



Vigilada Mineducación

**Nowcasting de la economía colombiana a través de indicadores de
confianza y expectativas de los agentes económicos.**

Nowcasting of the colombian economy through indicators of confidence and
expectations of the economic agents.

SARA GÓMEZ MONTOYA.

Trabajo de grado

Jesús Alonso Botero.

Docente Universidad EAFIT.

UNIVERSIDAD EAFIT
ESCUELA DE ECONOMÍA Y FINANZAS
ECONOMÍA
MEDELLÍN
2021

CONTENIDO

PLANTEAMIENTO DEL PROBLEMA Y JUSTIFICACIÓN	4
RESULTADOS ESPERADOS	8
OBJETIVOS.....	9
GENERAL	9
ESPECÍFICOS	9
MARCO TEÓRICO O MARCO CONCEPTUAL.....	10
REVISIÓN DE LITERATURA.....	12
DISEÑO METODOLÓGICO	16
ESTADÍSTICA DESCRIPTIVA Y PRIMER ACERCAMIENTO AL MODELO	18
ESTIMACIÓN Y RESULTADOS	26
CONCLUSIONES	28
REFERENCIAS	30

RESUMEN

En este proyecto de investigación se pretendió estimar el desempeño de la economía colombiana reflejado a través del ISE (Indicador de Seguimiento de la Economía) a partir de indicadores convencionales y no convencionales de expectativas de los agentes económicos a través del modelo de Nowcasting, el cual utiliza un conjunto de indicadores de alta frecuencia, estandarizadas y estacionarias para pronosticar el ISE mensual en Colombia y el PIB trimestral con el promedio de los 3 valores estimados del ISE. Esto se logró a través del enfoque de aprendizaje automático (Machine Learning) y modelos de factores dinámicos. El principal propósito de este proyecto fue contestar la pregunta de “¿Que tan explicativos y útiles son los indicadores de confianza y expectativas de los agentes económicos para la predicción del crecimiento de la economía?”. Esta pregunta fue respondida a través de correlaciones dinámicas, que mostraron movimientos procíclicos, adelantados, contemporáneos y rezagados de las variables.

Palabras clave: Nowcasting, Estimación, ISE, Expectativas, Colombia, Modelos de Factores dinámicos, Opiniones, Google Trends, Aprendizaje Automático, Correlaciones Dinámicas.

ABSTRACT

This research project was intended to estimate the performance of the Colombian economy reflected through the ISE (Indicator of Monitoring of the Economy) from conventional and unconventional indicators of expectations of economic agents through the Nowcasting model, which uses a set of high-frequency, standardized and stationary indicators to forecast the monthly ISE in Colombia and the quarterly GDP with the average of the 3 estimated ISE values. This was achieved through the machine learning approach and dynamic factor models. The main purpose of this project was to answer the question of "How explanatory and useful are the confidence indicators and expectations of economic agents for predicting economic growth?" This question was answered through dynamic correlations, which showed procyclical, advanced, contemporary and lagged movements of the variables.

Keywords: Nowcasting, Forecasting, ISE, Expectations, Colombia, Dynamic Factor Models, Opinions, Google Trends, Machine Learning, Dynamic Correlations.

PLANTEAMIENTO DEL PROBLEMA Y JUSTIFICACIÓN

Una de las aspiraciones del hombre desde el comienzo de los tiempos siempre ha sido predecir el futuro. Desde memorables personajes de las culturas ancestrales como chamanes, brujos y videntes; hasta rituales de conexión con los antepasados e interpretación de acontecimientos de la naturaleza e incluso de los mismos sueños. Todos estos ejemplos muestran cómo los seres humanos han idealizado el poder de tener el futuro en sus manos y anticiparse a los acontecimientos venideros. Ahora, con todas las evoluciones tecnológicas y los avances científicos, que nos han permitido incluso explorar vida en otros planetas, clonar organismos vivos, replicar elementos naturales por medio de la biotecnología; ideas utópicas acerca de predecir el futuro no parecen tan descabelladas, e inclusive, no parecen inalcanzables. Viéndolo de esa manera entonces, ¿qué tal si, en efecto, los seres humanos llegáramos a ser, de alguna forma, capaces de predecir el futuro? Incluso ejemplos de la vida real como la predicción del clima o de los fenómenos naturales y las estimaciones de rendimientos en el mercado accionario, pintan un panorama esperanzador en cuanto a la consecución del objetivo de la predicción. Lo más positivo de esta idea, es que algunos componentes de ambigüedad sean cada vez menores gracias a la abundante disponibilidad de información que existe actualmente y a los avances de estadísticas y mediciones cualitativas de lo que antes no se creía capaz de ser medido. Subsisten, sin duda, elementos de “incertidumbre radical”, pero la abundancia de información permite la interpretación de elementos antes impredecibles, y la identificación de nuevos choques que afecten el funcionamiento de los sistemas. Así que ahora el hombre no solo se preocupa en saber qué pasará, sino en cuál es la probabilidad de que eso en realidad suceda y cuáles son los demás posibles escenarios que podrían presentarse. Por así decirlo, en convertirse en una especie de “Nostradamus” pero que prediga cada acontecimiento con respaldo matemático, científico y probabilístico.

En la economía, esa especie de adivino, ha comenzado a ser reemplazado por los modelos de nowcasting o predicción inmediata. Estos adjudican su nombre al hecho de ser una herramienta de diagnóstico, en este caso de la economía, en “tiempo real”; combinando entre el presente (now) y el futuro (forecasting) (Banbura, et al. 2013) o simplemente como la “predicción inmediata”. Estos modelos han adquirido gran relevancia durante los últimos años por las grandes ventajas que tiene al reducir el margen de error en la predicción y estimación de modelos, su capacidad de ajuste por medio de iteraciones y la solución de rezagos de información.

Para entenderlo de forma más clara, el Nowcast o “predicción inmediata”, que pronostica el movimiento de los indicadores macroeconómicos inmediatos antes de su publicación, se ha estimado a través de varios enfoques. En ellos, los datos de intensidad de búsqueda obtenidos de Google Trends se han utilizado para mejorar

la precisión del nowcast de diversos indicadores económicos, cómo el desempleo en los Estados Unidos (Nagao et al., 2019); el sector automotriz en economías emergentes (Carrière-Swallow et al., 2013); el PIB en Alemania y Finlandia (Heikkinen, 2019); el PIB de los Estados Unidos (Koop et al., 2019) y el PIB en un conjunto de países (Woloszko, 2020), entre otros más que han investigado si los datos de los motores de búsqueda de internet como Google pueden ayudar a mejorar el Nowcast o previsiones a corto plazo de las variables macroeconómicas.

Dado el contexto de alta incertidumbre que se vive en el mundo desde la pandemia a comienzos de 2020, las autoridades económicas y los gobiernos se vieron obligados a ser rápidos y eficientes a la hora de tomar decisiones fundamentadas con datos reales e implementar políticas públicas que ayuden a contrarrestar las crisis y los choques súbitos de la economía. Sin embargo, el retraso en la publicación de variables macroeconómicas claves, como lo son el crecimiento del PIB, el desempleo, balanza de pagos, entre otros; presenta una limitación para los agentes responsables de la toma de decisiones al restringir su capacidad para evaluar con precisión las condiciones actuales. Estos lapsos de tiempo hacen que el Nowcast (predicción inmediata) sea de vital importancia para evidenciar las condiciones económicas en tiempo real. En países emergentes el problema se agudiza pues estos indicadores suelen ser publicados con frecuencia trimestral o mensual, con grandes retrasos; impidiendo que se tomen acciones pertinentes y rápidas.

Otra de las enseñanzas de la pandemia es la importancia de las personas y sus relaciones e interacciones. Acostumbrados a una dinámica de contacto físico, relacionamiento y convivencia, el COVID 19 le enseñó al mundo que no estaba preparado para afrontar, más allá de una crisis de salud mundial, el confinamiento masivo de personas y el freno a su relacionamiento. Por un lado, develó la vulnerabilidad del sistema social y económico ante la ausencia del contacto humano; pero por otro, puso el bienestar y la confianza de las personas como una de las cosas más importantes para la agenda política y económica. Si bien antes se le daba prioridad al consumidor y se estaba recorriendo a una importante velocidad el campo de las expectativas humanas y sus preferencias, ahora estos agentes son parte primordial del ciclo de la producción. Ahora es el mundo el que se adapta a las personas, para brindarles experiencias, bienestar y cubrir sus necesidades, que cada vez son más específicas. Y, como se mencionaba anteriormente, el argumento de que los deseos humanos y sus opiniones eran completamente subjetivas; cada vez es menor relevante.

