bloque en su conjunto.

Con la intención de testear la utilidad de los datos provenientes de Google Trends, la estrategia consiste en seguir el procedimiento propuesto por Gianonne et al (2008), estimando modelos con la muestra entera (2011 a 2020) incluyendo variables de Google Trends, y luego realizando predicción fuera de la muestra con el fin de visualizar cómo afecta la cadencia de publicación de las diferentes variables a la mejora del Nowcasting.

El ejercicio inicia con la aplicación de la metodología propuesta por Bańbura et al (2010), en este caso estimando  $\theta$ , un vector que recopila todos los parámetros del modelo de factores, a través de lo que se conoce como algoritmo de maximización de expectativas. Este algoritmo consta de dos pasos, según describen los autores: un primer paso está asociado al cálculo de la esperanza de la función de logverosimilitud condicional a los datos, que se calcula de manera recursiva, basándose en la anterior estimación de Theta,  $\theta_j$ . Cabe aclarar que  $\theta_0$  surge de un análisis de componente principal. Un segundo paso consiste en estimar el valor de los parámetros en el siguiente periodo,  $\theta_{j+1}$ , para lo cual se recurre a la maximización de la esperanza de la función de verosimilitud respecto al vector de parámetros  $\theta$ . La estructura de bloques junto al desglose de componentes de  $\theta$ , y una profundización del algoritmo EM pueden revisarse en profundidad en los anexos B y C de Bańbura et al. (2010). Para la estimación de los modelos se empleará el software R, haciendo uso intensivo del paquete “Nowcasting” de De Mattos, De Valk y Ferreira (2019), quienes realizan un procedimiento muy

similar al planteado en esta tesis, pero sin testear la validez de una base de datos adicional a las variables tradicionales de Nowcasting tanto para el caso estadounidense (replican los artículos de Giannone et al), como para el caso brasileño.

Para comparar el ajuste de los modelos, se recurrirá a la comparación de la raíz cuadrada del error cuadrático medio de los modelos estimados dentro de la muestra, y a un análisis de predicción fuera de muestra.

## Modelo de Análisis

### Estrategia empírica

La estrategia empírica consta de los siguientes pasos:

Revisión bibliográfica introductoria a la temática de Nowcasting, construcción de indicadores líderes, estimación de modelos de factores dinámicos y uso de Google Trends.

En esta etapa se procedió a la búsqueda por palabra clave de publicaciones arbitradas mediante las plataformas Timbó y Google Scholar. La finalidad ha sido recabar datos acerca de la manera en la que los investigadores abordan el problema de Nowcasting y las variables que emplean en sus modelos de Nowcasting del PIB. A lo largo de esta revisión se recabó información útil a la hora de seleccionar las primeras variables a incluir en el Modelo, así como se está avanzando en el estudio de herramientas de modelización y de selección de variables, las cuales demandarán una búsqueda especial de antecedentes en modelización mediante Factores dinámicos. Debido a la popularidad con la que cuenta el Nowcasting entre Bancos Centrales, he decidido incluir documentos de trabajo publicados por Bancos Centrales, aun cuando no hayan sido arbitrados. La naturaleza del tema exige tener presente teoría y práctica por igual.

Elección de las variables a incluir. Para resolver este pendiente se procedió a incluir las variables que típicamente se incluyen en la predicción del PIB uruguayo, tomando como referencia la elección realizada por Rodríguez (2014), pero no limitándose a ella. De la misma manera, para el caso de las variables de Google Trends se sigue un criterio de elección por juicio humano. El mismo se basa en la relación entre término de búsqueda y la variable que se pretende adelantar. Es ente sentido, parece razonable tratar de adelantar las ventas de autos, variable muy considerada a la hora de tratar de

adelantar el PIB, por el término “Autos” o por la búsqueda de “Automotora”. En aquellos casos donde haya más de un término que pueda ser de buen criterio incluir, como por ejemplo puede ocurrir en el intento de adelantar las cifras de ventas de casas, optaré por la construcción de un índice compuesto de los términos pertinentes, los cuales podrían ser, en ese caso, “Inmobiliaria” y “Casas”. Recordemos que la metodología de estimación será la que finalmente decida qué variable tiene poder explicativo suficiente para integrar el indicador compuesto final y con qué peso, por lo tanto, estas variables, hasta efectuada la estimación, son potenciales variables incluidas, excepto sus correlaciones respecto al PIB sean exactamente cero, en cuyo caso la variable es excluida.

En tercer lugar, se procedió a realizar un análisis sobre la estacionariedad de las series tentativas de ser incluidas, habida cuenta que el procedimiento requiere que las mismas sean estacionarias. Para ello se procedió a su modelación mediante la función “Autoarima” del paquete Forecast de R, complementado con una prueba de raíz unitaria en frecuencia principal (Test Aumentado de Dickey Fuller), incluido en la función. Adicionalmente se recurrió al análisis gráfico de las mismas, y en casos de dudas sobre la estacionariedad se siguió el siguiente orden de acciones: gráfico de las funciones de autocorrelación y de autocorrelación parcial.

Así mismo, se realizó la estandarización de las series incluidas ya que el análisis de componente principal, realizado en el primer paso de los modelos que aplican la metodología de dos pasos, no es invariante en la escala, y, por tanto, de no efectuarse, las series con mayor varianza dominarían a las de menor. Las estandarizaciones se realizaron con las medias y las varianzas del periodo 2011-2020, dando esto como resultado la existencia de dos sets según periodo de tiempo que incluyera el modelo. Al análisis de estacionariedad le sigue la determinación sobre qué transformación es requerida para hacerla estacionaria, según el criterio de los autores de Valk, de Mattos y Ferreira (2019).

#### Base de datos

La conformación de la base de datos siguió una estrategia de Banco de Datos, por oposición a la idea de Big Data. Para conformar los 4 bloques que finalmente fueron incluidos se siguió un criterio de fuente de datos para el caso de las variables de Google, monetarias del BCU y de empleo del INE. La excepción la conformó el bloque de variables tradicionales, donde se agruparon diez variables de diversas fuentes. Por detalles sobre las variables, sus fuentes y rezagos de publicación, en el anexo 1 se adjuntan detalles.

#### Tipología de modelos

Los modelos sobre los que finalmente se presentarán resultados se definen de la siguiente manera:

Modelos estimados según bloques incluidos.						
Nombre de los bloques						
		Tradicional	Google	Monetarias (BCU)	Laborales (INE)	Orden VAR
Grupo 3	Modelo 7	✓		✓		1
	Modelo 8	✓		✓		2
	Modelo 9	✓		✓		3
Grupo 4	Modelo 10	✓			✓	1
	Modelo 11	✓			✓	2
	Modelo 12	✓			✓	3
Grupo 1	Modelo 1	✓	✓			1
	Modelo 2	✓	✓			2
	Modelo 3	✓	✓			3
Referencias:						
✓ Bloque incluido						

En el plano puramente estimativo se procede a los siguientes pasos:

1. Se estimaron 5 grupos de modelos, donde se fueron variando los bloques de variables incluidas bajo el criterio de agregar bloques y revisar los cambios en el error de estimación que la inclusión de estos generaba. A su vez se estimaron 3 modelos por cada grupo, variando el orden de integración del VAR en los factores entre 1 y 3. En todos los casos el set de variables se limita a la ventana enero de 2011- diciembre de 2020. Si bien el método y los resultados derivados del paquete empleado lo permite, no serán considerados ni analizados los resultados del Nowcasting al inicio de la muestra, en la ventana enero de 2011 a junio de 2012, centrándome efectivamente en los resultados estimativos para el primer trimestre de 2021.

Modelos estimados según bloques incluidos.						
Nombre de los bloques						
		Tradicionales	Google	Monetarias (BCU)	Laborales	Orden VAR
Grupo 1	Modelo 1	✓	✓			1
	Modelo 2	✓	✓			2
	Modelo 3	✓	✓			3
Grupo 2	Modelo 4	✓		✓	✓	1
	Modelo 5	✓		✓	✓	2
	Modelo 6	✓		✓	✓	3
Grupo 3	Modelo 7	✓		✓		1
	Modelo 8	✓		✓		2
	Modelo 9	✓		✓		3
Grupo 4	Modelo 10	✓			✓	1
	Modelo 11	✓			✓	2
	Modelo 12	✓			✓	3
Grupo 5	Modelo 13	✓	✓	✓	✓	1
	Modelo 14	✓	✓	✓	✓	2
	Modelo 15	✓	✓	✓	✓	3
Referencias:						
✓ Bloque incluido						

Cabe aclarar que las series no siguieron el mismo tratamiento que realizaran Stock y Watson (2016) por el hecho de que la ventana elegida, consecuencia de la alta prevalencia de ceros en las series de la fuente de Google anterior a 2011, es muy acotada en el tiempo, y la exclusión de tendencias de largo plazo por el método de filtrado de doble peso con bandwidth corto no hubiera resultado procedente.

2. Se estiman los modelos de Factores dinámicos aproximados, siguiendo el algoritmo de maximización de expectativas. En cada caso se emplean los periodos ya definidos y se realizan con todas las variables cuyas correlaciones no sean cero respecto al PIB.
3. Comparación entre las predicciones fuera de la muestra y del comportamiento de los modelos estimados entre sí. En esta etapa se procederá a la comprobación de la capacidad predictiva de los modelos de Nowcasting, tanto entre sí como respecto a un modelo benchmark que en este caso es un autoarima (0,1,1) (0,1,0) sobre el IVF estandarizado. Para la comparación entre modelos se emplea el Criterio de la raíz del error cuadrático medio, dentro y fuera de la muestra.
4. Se procede a un análisis leave-one-out sobre las variables integrantes del bloque de Google Trends. Para ello se estiman 32 modelos por cada una de las 6 especificaciones que contienen el referido bloque. El resultado de esta etapa son dos matrices con el error cuadrático medio

dentro y fuera de la muestra, con el agregado de los respectivos errores de predicción de los modelos estimados con todas las variables, según corresponda.

## Principales herramientas econométricas utilizadas

La econometría aplicada se ubica en el campo de las series temporales. Se realizó una preselección de variables a incluir en base a antecedentes, conformándose de esa manera el bloque de variables tradicionales. Luego se complementó el panel con la agregación de términos de Google. La lista de términos preseleccionados se basó en incluir términos asociados al consumo masivo, como ser portales de compras frecuentemente utilizados por uruguayos, nombre de productos masivos y pequeñas variantes en su forma de escritura, términos empleados en la búsqueda de trabajo, subsidios, apoyos económicos y organizaciones públicas asociadas a esos vocablos, como ser ministerios o empresas públicas. Finalmente se incluyeron términos relativos a la búsqueda de descuentos y ofertas, recomendados por H. Varían como buenos indicadores adelantados del ciclo, así como términos que se vinculen con algunas de las principales fuentes de inestabilidad de la economía doméstica, a saberse: dolar (sin tilde por ser la variación con mayor correlación respecto a la variable relevante), Argentina, Brasil, , inflación (e “inflación” sin acento debido a la generación de series diferentes por parte de Google ante ambos términos).

## Formalización del problema en términos dinámicos

A continuación, se presenta la representación de espacio-estado de un sistema dinámico genérico siguiendo la notación de Bańbura, Giannone, Modugno & Reichlin. (2013)

$$Y_t^{KY} = \mu + \zeta(\theta)X_t + G_t, G_t \sim i. i. d. N(0, \Sigma_G(\theta)) \quad (6)$$

$$X_t = \varphi(\theta)X_{t-1} + H_t, H_t \sim i. i. d. N(0, \Sigma_H(\theta)) \quad (7)$$

Donde la ecuación (6) es conocida como Ecuación de observación (de señal o de medida), debido a que relaciona la variable observada  $Y_t^{KY}$  con la variable latente inobservada de estado  $X_t$ . La ecuación (7) es conocida como Ecuación de transición, o de estado, porque define como varía en el tiempo la variable latente (Scott & Varian, 2013). La notación de las ecuaciones se encuentra simplificada, debiéndose alertar que tanto los coeficientes de la variable latente como los componentes de la matriz de covarianzas de las perturbaciones pueden variar en el tiempo. El supuesto de normalidad de las perturbaciones es necesario en orden de poder aplicar el Filtro de Kalman.



Tal y como he presentado previamente, el problema consiste en realizar predicción de muy corto plazo, utilizando variables de frecuencia mensual, a una variable de frecuencia superior, en este caso el IVF del PIB uruguayo. En ese sentido, para poder llegar a buen puerto se debe dejar sentada la característica real de que la variable PIB, aunque observada trimestralmente, es una variable de flujo que podría observar en intervalos de amplitud menor si fuéramos capaces de realizar mediciones de mayor frecuencia. De hecho, el PIB que observamos trimestralmente no es más que la suma de los tres PIB mensuales no observados. Con este argumento tan simple es que podemos pensar el problema como uno de variables omitidas, donde no somos capaces de observar las mediciones de determinados meses del PIB.

$$PIB_t^Q = PIB_{t-1}^M + PIB_{t-2}^M + PIB_{t-3}^M \quad t=3, 6, 9, \dots \quad (8)$$

Tomando como punto de partida que la variable a estimar será el PIB estandarizado dentro de la muestra, al cual se le realizan manualmente las transformaciones necesarias para volverla estacionaria, lo que implica en este caso la realización de una diferencia trimestral y otra anual. serie que obtengo de la siguiente manera:

$$PIB_t^e = (PIB_t^Q - \mu)/\sigma \quad (9)$$

$$y_t^q = \Delta\Delta_{12} PIB_t^e \quad (10)$$

Finalmente, la mensualización de la serie se realiza a través de la función “qtr2month” de la librería Nowcasting, de modo que los datos del primer trimestre del año pasan a corresponder al mes 3 del trimestre, en tanto que los meses 1 y 2 son completados automáticamente por la función con NA.