Poniéndolo en términos económicos, hemos pasado de los famosos “animal spirits” de Keynes a indicadores de confianza y expectativas de los agentes, y a poder medir las preferencias en áreas como la economía del comportamiento y el economic marketing. Hasta las redes sociales más famosas de hoy en día como Instragram, Twitter o Tiktok son capaces de cuantificar a una persona y su personalidad en términos de likes, artículos o videos compartidos o guardados, búsquedas y tiempo que pasan en cada publicación. Google inclusive, con sus

indicadores de movilidad y Google Trends, cuantifica la actividad y las preferencias de las personas a través de sus búsquedas e intereses.

En este proyecto de investigación se pretendió estimar el desempeño de la economía colombiana reflejado a través del ISE (Indicador de Seguimiento de la Economía) a partir de indicadores macroeconómicos convencionales como el IPI y la inflación, e indicadores de expectativas y opinión de los diferentes agentes económicos como los Índices de Confianza publicados por Fedesarrollo, la Encuesta Mensual Manufacturera con enfoque Territorial publicada por el DANE en sus indicadores de expectativas de producción, ventas y empleo; y el Banco de la República en su Encuesta Mensual de Expectativas Económicas. Por otro lado, se extraen variables de compras de tarjeta de crédito y débito, tomados de la superintendencia financiera y el número de autos vendidos publicados en el informe mensual de Andemos. Entre los indicadores de más alta frecuencia se encuentra la Demanda de Energía SIN (XM) y la Tasa Representativa de mercado (TRM). Por último, datos no convencionales, como Google Trends y el Índice de Movilidad de Google y el Apple Transportation Index. La estimación se realizó a través del modelo de Nowcasting, el cual utiliza un conjunto de indicadores y variables económicas y financieras de alta frecuencia, estandarizadas y estacionarias para pronosticar el ISE mensual en Colombia y el PIB trimestral con el promedio de los 3 valores estimados del ISE. Esto se logra a través del enfoque de aprendizaje automático (Machine Learning) y modelos de factores dinámicos, de forma rápida, precisa y oportuna; alimentándose de la información publicada por las entidades competentes.

La herramienta del Nowcasting permite estimar de manera precisa y eficiente el desempeño de la economía dado que es capaz de captar el comportamiento en tiempo real de variables que tienen más frecuencia de información. El uso de datos no convencionales se considera innovador en Colombia, ya que estos datos son contemporáneos e incluso adelantados. Además de mostrar las expectativas de los individuos que pueden anticiparse a futuros escenarios y determinar la dinámica de la economía. Por tanto, herramientas como Google Trends se configuran como herramientas muy útiles para futuras predicciones.

La economía colombiana sigue mostrando señales de rápida recuperación económica, según la última publicación del DANE, el indicador de seguimiento de la economía (ISE) experimentó una variación anual en agosto respecto al mismo mes del año inmediatamente anterior de 13,2 %, e incluso siendo positivo respecto al mismo mes del año 2019 (prepandemia) con un 2,1 %, continuando con el dinamismo registrado en junio y julio, lo que ratifica las proyecciones alcistas por diferentes analistas respecto al crecimiento económico del 2021, que muy probablemente pasará la barrera del 9%, por ejemplo, Fedesarrollo subió su perspectiva a 9,5%; Bancolombia a 10% y Banco de la República 8,4%.

En el país se mencionan diferentes acercamientos o trabajos realizados sobre esta temática para diferentes propósitos. Entre los más destacados se encuentran

Cristiano et al. (2012), donde los autores utilizan un modelo de espacio de estado para el pronóstico inmediato y aplican un algoritmo basado en varianza del PIB para seleccionar variables relevantes, las cuales explican su comportamiento. A su vez, Pérez-Castañeda (2009) y Cárdenas et al. (2020) aplican modelos de muestreo de datos mixtos (MIDAS) para pronosticar el PIB y cambios en el IPC de Colombia. León y Ortega (2018) utilizaron un modelo de red neuronal exógena, autorregresiva no lineal para pronosticar el ISE considerando transacciones electrónicas, las cuales son factores no convencionales y de alta frecuencia.

El principal propósito de este proyecto fue contestar la pregunta de “¿Que tan explicativos y útiles son los indicadores de confianza y expectativas de los agentes económicos para la predicción del crecimiento de la economía?”

RESULTADOS ESPERADOS

El principal resultado esperado era que el dato proyectado por el modelo para los últimos tres meses del año, fuera lo más similar al dato real una vez sea publicada la información por las entidades responsables. Además de eso, que los resultados obtenidos por la estimación fueran comparables con las estimaciones publicadas por entidades como el Banco de la República, Davivienda y Bancolombia quienes también cuentan con modelos de nowcasting para el seguimiento de la economía en tiempo real. Esto indicaría que un modelo construido a partir de indicadores de expectativas de los agentes económicos es lo suficientemente preciso y eficiente como los modelos de nowcasting con datos tradicionales y con los datos reales de crecimiento.

OBJETIVOS

GENERAL

El principal objetivo de este proyecto fue proveer un pronóstico en tiempo real del crecimiento de la economía colombiana por medio de la implementación de la herramienta del Nowcasting y el Machine Learning, con indicadores convencionales y no convencionales que representaran la confianza y las expectativas de los agentes económicos.

ESPECÍFICOS

1. Evaluar qué tan significativos y relevantes son las expectativas y opiniones de los agentes económicos, específicamente los consumidores y las empresas, para la estimación del PIB de Colombia en el 2021.
2. Analizar las correlaciones que tienen indicadores convencionales como la inflación, los indicadores de confianza y el IPI, datos de alta frecuencia como la demanda de energía e indicadores no convencionales como Google Trends, Google Mobility y Apple transportation con el principal indicador de crecimiento económico, el ISE.
3. Estudiar en tiempo real el desempeño de la economía de forma eficiente y precisa, eliminando los rezagos de información de los diferentes indicadores económicos por el retraso en su publicación por parte de las entidades competentes.

MARCO TEÓRICO O MARCO CONCEPTUAL

Este trabajo se propuso pronosticar en tiempo real el comportamiento de la economía colombiana a través del ISE, indicador de referencia de crecimiento, utilizando datos de alta frecuencia, convencionales y no convencionales; donde las variables utilizadas, son variables que reflejan las expectativas de los principales agentes económicos (los consumidores y las empresas). Dentro de los datos utilizados, se encuentran datos convencionales como la inflación, el ISE, el IPI y la inflación; y además de ellos, indicadores de expectativas como lo son los indicadores de confianza publicados por Fedesarrollo, además de sus expectativas de crecimiento y producción para el próximo trimestre y el próximo semestre. También fueron utilizadas diferentes encuestas como la encuesta a especialistas del Banrep, extrayendo datos como la expectativa de tasa de cambio a 12 meses y a fin de mes y las expectativas de inflación a doce meses y a fin de mes. Además, la encuesta mensual manufacturera con sus resultados de ventas, producción y empleo. Por último, se utilizó la información de transacciones con tarjetas de crédito y débito publicada por la superintendencia financiera y el número de autos vendidos, publicados por Andemos, la asociación nacional de movilidad sostenible.

En cuanto a los datos de alta frecuencia, se utilizaron la demanda de energía extraída del portal XM y el EMBI o indicador de riesgo país, extraído de la plataforma Bloomberg. Respecto a los datos no convencionales, se extrajeron los indicadores de Google Trends, al igual que búsquedas en Youtube proveídas por el mismo buscador, Google Mobility y Apple Transportation.

Para la estimación, fueron implementados modelos de Factores Dinámicos, donde se pretendía crear un índice coincidente que replicara la información de la variable de interés, con lo cual se pretendía construir un factor que capturara el comportamiento del ISE. Tomando los factores estimados como datos, regresándolos (y una constante) cada uno (uno a la vez) en cada una de las variables observadas, y registrando los coeficientes de determinación (valores R^2), podemos obtener una sensación de las variables para las cuales cada factor explica una porción sustancial de la varianza y las variables para las cuales no lo hace. Para esta estimación se debió hacer un proceso de transformación de los datos, que requerían ser estandarizados y estacionarios.

El modelo de factores dinámicos propuesto por Lütkepohl & Helmut (2007), cuenta con 1 factor no observado que se supone que sigue un proceso AR(2). Se supone que las innovaciones ε_t son independientes (de modo que Σ es una matriz diagonal) y el término de error asociado a cada ecuación, μ_t se supone que sigue un proceso AR(2) independiente.

Donde hay k_{endog} series observadas y k_{factors} factores inobservados (2 pues siguen un factor de orden 2). Así, y_t es una k_{endog} x vector 1's y f_t es un vector de k_{factors} x un vector de 1's.

x_t son vectores endógenos según las variables endógenas del modelo, de tamaño k_{exog} x 1. η_t y ε_t son errores de tipo ruido blanco. Para identificar los factores, debe expresarse $\text{Var}(\eta_t) = \text{Identidad}$ y la $\text{Var}(\varepsilon_t) \equiv \Sigma$.

Luego se realiza el proceso de estimación por medio del Machine Learning, donde se estima el modelo por medio de iteraciones del modelo de pronóstico y se construyó en una primera instancia una matriz de entrenamiento que intenta pronosticar el resultado correcto de la estimación y minimizar el error de validación. Posterior a esto, el código escogió cuál fue el modelo más eficiente, es decir, con menor error de validación; y sobre este comienza el modelo de estimación para los tres periodos siguientes para el ISE.

Por último, se corre el nowcast a través de un proceso de aprendizaje automático y se ingresan las variables con el valor de su última observación y el promedio del valor arrojado por el ISE en la etapa anterior.

REVISIÓN DE LITERATURA

Los tipos de modelos que se utilizan en esta metodología son, en primer lugar, los modelos de Factores Dinámicos (MFD) que se basa en el principio de que pequeños números de factores explican grandes fracciones de la varianza de muchas series macroeconómicas. Este enfoque expresa que existen gran cantidad de series de tiempo que tienen factores internos que describen el comportamiento de esas series, como lo señalan los trabajos de Giannone, Reichlin & Small (2008) Sargent y Sims (1977) y Stock & Watson (2004). El crecimiento significativo de los potenciales computacionales, permitió la reconsideración de muchos de estos. De allí surgieron otras técnicas de pronóstico basadas en algoritmos de aprendizaje automático y minería de texto; las cuales mostraron que el uso de esos grandes conjuntos de datos aumentó la precisión del pronóstico con cada inclusión de nuevos datos. Los principales ejemplos de estos modelos estudiados incluyen estructuras de factores débiles, test para el número de factores y modelos de estabilidad y rupturas. La idea general es que los observados (factores económicos) dependen de un estado (factor latente) no observado, con errores ruido blanco.

En segundo lugar, los Modelos de Ecuaciones Puente (EP), implementados por la Organización de las Naciones Unidas para la Agricultura y la Alimentación (FAO). Estos modelos lo que hacen es buscar las elasticidades que hay entre las diferentes variables del modelo para periodos de tiempo. Por medio de estas participaciones trimestrales, se estima un puente o una relación entre sí para estimarlas de manera mensual. No se trata de calcular factores sino indicadores mensuales que se usan como variables explicativas. En tercer lugar, se encuentran los modelos de Análisis de Componentes Principales (PCA) desarrollados por Liu, Matheson & Romeu (2011) para economías emergentes. Estos modelos se proponen reducir la dimensión de las series, no solo en un factor sino agrupar en una sola serie, para estimar un indicador más adelantado.

En la literatura más moderna se encontraron investigaciones que incursionan en el campo de la inteligencia artificial fundamentalmente el Machine Learning como los trabajos de Bolhuis & Rayner (2020). Esta rama de la inteligencia artificial permite identificar patrones entre los datos para hacer predicciones y por medio de las iteraciones, corregir errores y aprender sobre los resultados anteriormente obtenidos para encontrar valores más óptimos y eficientes, pues es capaz de convertir una muestra de datos en un programa informático capaz de extraer inferencias de nuevos conjuntos de datos para los que no ha sido entrenado previamente. Es importante destacar que, en general, en la literatura se combinan los distintos enfoques haciendo un análisis comparativo. La diferencia con técnicas anteriores está en su capacidad para adaptarse a los cambios en los datos a medida que van entrando en el sistema. Los artículos pioneros como Choi y Varian (2009, 2011) han llevado a una explosión del trabajo de predicción inmediata utilizando datos de Google, incluidos, entre muchos otros, Akitas & Zimmermann (2009), Carriere-Swallow y Labbe (2011). Entre estos modelos, existen diversas

publicaciones que involucran las opiniones de los agentes económicos, especialmente de los consumidores y las empresas, como factor fundamental en la predicción de los diferentes indicadores macroeconómicos. Comenzando por la teoría keynesiana en la cual se hacía alusión a los “animal spirits” que representaban las expectativas de los consumidores y determinaba movimientos en indicadores como el consumo, hasta propuestas más evolucionadas como las llamadas ETS (Economic Tendency Surveys) y herramientas como Google Trends y The Google Big Query, que permiten cuantificar de mejor manera más certera las opiniones de los agentes.

En el caso del Banco de la República, la entidad se unió al uso de la metodología del Nowcast para el pronóstico del desempeño de la economía colombiana en el corto plazo. En este artículo se aplican dos tipos de metodologías de pronóstico de corto plazo, las cuales son Factor-MIDAS; la cual consiste en una aplicación de datos mixtos (con diferentes frecuencias). Y modelos DFM, que son modelos de factores dinámicos para la obtención de pronósticos a través de unificación de correlaciones dinámicas de variables explicativas que formen factores de estimación eficiente. El trabajo propone un proceso heurístico para la selección de variables relevantes y se evalúa el ajuste de los modelos comparándolo respecto a metodologías usuales de proyección. En general, los pronósticos de los modelos Factor MIDAS y del DFM superan los generados por metodologías tradicionales, con resultados más precisos en la medida que aumenta el flujo de información (Galeano-Ramírez, 2021). Dentro de los indicadores no convencionales para este modelo, los autores utilizan datos de transacciones electrónicas y otros indicadores que no habían sido incluidos en publicaciones anteriores.

Dentro de los datos utilizados, esta investigación abarcaba la producción industrial, las ventas minoristas, las transacciones electrónicas, los índices de precios, las tasas de interés, los términos de intercambio, las exportaciones, las importaciones, indicadores de cartera, series del mercado laboral, agregados monetarios y expectativas sobre actividad y precios. Las variables involucran información cualitativa obtenida de encuestas (datos suaves) e información cuantitativa de mediciones directas de la actividad económica (datos duros). Todos de ellos se ajustan estacionalmente y se transforman para asegurar la estacionariedad. (Galeano-Ramírez, 2021). Y con estos se siguen los pasos de 1) encontrar las correlaciones dinámicas y la regresión de Lasso 2) La minimización del error cuadrático medio 3) replicación de los patrones de los disponibles de información en la primera quincena de cada mes en un proceso iterativo fuera de muestra.

Claveria et al (2020) utilizan ETS, las cuales preguntan si la variable de estudio se espera que crezca, permanezca estable o decrezca, para la creación de indicadores precisos de confianza y expectativas de los consumidores. En este caso, para la construcción de los indicadores, los resultados se muestran en balances de series que son obtenidas por el porcentaje de respuestas positivas menos el porcentaje de respuestas negativas. En su paper diseñan un experimento de pronóstico recursivo fuera de la muestra en el que calculan iterativamente los indicadores para predecir

el crecimiento económico. Los pronósticos obtenidos se comparan luego con modelos de series de tiempo univariantes y modelos convencionales de series de tiempo. La metodología propuesta se basa en la programación genética (GP), la cual es una técnica de búsqueda de computación basada en la aplicación de algoritmos evolutivos; permitiendo que en el modelo evolucione simultáneamente la estructura y los parámetros de expresiones, permitiendo la formalización de las interacciones entre las variables que mejor se ajustan a una serie de referencia. (Claveria et al, 2020).