## Modelos a estimar presentados en su representación de Espacio-Estado estático

Las siguientes ecuaciones reescriben el problema en términos matriciales, siendo conocida esta expresión como la forma estática del problema, debido a que la notación empleada en esta ocasión genera la falsa idea de que los factores son calculados en un único momento  $t$ , aunque dentro de la matriz  $F$  se encuentren tanto factores contemporáneos como retrasados, según explican Stock y Watson (2011).

$$Y_t = \mu + \Lambda F_t + E_t, \quad E_t \sim i.i.d. N(0, \Sigma_E) \quad (11)$$

$$F_t = \phi(L)F_t + U_t, \quad U_t \sim i.i.d. N(0, \Sigma_U) \quad (12)$$

Al respecto del anterior planteo, es necesario que  $U_t$  se iguale con  $\eta_t$ , para lo que es necesario multiplicar  $\eta_t$  por una matriz de ceros y unos elegida específicamente para dicho propósito.

Al igual que en Giannone, Reichlin, & Small (2008), mi vector  $Y_t$  únicamente contendrá en primera instancia, variables de frecuencia mensual observadas. A partir de ello obtendré factores comunes que deberán ser reestimados en un segundo paso, explotando la capacidad del Filtro de actualizarse óptimamente a medida que se obtienen nuevas novedades.

La siguiente ecuación (13) es una muestra de lo que Bańbura et al (2013) definen como Modelo puente factorial. Esta ecuación hace referencia a la más importante componente del vector  $Y_t$ . La fase de estimación expuesta en (13) nos permite visualizar con más claridad la forma en la que la variable crecimiento económico es regresada a través de factores dinámicos. Es de precisión aclarar que, dado el método de estimación determinado, EM, los modelos puentes quedan por fuera del alcance de esta tesis, sin embargo, la ecuación (13) es también muy ilustrativa con el fin de visualizar la forma en la que el set de actualizaciones colabora en mejorar la estimación de la variable objetivo.

$$y_{t,1}^{k_1} = \alpha + \beta F_{t|\Omega_v}^{k_1} + e_t^{k_1}, \quad t = k_1, 2k_1, \dots \quad (13)$$

Donde  $Y_t$  es un vector que contiene todas las variables de más alta frecuencia y posee una estructura de factor que se refleja en la ecuación (12) mediante la presencia de la variable de estado  $F_t$ . Asimismo, la variable de estado sigue un proceso VAR, que en la sección de estimación de esta tesis oscila entre modelos de orden 1 y de orden 3.

Otra ventaja de la ecuación (13) es que permite ver con más claridad que el mecanismo subyacente de estimación de un Modelo de Factores Dinámicos busca modelar la distribución de los regresores, tal y como afirman Scott y Varian (2013).

## Software econométrico utilizado

Para esta tesis se utilizó el software R, haciendo uso extensivo del paquete “Nowcasting”<sup>1</sup> (Ferreira, de Mattos, & de Valk, 2019) y “Gtrends”, una API para acceder a Google Trends a través de R. Los scripts pueden ser consultados en el link a la plataforma Github que se adjunta al pie de página<sup>2</sup>. El paquete “Nowcasting” consta de una serie de funciones que permiten la creación de un panel balanceado, la estimación de modelos de factores dinámicos tanto haciendo uso de modelos puentes, como mediante el algoritmo de maximización de expectativas. Luego, también cuenta con la función PRTDB, la cual permite crear un pseudo panel a una fecha determinada impuesta por el investigador, de forma de testear fuera de la muestra la capacidad predictiva del sintético de Nowcasting construido, y la evolución mensual del Nowcasting.

Quizá la limitación más importante con la que cuenta esta herramienta sea la imposibilidad de introducir variables de frecuencia mayor a la mensual, lo cual constituye una limitante importante para el testeo de la capacidad predictiva de las variables de Google Trends en las frecuencias diarias y semanales. De todas maneras, las series mensuales de dicha fuente son aceptadas, estando a disposición de la carga de factores con un retraso de cero días, algo que ninguna otra fuente puede ofrecer. Es importante aclarar que la función Nowcast, del paquete Nowcasting, se encarga de la estimación de los diferentes modelos una vez fue establecida la ecuación a estimar, esto es, definiendo  $y_t^m$  como regresada y el set de datos correspondiente a cada modelo como la regresora. De esta forma, el usuario se evita la programación del algoritmo EM, dejando en manos de la función de R la maximización de la función de verosimilitud y la aplicación del filtro de Kalman.

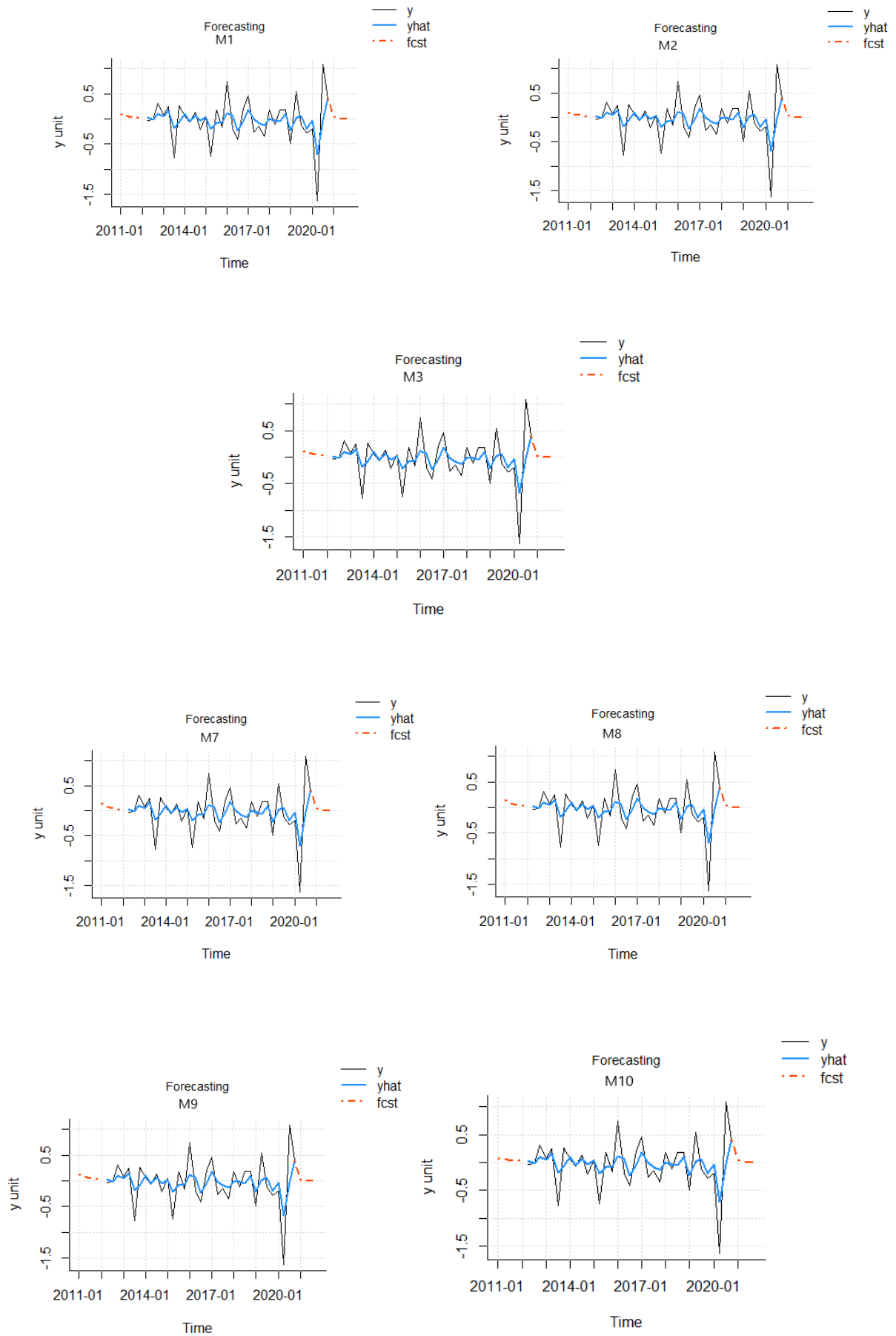
## Resultados Obtenidos

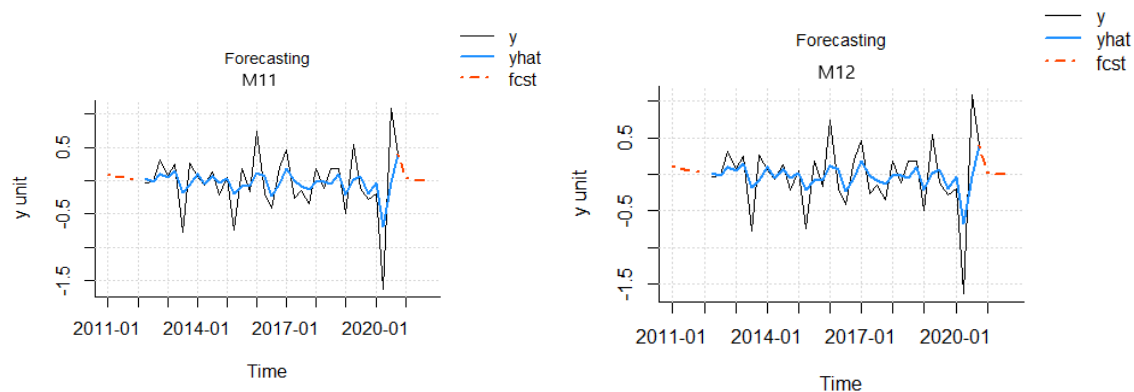
A continuación, se presentan la visualización de los modelos estimados para el caso de los diferentes modelos previamente definidos en la sección “Modelo de análisis”.

---

<sup>1</sup> El código fuente del Paquete se puede encontrar en la siguiente web: <https://rdrr.io/cran/nowcasting/src/R/nowcast.R> (disponible al 18/10/2021). El código fuente de la función BPanel se encuentra disponible en: <https://rdrr.io/github/nmecsyst/nowcasting/src/R/Bpanel.R>. En tanto el código fuente del algoritmo EM se encuentra en el siguiente link: [https://rdrr.io/cran/nowcasting/src/R/method\\_EM.r](https://rdrr.io/cran/nowcasting/src/R/method_EM.r)

<sup>2</sup> [https://github.com/GermanHernandezMagallanes/Tesis\\_master\\_Nowcasting](https://github.com/GermanHernandezMagallanes/Tesis_master_Nowcasting)





Como es posible apreciar, el sintético surgido de la proyección de la variable objetivo en los factores, en azul, describe con mucha fidelidad la trayectoria de la transformación estacionaria del PIB definida en la ecuación (10) para todos los modelos presentados. En rojo punteado es posible visualizar el resultado de la proyección fuera de la muestra. Estos resultados visualizados son sumamente alentadores pero deben ser observados con mayor detenimiento mediante cifras objetivas que se presentan a continuación, en forma de matriz de errores absolutos dentro y fuera de la muestra.

Error absoluto medio dentro y fuera de muestra por modelo			
Modelo	RMSE en muestra	Error Absoluto fuera de muestra	Bloques incluidos por modelo
M7	0.3513	0.0338	Tradicionales y monetarias
M8	0.3518	0.0301	Tradicionales y monetarias
M9	0.3529	0.0119	Tradicionales y monetarias
M10	0.3512	0.0334	Tradicionales y laborales
M11	0.3518	0.0300	Tradicionales y laborales
M12	0.3532	0.0079	Tradicionales y laborales
M1	0.3513	0.0335	Tradicionales y Google
M2	0.3518	0.0300	Tradicionales y Google
M3	0.3532	0.0083	Tradicionales y Google
ARIMA (0,1,1)(0,1,0)	0.7257	0.3275	

En primer lugar, se debe aclarar el motivo por el cual se omitieron varios modelos (M4, M5 y M6). Los mismos, mostraban errores de predicción superiores a otros modelos más parsimoniosos, contrario a lo que podría suponerse. Esta discrepancia requiere un estudio en mayor profundidad, pudiéndose atribuir, como hipótesis, a consecuencias asociadas a las etapas no lineales del método y la forma en la que el mismo realiza la asignación fuera de la muestra para los bloques.

De la tabla de errores se desprenden dos grupos de modelos con mejor ajuste dentro y fuera de la muestra, ya sea estén resaltados en celeste o en verde, respectivamente. Es inmediato notar el primer y más resaltable éxito de este trabajo: cualquier modelo que se elija supera ampliamente en ajuste a un ARIMA (0,1,1) (0,1,0), tanto dentro de la muestra como fuera de la misma, en especial en este último rubro, donde los mejores modelos logran un error absoluto casi nulo en un horizonte de predicción de un trimestre.

El mejor ajuste dentro de la muestra lo logran los modelos 1, 7 y 10, los cuales tienen en común el hecho de tratarse de VAR de orden 1, en tanto los modelos que minimizan el error fuera de la muestra, los modelos 3, 9 y 12 comparten ser VAR de orden 3.

De los resultados dentro de la muestra se desprende que no hay mayor ganancia o pérdida entre las distintas especificaciones de los modelos de factores dinámicos ajustados en este trabajo de incluir o excluir a las variables de Google.

Cuando volvemos al análisis de los resultados fuera de la muestra, los dos mejores modelos dentro de la terna que mejor ajusta, específicamente el modelo 12 y el modelo 3 pueden ser utilizados para ilustrar una ganancia a partir del uso de las variables de Google Trends como insumo para predictores oportunos del PIB, en cuanto no existe gran diferencia entre los errores absolutos alcanzados por cada uno, pero sí destaca el hecho de que el retraso de publicación entre ambos bloques desequilibra la balanza a favor de los índices de búsqueda por sobre los datos del Instituto de Estadística.