Tilly et al (2021) avanza su investigación incorporando una amplia gama de emociones desde periódicos de todo el mundo hasta pronósticos macroeconómicos utilizando la Global Database of Events, Language and Tone (GDELT), la cual es una investigación que analiza artículos de noticias globales y extrae elementos como temas, emociones y ubicaciones. Estos emplean una metodología basada en aprendizaje automático para identificar artículos que son relevantes para los índices macroeconómicos en cuestión, y logran demostrar que las emociones expresadas en esas noticias agregan valor a las previsiones de producción industrial y precios al consumidor. También analizan la correlación de estas variables con el GDP para agrupar las más de 600 puntuaciones de emociones disponibles en un número menor de factores interpretables y así reducir la dimensionalidad.

En la investigación de J Vuchelen (2004), mencionan que el sentimiento del consumidor es el promedio aritmético de balances de opinión sobre tres aspectos. Dos se relacionan con el presente y situación financiera esperada de los hogares; dos preguntas se relacionan con el presente y situación económica general esperada y una última pregunta es sobre compras importantes. Lo que proponen es cuantificar las condiciones económicas esperadas y la incertidumbre. El punto de partida es una especificación de regresión que sintetiza los datos publicados e intenta simular el sentimiento del consumidor. Estos prueban dos grupos de variables, donde el primer grupo incluye variables como los indicadores del ciclo económico, el precio del petróleo, la tasa de inflación, tasas de interés y desempleo que se supone que contienen información sobre el estado general actual y esperado de la economía. El segundo grupo de variables explica condición económica actual y esperada de los hogares. La condición de los hogares se reflejaría en los precios de las acciones, las tasas de interés y las compras que se hacen por tarjetas de crédito. Estos asumen que la mayor parte del tiempo la variable dependiente que es el sentimiento del consumidor es una variable rezagada y además de eso autorregresiva.

En el caso de este trabajo, se usaron como referencia este segundo tipo de variables, las cuales reflejan la condición de los hogares y sus expectativas a través de precios, compras y demanda. Estas condiciones de los agentes son determinadas a través de su ingreso; lo que significa que principalmente estas expectativas de los agentes económicos se forman por las expectativas de ingreso

que tendrán en el futuro. Lo que explica que los agentes no pretenden formar expectativas meramente de crecimiento sino más de precios, ingreso, disponibilidad de consumo y movilidad; determinantes convencionales y directos del crecimiento económico.

DISEÑO METODOLÓGICO

El modelo del Nowcasting se basó en la metodología de factores dinámicos (MFD) en su primera fase y con Machine Learning en la segunda etapa, con el objetivo de monitorear las variables económicas en tiempo real e intentar predecirlas en un futuro cercano.

En primer lugar, se recolectaron datos convencionales, como los principales indicadores líderes del PIB como el indicador de seguimiento de la economía (ISE), los indicadores de confianza (ICC, ICI, ICCO), la inflación mensual y datos de la encuesta mensual manufacturera como producción, ventas y empleo; entre otros indicadores de expectativas publicados por diferentes entidades. Luego se recolectaron otros indicadores de más alta frecuencia (demanda energía, índice de movilidad de Google, índice de transporte de Apple), y datos no convencionales, como Google Trends Colombia; los cuales se encontraron con frecuencia semanal, mensual y diaria.

Luego se encontraron variaciones interanuales para que fueran comparables entre sí. El propósito de usar modelos de factores dinámicos fue ponderar el comportamiento de variables de estudio para lograr que se ajustasen al comportamiento de la variable de interés; la cual, en este caso, es el ISE de Colombia. Estas variables y el error del modelo siguen un comportamiento autorregresivo; lo que quiere decir que su comportamiento actual depende de su comportamiento en el pasado, como lo sugiere la literatura. Entonces, para este tipo de modelos se necesita que los datos sean estacionarios (es decir, que su media y su varianza no dependan del tiempo) y estandarizados (de manera que todas las variables se comporten con distribución normal con media 0 y varianza 1).

Como siguiente paso se estudiaron las relaciones que se tienen con el ISE, se estimaron las correlaciones dinámicas de las variables como indicadores, expresándolas en términos logarítmicos para encontrar sus elasticidades, y sobre ellas aplicándoles un filtro de Hodrick y Prescott para eliminar su componente cíclico. Luego, se encontraron sus valores en rezagos hasta 5 periodos atrás y 5 periodos hacia adelante, y sobre estos se estiman las correlaciones dinámicas que tiene cada variable en cada periodo con el ISE.

Una vez cargados los datos en el modelo, este fue escogiendo las variables según su error de pronóstico, eliminando así variables a lo largo del modelo para eliminar los componentes de error de la estimación. Usando diferentes paquetes del lenguaje Python, se procesaron los datos como se mencionó anteriormente, se les hizo un proceso de estacionarización y suavización de las variables y posteriormente se programaron los modelos de factores (MFD), para continuar con el Machine Learning (ML) y, estimar el Nowcast. Esta primera fase, arrojó el valor del ISE estimado por Google Trends; el cual es el principal ponderador para el resultado final del Nowcast. La segunda fase del modelo con ML consistió en un proceso iterativo, en el que el programa fue aprendiendo y minimizando el error de la

estimación con respecto al dato real, encontrando diferentes ajustes necesarios, hasta que no encuentra una mejor manera de realizar lo que se le está pidiendo; con el fin de tener predicciones posteriores más ajustadas. En esta fase, se introdujo el valor mensual del ISE arrojado en la primera fase de la estimación y se introdujeron los demás indicadores convencionales para estimar el modelo de Nowcast a través de Machine Learning.

Como resultado de los procesos anteriores, se construyó un índice de la actividad económica de tendencias de Google (Nowcast_GT) utilizando consultas de búsqueda para Colombia incluyéndose como regresores en una familia de modelos simples de dinámica de factores, series de tiempo ARIMA, y modelos de aprendizaje supervisado de predicción inmediata (Nowcast). Fueron obtenidos 2 valores para los periodos continuos al ISE (octubre, noviembre), y el promedio de estos 2 resultados conjunto con el ISE arrojado para el mes de septiembre calculado a partir de Google Trends, fue utilizado como el pronóstico de crecimiento para el cuarto trimestre del 2021.

ESTADÍSTICA DESCRIPTIVA Y PRIMER ACERCAMIENTO AL MODELO

Para este ejercicio econométrico y estadístico simple, se utilizó una muestra de un total de 30 variables en términos reales con frecuencia mensual con datos comenzando desde enero del 2010 hasta el mes de agosto del 2021.

La ecuación que describe el modelo en términos reales fue:

$$\begin{aligned}
 ISE_t = & c + \beta_1 HotelGoogle_t + \beta_2 RestauranteGoogle_t + \beta_3 CarroGoogle_t \\
 & + \beta_4 AutomovilGoogle_t + \beta_5 ExpectativasdeInflacionMensual_t \\
 & + \beta_6 TRMFinDeMes_t + \beta_7 ICC_t + \beta_8 ICI_t + \beta_9 ICCO_t \\
 & + \beta_{10} SituacionEmpresa_t \\
 & + \beta_{11} ExpectativasSituacionPorxSemestre_t \\
 & + \beta_{12} ExpectativasSituacionPorxSemestre_t + \beta_{13} VentasReal_t \\
 & + \beta_{14} XM_t + \beta_{15} EMBI_t + \beta_{16} ComprasTarjetaDebito_t + \mu
 \end{aligned}$$

Tabla1. Regresión de las variables en términos reales.

Source	SS	df	MS	Number of obs	=	150
Model	20470.733	16	1279.42081	F(16, 133)	=	259.97
Residual	654.543731	133	4.92138143	Prob > F	=	0.0000
				R-squared	=	0.9690
				Adj R-squared	=	0.9653
Total	21125.2767	149	141.780381	Root MSE	=	2.2184

ISE	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
HotelGoogle	-.065668	.0273837	-2.40	0.018	-.119832	-.0115041
RestauranteGoogle	.6205334	.0859984	7.22	0.000	.4504319	.7906349
CarroGoogle	.0552605	.0348394	1.59	0.115	-.0136506	.1241716
AutomovilYoutube	.9249642	.3030949	3.05	0.003	.3254542	1.524474
EXPECTATIVASDEINFLACIONMENSUA	-913.4949	94.89861	-9.63	0.000	-1101.201	-725.7891
TRMAFINDEMES	-.0044163	.001606	-2.75	0.007	-.0075929	-.0012396
ICC	-.0945678	.0306868	-3.08	0.003	-.1552651	-.0338705
ICI	.2887129	.0864191	3.34	0.001	.1177793	.4596464
ICCO	-.4770029	.1212443	-3.93	0.000	-.7168194	-.2371864
SITUACIONACTUALDELAEMPRESA	.3134809	.060299	5.20	0.000	.1942118	.4327501
EXPECTATIVASSITUACIONPROXSEME	.1543095	.0586613	2.63	0.010	.0382797	.2703394
EXPECTATIVASPRODTRIMESTRE	-.2630111	.027728	-9.49	0.000	-.317856	-.2081663
VENTASREALES	-.1130606	.0635972	-1.78	0.078	-.2388534	.0127322
XM	3.23e-07	4.71e-08	6.86	0.000	2.30e-07	4.16e-07
EMBI	.0096705	.0052728	1.83	0.069	-.0007589	.0200999
COMPRSTD	1.70e-07	6.16e-08	2.76	0.007	4.83e-08	2.92e-07
_cons	42.70827	7.796628	5.48	0.000	27.28684	58.1297

1

¹ Tabla elaborada por el autor.