A partir del análisis de robustez del tipo Leave-one-out, llevado a cabo con el fin de testear la sensibilidad de los resultados de los modelos que incluyen un bloque dedicado a las variables de Google, no se concluye que alguna de ellas presente un poder explicativo muy superior a las restantes. Las salidas son presentadas en el anexo 4.

## Conclusiones

Los resultados obtenidos resultan esperanzadores en cuanto muestran un desempeño muy similar entre los modelos que incluyen al bloque de Google, respecto a aquellos que únicamente se basan en fuentes más tradicionales. Es posible que de haberse contado con información más completa de Google los resultados hubieran propendido a arrojar mejores niveles de ajuste aún a los efectivamente logrados.

En base a los resultados obtenidos se concluye que las series con origen en Indicadores de búsqueda de Google realizan un aporte significativo en el marco de los modelos de factores dinámicos empleados, cuya magnitud deberá ser estudiada en futuros desarrollos.

### Limitaciones y posibles extensiones

La presente tesis intentó hacer uso de una manera novedosa respecto de las fuentes tradicionales de predicción y adelantamiento del PIB, incorporando series no utilizadas anteriormente, pero atractivas por su publicación instantánea. Lamentablemente el grado de avance del paquete utilizado no permitió ir más allá de un adelantamiento mensual de la variable objetivo. La introducción de datos de frecuencias horaria, diaria y semanal no fue posible por limitaciones de software, siendo dicho refinamiento teóricamente posible y deseable. Además, la elección de los términos a utilizar, aunque sustentada en antecedentes, no constituye una forma exhaustiva de selección, de modo que muchos términos con capacidad de adelantamiento pudieron haber sido dejados de lado.

Una limitación importante de esta tesis discurrió por el número de series empleadas, siendo una cantidad inferior a la utilizada por los trabajos de Giannone et. al, Ferreira, et al y por casi todos los trabajos consultados. Esta estrategia elegida está en línea con los trabajos de primera generación citados por Stock y Watson (2011) pero se aleja de los trabajos más recientes, entre los que se encuentran los de Giannone, más que nada en cuanto al tamaño del panel con el que se trabaja. Esto puede traer aparejados inconvenientes vinculados a una reducción de dimensionalidades donde la cantidad de factores incluidos no sea lo suficientemente ínfima con relación al número total de variables incluidas. Dada la cantidad de variables incluidas, no se puede asegurar que la distribución para muestras grandes de los estimadores de los factores comunes se cumpla, con lo que esta es otra línea de investigación y posible fuente de refinamiento que se abre con este trabajo.

Tres líneas de relevancia principal que deberán ser atendidas en futuros desarrollos radican en la obtención de series de Google sin transformar, esto es, expresadas en número absoluto de búsquedas, en la posibilidad de experimentar con una mayor cantidad de vocablos y en ser capaz de agruparlos en sub-bloques temáticos, de manera que se pueda medir la utilidad de, por citar un ejemplo, un sub-bloque de búsquedas vinculado a vocablos de índole comercial, que podría estar conformado por palabras tales como autos, nombres de tiendas web, términos relacionados con descuentos y ofertas, entre otras y con todas sus posibles variantes pasibles de ser agregadas en términos absolutos. El uso de expresiones regulares podría ampliar el espectro de elección de variables, así como también lo haría el establecimiento de un índice de frecuencia en el uso de determinadas palabras en contextos recesivos o expansivos, ya sea a partir de un ranking dentro del universo total de vocablos buscados por los uruguayos en Google, como en otros medios. El empleo de técnicas de Machine Learning, que permitan juzgar con mayor propiedad la totalidad de la capacidad de la base de datos y no únicamente una porción, que, si bien no es arbitraria, incurre en muchos riesgos de omisión de variable relevante podría ser un buen desarrollo expansivo de este trabajo. Adicionalmente, la obtención de las series de Google sobre la cantidad absoluta de búsquedas de cada término también aportaría una ventaja con la que no contó este trabajo: poder agregar aquellas variables que puedan ser escritas de manera variable en una sola, ya que de esta manera no hubiera sido necesario incluir dos variables para lo que claramente es una sola intención de búsqueda, como fue en el caso de la variable “El país” y su similar con omisión de acento gráfico.



## Referencias Bibliográficas

- Bañbura, M., Giannone, D., & Reichlin, L. (2010). *Nowcasting*. European Central Bank.
- Bañbura, M., Giannone, D., Modugno, M., & Reichlin, L. (2013). Now-Casting and the real-time data flow. En G. Elliot, & A. Timmermann, *Handbook of Economic Forecasting 2, Part A* (págs. 195-237). Amsterdam: Elsevier.
- Bolhuis, M., & Rayner, B. (2020). *Deus ex Machina? A framework for Macro Forecasting with Machine Learning*. International Monetary Fund Working Paper.
- Bontempi, M., Golinelli, R., & Squadrani, M. (2016). *A New Index of Uncertainty Based On Internet Searches: A Friend or Foe of Other Indicators?* Bolonia: Departamento de Economía de la Universidad de Bolonia.
- Brons, K., Lange, R. J., Niesert, R. F., Oorschot, J. A., & Veldhuisen, C. P. (2020). Can Google search data help predict macroeconomic series? *International Journal of forecasting*, 1163-1172.
- Browning, K., & Collier, C. G. (1989). Nowcasting of Precipitation Systems. *Review of Geophysics*, 345-370.
- Brum, C., & Rodríguez, G. (2016). Modelos puente para proyectar el PIB en el corto plazo. Enfoque sectorial. *Documentos de Trabajo, Banco Central del Uruguay*.
- Bruno, G., & Lupi, C. (2004). Forecasting industrial production and the early detection of turning points. *Empirical Economics*, 647-671.
- Burns, A. F., & Mitchell, W. C. (1946). *Measuring Business Cycles*. National Bureau of Economic Research.
- Carlos, P., & Monge, M. (2020). A real time leading economic indicator based on text mining for the Spanish economy. Fractional cointegration VAR and Continuous Wavelet Transform analysis. *International Economics*. Obtenido de [www.elsevier.com/locate/inteco](http://www.elsevier.com/locate/inteco)
- Carrière-Swallow, Y., & Labbé, F. (2010). *Nowcasting with google trends in an emerging market*. Obtenido de Banco Central de Chile:  
[https://www.bcentral.cl/documents/33528/133326/bcch\\_archivo\\_096838\\_es.pdf/fbe6540d-b49a-8256-9616-4e1d477b8178?t=1573280442457](https://www.bcentral.cl/documents/33528/133326/bcch_archivo_096838_es.pdf/fbe6540d-b49a-8256-9616-4e1d477b8178?t=1573280442457)
- Chauvet, M., & Potter, S. (2013). Forecast Output. En G. Elliott, & A. Timmermann, *Handbook of Economic Forecasting 2, Part A*. North Holland.
- Choi, H., & Varian, H. (2011). Predicting the present with Google Trends. 1.
- Cristiano, D., Hernández, M., & Pulido, J. (2014). *Pronóstico de corto plazo en tiempo real para la actividad económica colombiana*. Banco Central de Colombia.
- Davidowitz, S. V. (2014). *A Hands-on Guide to Google Data*.

- Dempster, A., Laird, N., & Rubin, D. (1977). Maximum likelihood estimation from incomplete data. *Journal of the Royal Statistical Society*, 1-38.
- Engle, R., & Watson, M. (1981). A One-Factor Multivariate Time Series Model of Metropolitan Wage Rates. *Journal of the American Statistical Association* .
- Ferreira, P., de Mattos, D., & de Valk, S. (2019). Nowcasting: An R Package for predicting Economic Variables Using Dynamic Factor Models. *The R Journal* .
- Fiorentini, G., Galesi, A., & Sentana, E. (2018). A spectral EM algorithm for dynamic factor models. *Journal of Econometrics*, 249-279.
- Forni, M., Hallin, M., Lippi, M., & Reichlin, L. (2000). The generalized dynamic-factor model: Identification and estimation. *Review of Economics and Statistics*, 540-554.
- Galaso, P., & Rodríguez López, S. (2014). A Composite Leading Cycle Indicator for Uruguay. *Documentos de trabajo, Instituto de Economía*.
- Giannone, D., Reichlin, L., & Small, D. (2008). Nowcasting: The real-time informational content of macroeconomic data. *Journal of Monetary Economics*, 665-676.
- Grover, S., & Kliesen, K. a. (2016). A Macroeconomic News Index for constructing Nowcast of U.S. Real Gross Domestic Product Growth. *Federal Reserve Bank of Saint Louis Review*, 277-96.
- Hamilton, J. D. (1994). *Time Series Analysis*. Princeton, New Jersey: Princeton University Press.
- Heikkinen, J. (2019). Nowcasting GDP Growth using Google Trends. (*Tesis de Maestría*). Escuela de Negocios y Economía, Universidad de Jyväskylä, Jyväskylä, Finlandia.
- Hyunyoung Choi, H. V. (2011). Predicting the present with Google Trends. 1.
- Lanzilotta, B., Mordecki, G., & Umpiérrez, V. (2018). *Political economic uncertainty in a small & open economy: the case of Uruguay*. Montevideo: Instituto de Economía.
- Mariano, R. S., & Murasawa, Y. (2003). A new coincident index of business cycles based on monthly and quarterly series. *Journal of applied econometrics*, 427-443.
- Mills, T. C. (2019). *Applied Time Series Analysis: A Practical Guide to Modelling and Forecasting* . Academic Press .
- Mitchell, W. C., & Burns, A. F. (1938). *Statistical Indicators of Cyclical Revivals*. New York: National Bureau of Economic Research.
- Mostbeck, J. (2019). Master Thesis Nowcasting US GDP Growth in 'Pseudo' Real Time Using Various Econometric Models.
- Prabhu, T. (29 de febrero de 2020). *Google Trends API for Python*. Obtenido de Toward data sciences : <https://towardsdatascience.com/google-trends-api-for-python-a84bc25db88f>
- R Core Team. (2021). R: A Language and Environment for statistical computing. Vienna, Austria. Obtenido de <https://www.R-project.org/>

- Rodriguez, H. (2014). Un indicador de la evolución del PIB uruguayo en tiempo real. *Documento de trabajo del Banco Central del Uruguay 2014/009*. Montevideo: Banco Central del Uruguay.
- Scott, S., & Varian, H. (2013). Predicting the Present with Bayesian Structural Time Series.
- Scott, S., & Varian, H. (2014). Bayesian Variable Selection for Nowcasting Economic Time Series.
- Stephens-Davidowitz, S., & Varian, H. (2014 (Revisado en 2015)). A Hands-on Guide to Google Data.
- Stock, J. H., & Watson, M. W. (1988). *A probability model of the coincident economic indicators*. NBER.
- Stock, J. H., & Watson, M. W. (2020). *Introduction to Econometrics*. Harlow: Pearson Education Limited .
- Stock, J., & Watson, M. (2011). Dynamic Factor Models. En M. P. Clements, & D. F. Hendry, *Handbook on Economic Forecasting*. Oxford University Press.
- Stock, J., & Watson, M. (2016). Chapter 8 - Dynamic Factor Models, Factor-Augmented Vector Autoregressions, and Structural Vector Autoregressions in Macroeconomics. En J. B. Taylor, & H. Uhlig, *Handbook in Macroeconomics, Volumen 2* (págs. 415-525).
- The Economist. (2012). *What is Nowcasting?* Obtenido de Youtube:  
<https://www.youtube.com/watch?v=uKu4Lh4VLR8>
- Thorsrud, L. A. (2016). Nowcasting using news topics. Big data versus big bank. *Norgen Bank research*.
- Yau, R., & Hueng, J. (2017). Nowcasting GDP Growth for small Open Economies with a mixed-frequency Structural Model. *Springer Science + Business Media New York*.

# Anexo 1-Diccionario de variables utilizadas

## Tradicionales

Variable	Inicio	Fin	Frecuen	Transformación Auto. arima	Unidad de medida	Fuente	Observación
Índice de Volumen Físico del PIB Uruguay	ene-11	dic-20	Trimestral		0 Índice 2016 base 100	BCU	Retropolación de Pienika et al para valores entre 2011 y 2016 A través de Gabriela Mordecki, IECON
Exportaciones	ene-11	dic-20	Mensual		5 Millones de dólares corrientes	BCU	
Facturación total de energía eléctrica	ene-11	dic-20	Mensual		5 Mega Watts por hora	MIEM	
Faena	ene-11	dic-20	Mensual		0 Cabezas de ganado	INAC	
Importaciones	ene-11	dic-20	Mensual		2 Millones de dólares corrientes	BCU	A través de Gabriela Mordecki, IECON
Venta de derivados de petróleo al mercado interno y zf	ene-11	dic-20	Mensual		0 Miles de metros cúbicos	ANCAP/MIEM	
Índice de volumen físico de la industria manufacturera	ene-11	dic-20	Mensual		5 Índice 2006 base 100	INE	
Turismo receptivo	ene-11	dic-20	Mensual		4 Personas	Ministerio de turismo	A través de Gabriela Mordecki, IECON
Ventas totales de cemento	ene-11	dic-20	Mensual		0 Toneladas	Cámara de Industrias	
Autos vendidos en el mercado doméstico	ene-11	dic-20	Mensual		0 Unidades	ACAU/ASCOMA	
UBI promedio	ene-11	dic-20	Mensual		0 Puntos básicos	República AFAPs	