De lo cual se pudo concluir que Automóvil es la única palabra significativa para el ISE dentro de las búsquedas en la plataforma Youtube. Hotel y Restaurante fueron las palabras de Google Trends que resultaron significativas para el modelo; sin embargo, el coeficiente de la palabra hotel tiene un sentido contradictorio, puesto que muestra una relación inversa con el ISE. En el caso de las expectativas de inflación mensual y TRM a fin de mes, también tuvieron una relación negativa con el ISE, lo cual tiene sentido del lado de la Tasa Representativa del Mercado pues que decrezca, significaría que la moneda local ha tomado fuerza y por ende puede representar crecimiento económico. Las expectativas de inflación, desde la mirada de la confianza de los consumidores, si estos esperan que la inflación suba, contraerán su demanda pues esperan que los precios sean más caros y por ende una contracción en su poder adquisitivo, y este hecho contraería la demanda. Para la situación actual de la empresa y las expectativas de la situación del próximo semestre, estas resultaron significativas bajo todos los niveles de confianza y con betas positivos y mayores al 0.15.

En el caso de los índices de confianza publicados por Fedesarrollo, solo el ICI tuvo un β positivo entre los 3 indicadores que resultaron significativos para el modelo. Las relaciones negativas del ICC y el ICCO no fueron intuitivas, e incluso resultan contradictorias, pues se había planteado que tendrían una relación directa con la variable de interés del modelo. Esto sucede igual en el caso de las expectativas de producción del próximo semestre.

Por último, los datos de más alta frecuencia como el EMBI y el índice de XM que representa la demanda de energía total, solo el XM resultó significativo para el modelo en términos reales bajo todos los niveles de confianza. El EMBI solo resultó significativo bajo el 90%. Y las compras con tarjetas de débito resultaron con una sensibilidad positiva y significativa bajo todos los niveles de confianza.

Tabla 2. Regresión del ISE respecto a los indicadores de Google Trends.

. reg ISE HotelGoogle RestauranteGoogle EmpleoGoogle HotelYoutube MotocicletaYoutube PeliculasYoutube						
Source	SS	df	MS	Number of obs	=	151
Model	19660.6665	6	3276.77775	F(6, 144)	=	269.08
Residual	1753.57393	144	12.1775967	Prob > F	=	0.0000
				R-squared	=	0.9181
				Adj R-squared	=	0.9147
Total	21414.2404	150	142.761603	Root MSE	=	3.4896

ISE	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
HotelGoogle	.1972092	.0498651	3.95	0.000	.098647	.2957714
RestauranteGoogle	.712578	.0980713	7.27	0.000	.5187328	.9064233
EmpleoGoogle	-.49503	.0556384	-8.90	0.000	-.6050035	-.3850565
HotelYoutube	-.6914218	.1244104	-5.56	0.000	-.9373284	-.4455153
MotocicletaYoutube	.937003	.2685315	3.49	0.001	.4062304	1.467776
PeliculasYoutube	.0850647	.0222131	3.83	0.000	.0411588	.1289706
_cons	82.45046	1.943325	42.43	0.000	78.60933	86.29159

2

² Tabla elaborada por el autor con cálculos propios.

En cuanto a los indicadores no convencionales de Google Trends, las palabras que resultaron significativas para el ISE son Hotel, Restaurante y empleo. Para el caso de la plataforma de Youtube, las palabras significativas fueron Hotel, Motocicleta y Películas. Empleo. Esta última, a pesar de tener un β negativo, tiene sentido interpretativo, pues entre menos se busque empleo en Google, quiere decir que hay mayor cantidad de personas empleadas en Colombia y no se requiere buscar tantas vacantes. En el lado de Hotel, este β no tuvo sentido económico, pues se esperaría que entre más se busquen hoteles quiere decir que las familias tienen mayor capacidad adquisitiva y planean viajes en el futuro; e incluso es una señal de reactivación económica específicamente del sector del turismo.

El siguiente modelo, el más adecuado para el estudio del nowcast, consistió en expresar todas las variables en sus variaciones porcentuales, expresándolas en términos logarítmicos para así hallar sus elasticidades.

Para este caso, el modelo tuvo la siguiente forma:

$$\begin{aligned} \ln ISE_t = & c + \beta_1 \ln HotelGoogle_t + \beta_2 \ln RestauranteGoogle_t + \beta_3 \ln EmpleoGoogle_t \\ & + \beta_4 \ln PeliculasGoogle_t + \beta_5 \ln ExpectativasdeInflacionMensual_t \\ & + \beta_6 \ln TRMFinDeMes_t + \beta_7 ICI_t \\ & + \beta_8 ExpectativasProduccionProxTrimestre_t + \beta_9 EmpleoReal_t \\ & + \beta_{10} XM_t + \beta_{11} EMBI_t + \beta_{12} ComprasTarjetaCredito_t + \mu \end{aligned}$$

De las variables mencionadas anteriormente, la única que no tuvo sentido económico en su resultado de la regresión es la variable de Expectativas de Producción para el próximo trimestre, la cual tiene una relación negativa con el ISE. Este modelo tiene un R cuadrado mayor que el modelo representado en términos reales, y representa mejor la estimación, ya que lo que nos proponemos es estimar el crecimiento económico, representado en término de variaciones porcentuales.

Tabla 3. Regresión de las variables en términos logarítmicos.

Source	SS	df	MS	Number of obs	=	141
Model	2.22399921	12	.185333268	F(12, 128)	=	379.96
Residual	.062435415	128	.000487777	Prob > F	=	0.0000
				R-squared	=	0.9727
				Adj R-squared	=	0.9701
Total	2.28643463	140	.016331676	Root MSE	=	.02209

lISE	Coef.	Std. Err.	t	P> t	[95% Conf. Interval]	
lHotelGoogle	-.013794	.0151181	-0.91	0.363	-.0437076	.0161197
lRestauranteGoogle	.094666	.0172085	5.50	0.000	.060616	.128716
lEmpleoGoogle	-.0539396	.0153466	-3.51	0.001	-.0843055	-.0235737
lPeliculasGoogle	.0643885	.0122546	5.25	0.000	.0401406	.0886364
lEXPECTATIVASDEINFLACIONMENSUA	-.0177075	.0026012	-6.81	0.000	-.0228544	-.0125606
lTRMAFINDEMES	-.0412381	.0302029	-1.37	0.175	-.1009998	.0185235
lEXPECTATIVASPRODTRIMESTRE	-.02861	.0035197	-8.13	0.000	-.0355743	-.0216456
lEMPLEOTOTAL	.4136525	.1303516	3.17	0.002	.1557296	.6715754
lXM	.540177	.0869106	6.22	0.000	.3682096	.7121444
lEMBI	.0332404	.0141482	2.35	0.020	.0052457	.0612352
lCOMPRASTC	.1752185	.0275514	6.36	0.000	.1207033	.2297337
ICI	.0009443	.0004141	2.28	0.024	.000125	.0017635
_cons	-10.64398	1.66933	-6.38	0.000	-13.94704	-7.34093

Aquella que tuvo un mayor coeficiente fue la variable de empleo real de la encuesta mensual manufacturera; la cual al incrementarse en un 1%, provocó un incremento del ISE del 0.4136%. La siguiente variable que presentó un mayor coeficiente fueron las expectativas de TRM a fin de mes, la cual tuvo un crecimiento negativo que representó un decrecimiento del 0.4123% del ISE por un incremento del 1% en las expectativas de la TRM.