## Google

Variable	Inicio	Fin	Frecuencia	Transformación	Unidad de medida	Fuente
ofertas	ene-11	dic-20	Mensual		0 Índice de frecuencia máxima base 100	Google
descuentos	ene-11	dic-20	Mensual		0 Índice de frecuencia máxima base 100	Google
mides	ene-11	dic-20	Mensual		2 Índice de frecuencia máxima base 100	Google
trabajo	ene-11	dic-20	Mensual		0 Índice de frecuencia máxima base 100	Google
vacantes	ene-11	dic-20	Mensual		0 Índice de frecuencia máxima base 100	Google
BPS	ene-11	dic-20	Mensual		0 Índice de frecuencia máxima base 100	Google
banco de prevision social	ene-11	dic-20	Mensual		0 Índice de frecuencia máxima base 100	Google
crisis	ene-11	dic-20	Mensual		0 Índice de frecuencia máxima base 100	Google
depression	ene-11	dic-20	Mensual		0 Índice de frecuencia máxima base 100	Google
Argentina	ene-11	dic-20	Mensual		0 Índice de frecuencia máxima base 100	Google
Brasil	ene-11	dic-20	Mensual		0 Índice de frecuencia máxima base 100	Google
dólar	ene-11	dic-20	Mensual		2 Índice de frecuencia máxima base 100	Google
dolar	ene-11	dic-20	Mensual		2 Índice de frecuencia máxima base 100	Google
inflacion	ene-11	dic-20	Mensual		0 Índice de frecuencia máxima base 100	Google
miami box	ene-11	dic-20	Mensual		0 Índice de frecuencia máxima base 100	Google
mercado libre	ene-11	dic-20	Mensual		0 Índice de frecuencia máxima base 100	Google
automotoras	ene-11	dic-20	Mensual		0 Índice de frecuencia máxima base 100	Google
autos	ene-11	dic-20	Mensual		2 Índice de frecuencia máxima base 100	Google
automovil	ene-11	dic-20	Mensual		0 Índice de frecuencia máxima base 100	Google
auto	ene-11	dic-20	Mensual		0 Índice de frecuencia máxima base 100	Google
LTE	ene-11	dic-20	Mensual		0 Índice de frecuencia máxima base 100	Google
OSE	ene-11	dic-20	Mensual		0 Índice de frecuencia máxima base 100	Google
Antel	ene-11	dic-20	Mensual		0 Índice de frecuencia máxima base 100	Google
Ancap	ene-11	dic-20	Mensual		0 Índice de frecuencia máxima base 100	Google
el correo	ene-11	dic-20	Mensual		0 Índice de frecuencia máxima base 100	Google
correo uruguayo	ene-11	dic-20	Mensual		0 Índice de frecuencia máxima base 100	Google
El país	ene-11	dic-20	Mensual		2 Índice de frecuencia máxima base 100	Google
el pais	ene-11	dic-20	Mensual		2 Índice de frecuencia máxima base 100	Google
el observador	ene-11	dic-20	Mensual		2 Índice de frecuencia máxima base 100	Google
la mañana	ene-11	dic-20	Mensual		2 Índice de frecuencia máxima base 100	Google
la diaria	ene-11	dic-20	Mensual		2 Índice de frecuencia máxima base 100	Google
montevideo portal	ene-11	dic-20	Mensual		2 Índice de frecuencia máxima base 100	Google

## Monetarias

Variable	Inicio	Fin	Frecuencia	Transformación Auto.arima	Unidad de medida	Fuente
1.- EMISION	ene-11	dic-20	Mensual		4 Saldos a fin de mes en Millones de pesos	Banco Central del Uruguay
2.- CIRCULANTE EN PODER DEL SISTEMA BANCARIO	ene-11	dic-20	Mensual		2 Saldos a fin de mes en Millones de pesos	Banco Central del Uruguay
3.- CIRCULANTE FUERA DEL SISTEMA BANCARIO = (1) - (2)	ene-11	dic-20	Mensual		4 Saldos a fin de mes en Millones de pesos	Banco Central del Uruguay
4.- DEPOSITOS VISTA EN MONEDA NACIONAL (b)	ene-11	dic-20	Mensual		2 Saldos a fin de mes en Millones de pesos	Banco Central del Uruguay
5.- M 1 = (3) + (4)	ene-11	dic-20	Mensual		4 Saldos a fin de mes en Millones de pesos	Banco Central del Uruguay
6.- DEPOSITOS EN CAJA DE AHORROS EN MONEDA NACIONAL	ene-11	dic-20	Mensual		4 Saldos a fin de mes en Millones de pesos	Banco Central del Uruguay
7.- M 1' = (5) + (6)	ene-11	dic-20	Mensual		4 Saldos a fin de mes en Millones de pesos	Banco Central del Uruguay
8.- DEPOSITOS PLAZO EN MONEDA NACIONAL	ene-11	dic-20	Mensual		2 Saldos a fin de mes en Millones de pesos	Banco Central del Uruguay
Depósitos a plazo fijo (c)	ene-11	dic-20	Mensual		2 Saldos a fin de mes en Millones de pesos	Banco Central del Uruguay
Certificados de depósito	ene-11	dic-20	Mensual		2 Saldos a fin de mes en Millones de pesos	Banco Central del Uruguay
3.- M 2 = (7) + (8)	ene-11	dic-20	Mensual		4 Saldos a fin de mes en Millones de pesos	Banco Central del Uruguay
10.- DEPOSITOS DE AHORRO PREVIO	ene-11	dic-20	Mensual		4 Saldos a fin de mes en Millones de pesos	Banco Central del Uruguay
11.- M 2' = (9) + (10)	ene-11	dic-20	Mensual		4 Saldos a fin de mes en Millones de pesos	Banco Central del Uruguay
12.- DEPOSITOS EN MONEDA EXTRANJERA DE RESIDENTE (En millones de U\$S)	ene-11	dic-20	Mensual		2 Saldos a fin de mes en Millones de pesos	Banco Central del Uruguay
13.- M 3 = (11) + (12)	ene-11	dic-20	Mensual		2 Saldos a fin de mes en Millones de pesos	Banco Central del Uruguay
Tipo de cambio fin de período	ene-11	dic-20	Mensual		2 Pesos uruguayos por dólar estadounidense	Banco Central del Uruguay

## Laborales

Variable	Inicio	Fin	Frecuencia	Transformación Auto.arima	Unidad de medida	Fuente
Tasa de actividad	ene-11	dic-20	Mensual		2 % de la PEA sobre la PET	Instituto Nacional de Estadística
Empleo	ene-11	dic-20	Mensual		2 % de la PEA ocupada respecto a la PET	Instituto Nacional de Estadística
Tasa de desempleo	ene-11	dic-20	Mensual		2 % de la PEA desocupada	Instituto Nacional de Estadística

## Anexo 2: Transformaciones disponibles a través de la función Bpanel de “Nowcasting”

Transformaciones paquete "Nowcasting" (2019)	
Transformación	Significado
0	Serie sin transformar
1	Tasa de crecimiento mensual
2	diferencia mensual
3	Diferencia Mensual de la tasa de crecimiento interanual
4	Diferencia Mensual de la diferencia 12
5	Diferencia 12
6	Tasa de crecimiento interanual
7	Tasa de crecimiento trimestral (reservada para serie explicada)

## Anexo 3: Scripts

Los scripts son de dominio público, al igual que las variables utilizadas y las tablas con resultados mediante acceso a la plataforma Github siguiendo el link que se copia a continuación:

[https://github.com/GermanHernandezMagallanes/Tesis\\_master\\_Nowcasting](https://github.com/GermanHernandezMagallanes/Tesis_master_Nowcasting)

## Anexo 4: Salidas del análisis de robustez

	Ninguna Exc.	ofertas	descuentos	mides	trabajo	vacantes	BPS	banco de prevision social	crisis	depression
M1	0.3512617	0.3512484	0.3512576	0.3512691	0.3512576	0.3512617	0.3512785	0.3512376	0.3512617	0.3512656
M2	0.3518215	0.3518167	0.3518201	0.3518245	0.3518201	0.3518215	0.3518252	0.3518121	0.3518201	0.3518237
M3	0.3531621	0.3531815	0.3531657	0.3531427	0.3531621	0.3531621	0.3531348	0.3531723	0.3531657	0.3531621
M13	0.3513215	0.3513221	0.3513221	0.3513218	0.3513221	0.3513212	0.3513212	0.3513198	0.3513221	0.3513218
M14	0.3518163	0.3518160	0.3518160	0.3518160	0.3518160	0.3518165	0.3518166	0.3518183	0.3518159	0.3518163
M15	0.3530293	0.3530251	0.3530251	0.3530272	0.3530251	0.3530315	0.3530315	0.3530430	0.3530230	0.3530272

Argentina	Brasil	dólar	dolar	inflacion	miami box	mercado libre	automotoras	autos	automovil
0.3512656	0.3512656	0.3512656	0.3512656	0.3512656	0.3512656	0.3512617	0.3512376	0.3512484	0.3512177
0.3518227	0.3518227	0.3518227	0.3518227	0.3518227	0.3518237	0.3518215	0.3518093	0.3518167	0.3517983
0.3531621	0.3531545	0.3531584	0.3531545	0.3531545	0.3531506	0.3531584	0.3531824	0.3531723	0.3531716
0.3513221	0.3513221	0.3513221	0.3513221	0.3513221	0.3513212	0.3513209	0.3513221	0.3513212	0.3513198
0.3518160	0.3518160	0.3518160	0.3518160	0.3518160	0.3518165	0.3518171	0.3518158	0.3518166	0.3518181
0.3530251	0.3530251	0.3530251	0.3530251	0.3530251	0.3530293	0.3530337	0.3530251	0.3530315	0.3530430

	Ninguna Exc.	ofertas	descuentos	mides	trabajo	vacantes	BPS	banco de prevision social	crisis	depression	Argentina
M1	0.033503362	0.033456637	0.033489021	0.033528839	0.033489021	0.033503362	0.033560265	0.033418496	0.033503362	0.033516605	0.033516605
M2	0.029983297	0.029990952	0.029985320	0.029979914	0.029985320	0.029983297	0.029979554	0.029998973	0.029985320	0.029980633	0.029981748
M3	0.008307108	0.007973484	0.008249007	0.008611087	0.008307108	0.008307108	0.008732729	0.008139711	0.008249007	0.008307108	0.008307108
M13	0.033679682	0.033681351	0.033681351	0.033680524	0.033681351	0.033678820	0.033678820	0.033675125	0.033681351	0.033680524	0.033681351
M14	0.030065587	0.030067830	0.030067830	0.030067830	0.030067830	0.030064416	0.030063212	0.030050822	0.030068904	0.030065587	0.030067830
M15	0.010290912	0.010351814	0.010351814	0.010321636	0.010351814	0.010259628	0.010259628	0.010094219	0.010381461	0.010321636	0.010351814

auto	UTE	OSE	Antel	Ancap	el correo	correo uruguayo	El país	el país	el observador	la mañana	la diaria	montevideo portal
0.3512177	0.3512656	0.3512656	0.3512656	0.3512576	0.3512691	0.3512249	0.3512656	0.3512576	0.3512617	0.3512656	0.3512617	0.3512725
0.3517983	0.3518237	0.3518237	0.3518237	0.3518167	0.3518237	0.3518024	0.3518237	0.3518201	0.3518215	0.3518227	0.3518215	0.3518245
0.3531765	0.3531545	0.3531545	0.3531584	0.3531723	0.3531506	0.3531815	0.3531545	0.3531657	0.3531621	0.3531545	0.3531584	0.3531388
0.3513228	0.3513206	0.3513206	0.3513218	0.3513223	0.3513215	0.3513223	0.3513221	0.3513218	0.3513218	0.3513221	0.3513218	0.3513215
0.3518156	0.3518173	0.3518173	0.3518162	0.3518159	0.3518163	0.3518159	0.3518160	0.3518162	0.3518162	0.3518160	0.3518160	0.3518163
0.3530112	0.3530360	0.3530360	0.3530272	0.3530230	0.3530293	0.3530170	0.3530251	0.3530272	0.3530272	0.3530251	0.3530272	0.3530272

Brasil	dólar	dolar	inflacion	miami box	mercado libre	automotoras
0.033516605	0.033516605	0.033516605	0.033516605	0.033516605	0.033503362	0.03341850
0.029981748	0.029981748	0.029981748	0.029981748	0.029980633	0.029983297	0.03000398
0.008427252	0.008366688	0.008427252	0.008427252	0.008488389	0.008366688	0.00794999
0.033681351	0.033681351	0.033681351	0.033681351	0.033678820	0.033677937	0.03368135
0.030067830	0.030067830	0.030067830	0.030067830	0.030064416	0.030059397	0.03006995
0.010351814	0.010351814	0.010351814	0.010351814	0.010290912	0.010227766	0.01035181

autos	automovil	auto	UTE	OSE	Antel	Ancap	el correo	correo uruguayo	El país	el pais	el observador	la mañana	la diaria	montevideo portal
0.033456637	0.03334761	0.033347607	0.033516605	0.033516605	0.033516605	0.033489021	0.033528839	0.033373336	0.033516605	0.033489021	0.033503362	0.033516605	0.033503362	0.033540147
0.029990952	0.03002330	0.030023298	0.029980633	0.029980633	0.029980633	0.029990952	0.029980633	0.030016121	0.029980633	0.029985320	0.029983297	0.029981748	0.029983297	0.029979914
0.008139711	0.00803834	0.007984451	0.008427252	0.008427252	0.008366688	0.008139711	0.008488389	0.007973484	0.008427252	0.008249007	0.008307108	0.008427252	0.008366688	0.008672138
0.033678820	0.03367512	0.033683754	0.033677029	0.033677029	0.033680524	0.033682163	0.033679682	0.033682163	0.033681351	0.033680524	0.033680524	0.033681351	0.033680524	0.033679682
0.030063212	0.03005234	0.030071941	0.030058057	0.030058057	0.030066725	0.030068904	0.030065587	0.030068904	0.030067830	0.030066725	0.030066725	0.030067830	0.030067830	0.030065587
0.010259628	0.01009422	0.010548962	0.010195312	0.010195312	0.010321636	0.010381461	0.010290912	0.010467357	0.010351814	0.010321636	0.010321636	0.010351814	0.010321636	0.010321636