Para el ISE en términos logarítmicos, tanto el ICCO como el ICC dejaron de ser significativas para el modelo. Y en este caso, las compras con tarjetas de crédito reemplazaron a las compras con tarjetas de débito.

Con un Durbin Watson de 1.517716, hubo evidencia a favor de correlación serial positiva, ya que el estadístico es menor a 2. El test estadístico de Breusch Godfrey también probó la existencia de correlaciones seriales.

Tabla 4. Resultados test estadísticos Breusch Godfrey.

```
estat bgodfrey, lags(1/5)
```

```
Number of gaps in sample: 7
```

```
Breusch-Godfrey LM test for autocorrelation
```

lags(p)	chi2	df	Prob > chi2
1	1.886	1	0.1697
2	4.064	2	0.1311
3	5.207	3	0.1573
4	5.569	4	0.2337
5	5.815	5	0.3246

```
H0: no serial correlation
```

3

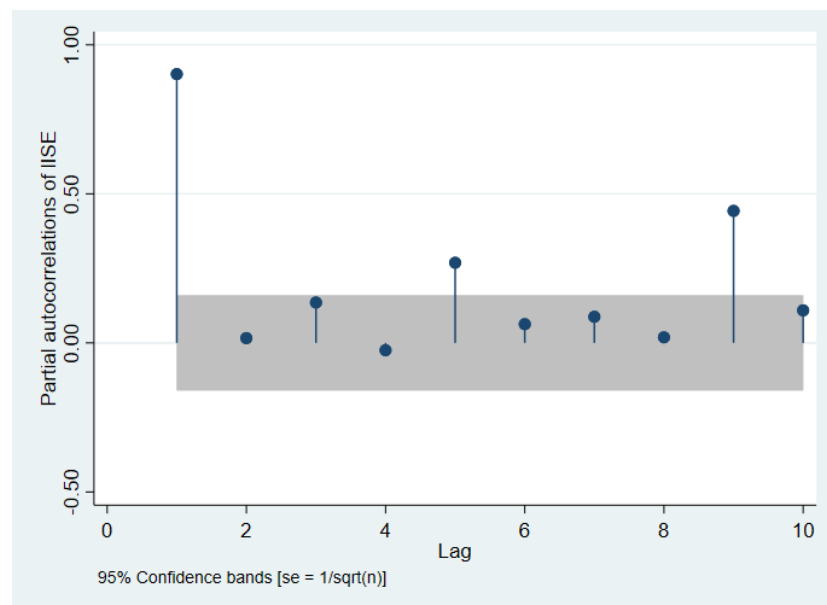
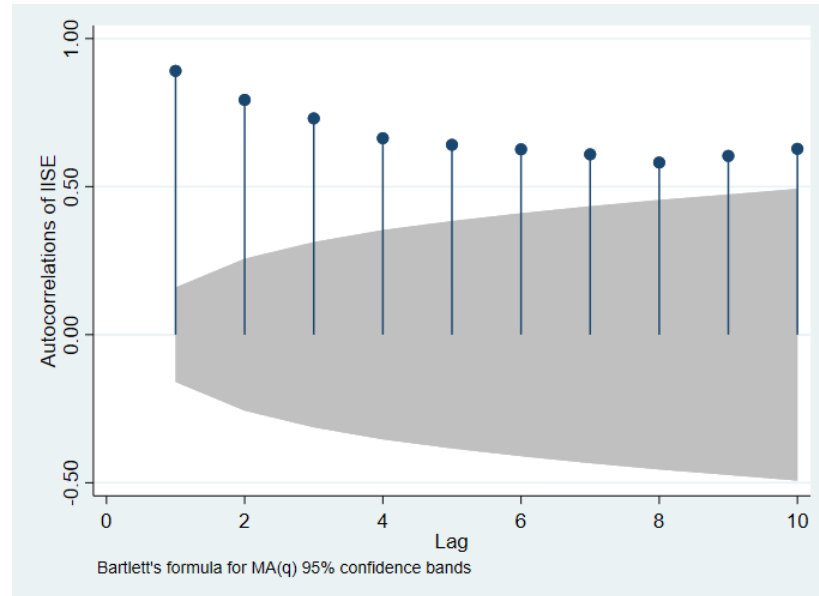
En el caso del Test de Breusch Pagan, que prueba la existencia de heterocedasticidad en el modelo, con un P-value de 0.0453, se rechazó la hipótesis nula de la existencia de heterocedasticidad en el modelo. El test de White confirma el resultado anterior de que el modelo es homocedástico, puesto que con un estadístico χ^2 de 0.4301, no se rechazó la hipótesis nula de homocedasticidad en el modelo.

El último test utilizado en el modelo es el test de Ramsey, el cual prueba si el modelo está correctamente especificado y si los β 's son insesgados y consistentes. Con un estadístico de 0.2253, no rechazo H_0 de que el modelo no tiene variables omitidas. Los β 's eran insesgados y consistentes. El modelo está correctamente especificado.

En cuanto a la identificación de autocorrelaciones para el modelo, se obtuvo que las series eran no estacionarias, lo que implica que poseen tendencia en el tiempo, y su varianza y media no variables en el tiempo; por ende, para poder correr la estimación requirieron de un proceso de estacionarización y transformación de datos. Corriendo los correlogramas para la variable del ISE, se pudo deducir que este seguía un proceso auto regresivo y de medias móviles por su comportamiento en el ac y el pac, pues presenta cortes en ambos gráficos.

Gráficos 1 y 2. Gráficos de autocorrelacion y correlacion parcial.

³ Elaboración del autor con cálculos propios.



4

Por ende, se dedujo que el modelo tiene presencia de componentes autorregresivos y de medias móviles, por lo que podría ser un ARMA o un ARIMA. En su estimación, se encontró que los modelos que más se acomodaban al comportamiento de la variable son ARIMA (1,1,1) y un ARMA (1,0,2). Esto implicó que los rezagos de un periodo de la variable determinan su comportamiento en el presente, y para el caso de su error, como lo expresan las medias móviles, los errores del modelo en su

⁴ Gráficos elaborados por el autor con cálculos propios.

primer o segundo rezago fueron significativos para la determinación de la variable. Esto implicó que tanto observaciones o valores pasados del crecimiento, como errores o choques aleatorios, son determinantes del desempeño de la variable en el presente.

Dados los resultados anteriores, se encuentra factible la estimación del modelo, dado que las series presentaron autocorrelación, son homocedásticas y estacionarias. Esto quiere decir que sus estimadores son insesgados y consistentes y que su varianza y media no dependen del tiempo. Por otro lado, los residuales del modelo siguen una distribución normal estándar, con media 0 y varianza 1.

Esto permitió avanzar al siguiente paso, el cual fue encontrar las correlaciones dinámicas entre las diferentes variables respecto al ISE. Estas correlaciones fueron halladas por medio de un proceso de transformación en donde se presentaron en variaciones interanuales mensuales y a cada una de estas variables se le aplicó el filtro de Hodrick y Prescott para eliminar el componente cíclico de los datos y poder estimar el comportamiento de las variables en periodos venideros.

Tabla 5. Correlaciones dinámicas de las variables hasta 5 periodos rezagados y 5 periodos adelantados respecto al ISE.

	I4	I3	I2	I1	t	f1	f2	f3	f4
Hotel	0.158	0.178	0.470	0.568	0.821	0.992	0.805	0.607	0.320
Moto	0.403	0.278	0.212	0.145	0.363	0.645	0.541	0.415	0.286
Usados	0.351	0.287	0.307	0.348	0.372	0.595	0.592	0.552	0.481
Empleo	0.478	0.297	0.273	0.143	-0.099	0.584	0.658	0.621	0.434
Restaurante	0.102	0.233	0.259	0.510	0.947	0.874	0.681	0.477	0.270
Películas	0.298	-0.029	0.017	-0.095	0.204	0.708	0.468	0.396	0.406
Casa	0.480	0.284	0.318	0.169	-0.045	0.223	0.192	0.132	0.218
Vuelos	0.132	0.158	0.285	0.425	0.762	0.950	0.807	0.625	0.301
PeI YouTube	0.407	0.171	0.191	0.069	0.080	0.180	0.007	0.044	0.085
XM	0.280	0.526	0.484	0.656	0.679	0.359	0.468	0.441	0.281
EMBI	0.324	0.279	0.181	0.058	0.043	0.102	0.189	0.278	0.302
ICI	0.427	0.539	0.505	0.494	0.488	0.499	0.599	0.499	0.309
Compras TC	0.082	0.209	0.257	0.509	0.986	0.826	0.537	0.436	0.237
Compras TD	0.144	0.207	0.200	0.263	0.976	0.876	0.613	0.424	0.212
Empleo	0.199	0.244	0.328	0.428	0.580	0.659	0.645	0.623	0.525
TRM	0.185	0.184	0.140	0.102	0.168	0.219	0.307	0.376	0.382
Sit acutal	0.178	0.238	0.343	0.609	0.784	0.754	0.698	0.562	0.341
Expect prod	0.747	0.736	0.501	0.171	0.011	0.157	0.358	0.362	0.273
ICC	0.337	0.354	0.454	0.569	0.674	0.588	0.427	0.289	0.149
ICCO	0.345	0.399	0.454	0.660	0.689	0.615	0.587	0.449	0.274
INFLACION	0.169	0.101	-0.053	-0.143	-0.055	0.379	0.749	0.749	0.753

5

⁵ Elaborado por el autor con cálculos propios.

Como se evidenció en los resultados, donde hay mayores correlaciones positivas es en el periodo t (presente) y los siguientes dos periodos. Lo que significó que estas variables tendrían a ser variables procíclicas rezagadas o contemporáneas al ISE. De aquí se concluyó que los agentes económicos formulan sus expectativas después de haber estudiado el panorama de la economía o evaluar el desempeño de ésta en el presente, y de allí formulan sus expectativas para el futuro. Con base a estos resultados, se categorizaron las variables como:

Procíclicas y adelantadas: Expectativas de producción del próximo semestre.

Procíclicas y contemporáneas: Restaurante, XM, Compras con tarjetas de crédito y débito, Situación actual de la economía, ICC, ICCO.

Procíclicas y rezagadas: Hotel, Inflación, Empleo, Vuelos, Moto, Usados, Películas,

Dentro de todas estas correlaciones, la variable con más alta correlación positiva para el modelo fueron las compras con tarjetas de crédito con una correlación de casi 1 en t respecto al ISE, seguida por compras con tarjetas de débito y la búsqueda de la palabra Restaurante. Como variables rezagadas, aquella que tuvo mayor correlación dinámica fue la palabra hotel seguida por vuelos, con correlaciones superiores al .95.

No se encontraron palabras que tuviesen una correlación negativa con el ISE. Mayores al 0.143 como lo fue la inflación. Sin embargo, los cuatro periodos previos al ISE no tuvieron correlaciones tan altas dentro de los indicadores de expectativas como lo fueron el periodo actual y los dos periodos siguientes a la formación del ISE.

En cuanto a las variables de expectativas convencionales como lo son los indicadores de confianza publicados por Fedesarrollo, El que tuvo mayor correlación dinámica fue el índice de confianza comercial, seguido por el índice de confianza del consumidor.

ESTIMACIÓN Y RESULTADOS

Como última fase del procedimiento de estimación, se transformaron todas las series mensuales y se suavizaron con el promedio móvil para construir un modelo dinámico de factores que sigue un proceso autorregresivo de orden 2 para los errores, seleccionando el de menor criterio bayesiano AIC y BIC. En síntesis, se encontró un factor que refleja de forma diaria del estado actual de la economía colombiana a partir de la búsqueda de los colombianos en Google. Este proceso de estimación se realizó a través de la herramienta Python. El proceso de estimación arrojó datos para 3 periodos adelante del último dato del ISE (agosto, septiembre y octubre). En esta estimación, los datos Arrojados para el modelo fueron de 11.96% para el mes de agosto, 9.760% para el mes de septiembre y 5.340 para el mes de octubre.

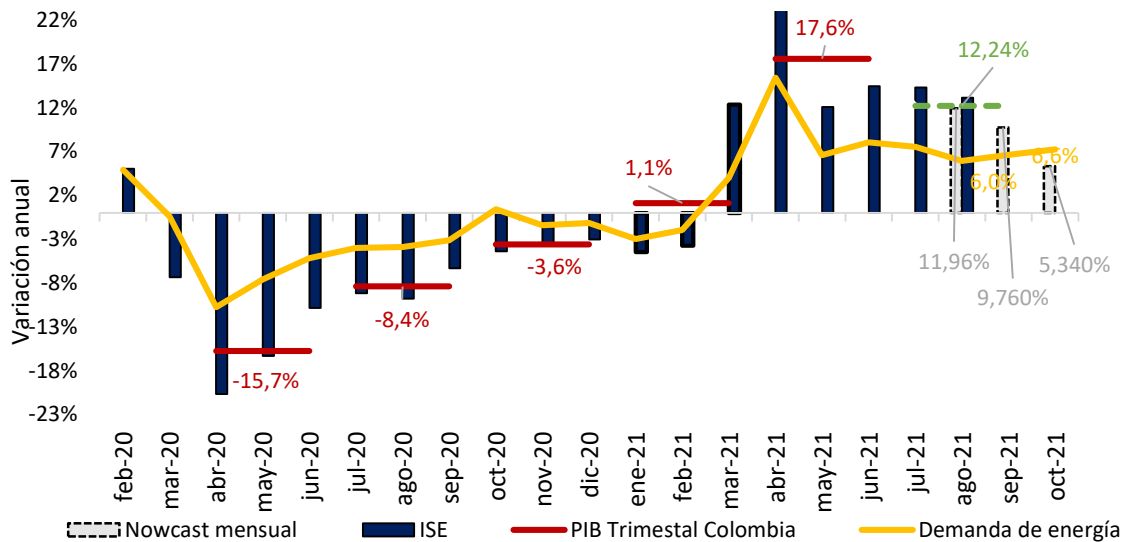
El mes de agosto fue estimado previamente al resultado del ISE publicado el 19 de octubre, el cual reportó una variación interanual mensual real del 13.2%. Esto permitió concluir que el pronóstico del nowcasting con expectativas de agentes económicos difirió del dato real en tan solo 1.24bps. Para el mes de septiembre se estimó un crecimiento del ISE del 9.76%.

Posterior a esto, se programó la segunda fase del modelo para encontrar el pronóstico trimestral del crecimiento de la economía en simulación del PIB. En este caso, se incorporó al modelo del nowcasting en su estimación y regresión las variables de expectativas de los agentes que se enunciaron en el modelo econométrico en términos logarítmicos. En este modelo se estimó el promedio del crecimiento del ISE para el tercer trimestre y cuarto trimestre del año.

En este procedimiento, se incorporaron las variables de expectativas de los consumidores, tanto convencionales como de alta frecuencia; las cuales fueron regresadas respecto al ISE y comprimidas en un vector de coeficientes, el cual fue utilizado para la predicción del PIB expresado en términos mensuales en un modelo que permitió la predicción de esta variable real de crecimiento mediante la minimización del error cuadrático medio.

Una vez encontrado el modelo más eficiente, se ejecutó la predicción del PIB con la introducción de los valores correspondientes a los meses de estimación que fueron arrojados en la primera fase para el ISE (un promedio entre agosto y septiembre para el tercer trimestre del año), y los valores reales de las variables para los meses a estimar. La estimación se realizó mediante un modelo OLS (Ordinary Least Squares Model) para cada trimestre a hallar, teniendo en total dos estimaciones, una para el tercer trimestre y otra para el cuarto trimestre con los datos disponibles para octubre.

Gráfico 3. Nowcasting de la economía colombiana.



6

Se encontró entonces en el nowcast un crecimiento estimado de crecimiento para el tercer trimestre del año de 12,235 %, y para el cuarto trimestre de 4,18%. Es importante destacar del resultado del tercer trimestre de 12,235% que continuó siendo positivo, netamente por un efecto base de una línea de comparación muy baja julio-septiembre del año 2020 (-8,38 %), mientras que para octubre-diciembre del 2020 se presentó un decrecimiento de apenas 3,59%.

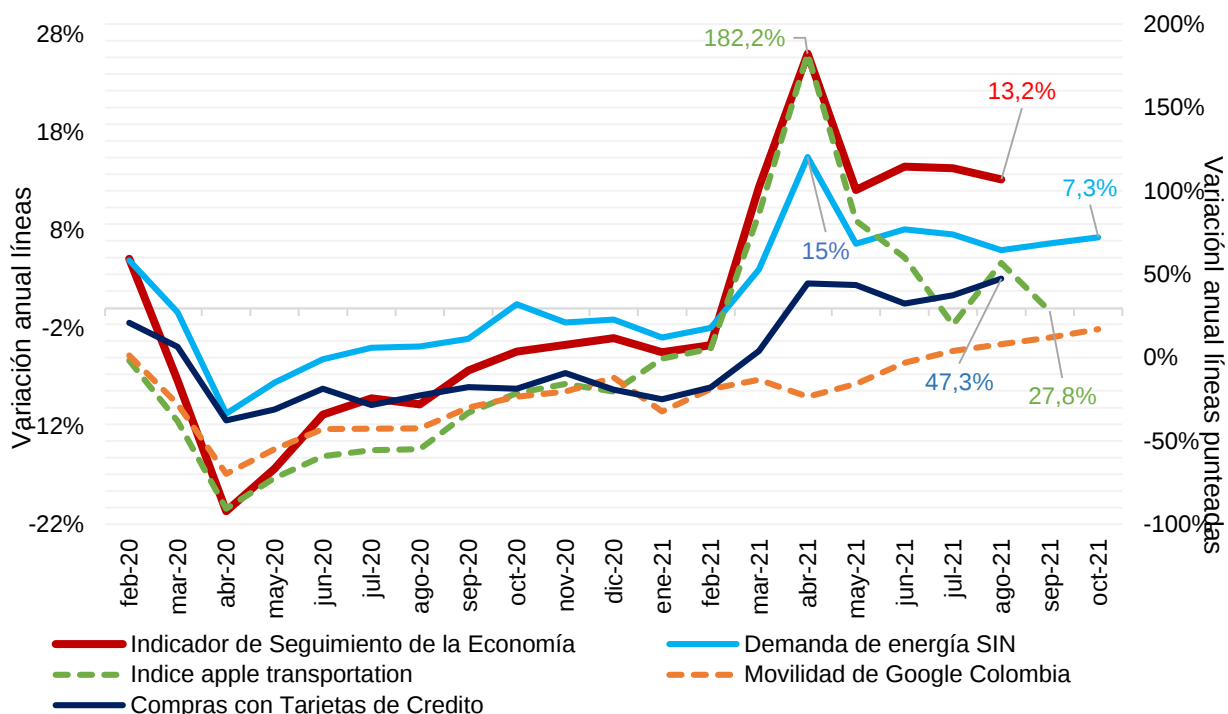
⁶ Elaborado por el autor con cálculos propios y datos tomados del DANE y el portal XM.

CONCLUSIONES

La herramienta del nowcasting, o de predicción inmediata, permitió estimar el comportamiento de la economía en tiempo real de manera eficiente y anticipada, difiriendo en solo 1.24bps del dato real publicado por el DANE. Al haber sido estimado con variables no convencionales, convencionales y de alta frecuencia que representaran las expectativas y opiniones de los agentes económicos; permitió afirmar que el comportamiento de estos agentes y su formación de opiniones y sentimientos sobre el desempeño de la economía son determinantes fundamentales para el crecimiento del país. Tanto la implementación del Machine Learning como la incorporación de indicadores no convencionales para el pronóstico del crecimiento son elementos innovadores en Colombia y poseen grandes ventajas para las metodologías de modelación y predicción por su bondad de ajuste y alta representatividad de la dinámica económica.

Se espera que la economía colombiana crezca en un rango entre el 8,25% y un 13,85% para el Q3 del 2021, con un valor más probable de 12,235%. Incluso sin datos disponibles de muchas variables, aquellas de alta frecuencia como el XM, google trends, google mobility e indicadores de Apple transportation; nos permitieron tener un seguimiento más cercano al crecimiento y encontrar modelos de factores que simularan el ISE con los datos disponibles. Adicional a esto, las correlaciones dinámicas de las variables demostraron que hay variables que son procíclicas, contracíclicas, adelantadas y rezagadas respecto al crecimiento, destacando principalmente las compras con tarjetas de crédito y débito y palabras como Hotel, Vuelos y Restaurante.

Gráfico 4. Gráfico de seguimiento de la economía colombiana y sus principales variables.



7

Como principal conclusión, y de acuerdo con el resultado esperado, es posible construir un modelo que estima de forma correcta y eficiente el comportamiento de la economía en tiempo real mediante indicadores de confianza y expectativas de los agentes económicos, combinando datos de diferentes frecuencias, convencionales y no convencionales.

⁷ Elaboración del autor con datos tomados del DANE, portal XM, Google trends, Google Mobility, Apple transportation y la superintendencia financiera de Colombia.

REFERENCIAS

- Askitas, N., & Zimmermann, K. F. (2009). Google econometrics and Unemployment forecasting.
- Asociación Nacional de Movilidad Sostenible (2021). Cifras y estadísticas.
- Banbura, M., D. Giannone, M. Modugno, and L. Reichlin. (2013). "Nowcasting and the Real-Time Data Flow." In G. Elliott and A. Timmermann, eds., *Handbook of Economic Forecasting*, Vol. 2. Amsterdam: Elsevier-North Holland.
- Bolhuis, M. A., & Rayner, B. (2020). *Deus ex Machina? A Framework for Macro Forecasting with Machine Learning*. IMF Working Papers.
- Bok, B., D. Caratelli, D. Giannone, A. Sbordone, and A. Tambalotti. (2017). "Macroeconomic Nowcasting and Forecasting with Big Data." Federal Reserve Bank of New York Staff Reports, no. 830, November.
- Cárdenas-Cárdenas, J. A., Caicedo-García, E., and González-Molano, E. R. (2020). Estimación de la variación del precio de los alimentos con modelos de frecuencias mixtas. *Borradores de economía*, Banco de la República, 1109.
- Cristiano, D., Hernández, M., and Pulido, J. D. (2012). Pronósticos de corto plazo en tiempo real para la actividad económica colombiana. *Borradores de economía*, Banco de la República, 724.
- Caruso, A. (2019). Macroeconomic news and market reaction: Surprise indexes meet nowcasting. *International Journal of Forecasting*, 35(4), 1725-1734.
- Claveria, O., Monte, E., & Torra, S. (2020). Economic forecasting with evolved confidence indicators. *Economic Modelling*, 93, 576-585.
- DANE, Departamento Nacional de Estadística. (2021). Estadísticas por tema.
- Diebold, F. X. (2003, February). Big data dynamic factor models for macroeconomic measurement and forecasting. In *Advances in Economics and Econometrics: Theory and Applications, Eighth World Congress of the Econometric Society*, (edited by M. Dewatripont, LP Hansen and S. Turnovsky) (pp. 115-122).
- Fedesarrollo, Encuesta Expectativas del Consumidor y Encuesta de Expectativa Empresarial. (2021).
- Fornaro, P., & Henri, L. (2020). Nowcasting Finnish real economic activity: a machine learning approach. *Empirical Economics*, 58(1), 55–71.
- Galeano-Ramírez, F. J., Martínez-Cortés, N., & Rojas-Martínez, C. D. (2021). Nowcasting Colombian Economic Activity: DFM and Factor-MIDAS approaches. *Borradores de Economía*; No. 1168.

León, C. and Ortega, F. (2018). Nowcasting economic activity with electronic payments data: A predictive modeling approach. *Revista de Economía del Rosario*, 21(2):381–407.

Lütkepohl, Helmut. 2007. *New Introduction to Multiple Time Series Analysis*. Berlin: Springer.

Pérez-Castañeada, G. C. (2009). Using the MIDAS approach for now- and forecasting colombian GDP. Master's thesis, Erasmus Universiteit Rotterdam.

Portal XM (2021). Información histórica. Demanda de energía SIN.

Stock, J. H., & Watson, M. W. (2004). Combination forecasts of output growth in a seven-country data set. *Journal of forecasting*, 23(6), 405-430.

Superintendencia Financiera de Colombia (2021). Informes mensuales de establecimiento de crédito.

Tilly, S., Ebner, M., & Livan, G. (2021). Macroeconomic forecasting through news, emotions and narrative. *Expert Systems with Applications*, 175, 114760.

Vuchelen, J. (2004). Consumer sentiment and macroeconomic forecasts. *Journal of economic psychology*, 25(4), 493-506.