

**UNIVERSIDAD CATÓLICA SANTO TORIBIO DE MOGROVEJO**  
**FACULTAD DE CIENCIAS EMPRESARIALES**  
**ESCUELA DE ECONOMÍA**



**GOOGLE TRENDS, INCLUSIÓN DE DATOS EN TIEMPO REAL PARA  
LA PREDICCIÓN DE VARIABLES MACROECONÓMICAS EN EL  
PERÚ**

**TESIS PARA OPTAR EL TÍTULO DE:  
ECONOMISTA**

**AUTORES**

**PILCO CORONADO, JEANCARLO JUNIOR  
SANDOVAL QUIROZ, KEYLA OFELIA**

**Chiclayo, 15 de Octubre de 2018**

**GOOGLE TRENDS, INCLUSIÓN DE DATOS EN TIEMPO  
REAL PARA LA PREDICCIÓN DE VARIABLES  
MACROECONÓMICAS EN EL PERÚ**

**POR:**

**Jeancarlo Junior Pilco Coronado**

**Keyla Ofelia Sandoval Quiroz**

Presentada a la Facultad de Ciencias Empresariales de la Universidad Católica  
Santo Toribio de Mogrovejo, para optar el Título de:

**ECONOMISTA**

**APROBADO POR:**

---

Mgtr. Antonio Gilberto Escajadillo Durand

Presidente de Jurado

---

Econ. Milagros Carmen Gamarra Uceda

Secretario de Jurado

---

Mgtr. Carlos Alberto León De La Cruz

Vocal/Asesor de Jurado

**CHICLAYO, 2018**

## **Dedicatoria**

En honor a nuestros padres que nos apoyaron incondicionalmente a lo largo de estos cinco años de estudios universitarios.

Y por sobre todo a Dios por habernos permitido llegar hasta este punto y lograr uno más de nuestros objetivos, además de su infinito amor y bondad.

## **Agradecimiento**

La presente investigación nos ha permitido aprovechar directa e indirectamente las competencias y experiencia de muchas personas a quienes agradecemos de todo corazón.

En primer lugar, a nuestro asesor Mgtr. Carlos León de la Cruz por haber depositado confianza en nosotros y viceversa. Sus conocimientos y trabajo han sido y seguirán siendo fuente de motivación.

Así también, nuestro más sincero agradecimiento a la Mgtr. Julia Gabriela Maturana Coronel por sus valiosas recomendaciones, su gran sabiduría, paciencia y apoyo profesional y emocional brindado durante todo este tiempo.

Agradecemos también, la colaboración del Dr. Ciro Eduardo Bazán Navarro. Fue una verdadera satisfacción contar con la ayuda de tan excelente profesional.

## **Resumen**

En el presente trabajo de investigación se realiza el análisis y evaluación del uso de información en tiempo real, brindada por Google Trends, para la predicción de variables macroeconómicas como Inflación e Inversión, mediante el cálculo de índices (IG) y el desarrollo de modelos ARMA (Autorregresivos de Media Móvil) y modelos VAR (Vectores Autorregresivos). Este estudio aplicado hace énfasis en la necesidad de entender más la actividad y el uso del tiempo de la población en línea ya que el internet contiene y ofrece una cantidad enorme y continua de datos que no presentan costo alguno a los cuales se puede acudir en cualquier momento deseado. Estos modelos fueron empleados y comparados al incluirse los índices calculados para mostrar la capacidad predictiva, los resultados obtenidos muestran que los modelos que incluyen los IG calculados presentan mejores cualidades predictivas en comparación con los modelos generales; en el caso de inversión los estadísticos de predicción se ven reducidos en mayor medida que en la Inflación, dando así un sustento a la importancia de tomar en cuenta este tipo de información como insumo para predicciones en el ámbito económico.

**Palabras clave:** Google Trends, Inflación, Inversión, Predicción.

**Clasificaciones JEL:** O12, O33

## **Abstract**

In this research the analysis and evaluation of the use of real time information, provided by Google Trends, for the prediction of macroeconomic variables such as inflation and investment by calculating index (IG) and the development of models is made ARMA (Autoregressive Moving Average) and VAR (Vector Autoregressive) models. This study applied emphasizes the need to understand more activity and time use of the online population as the Internet contains and offers a vast and continuous amount of data that do not present any cost which can go anywhere desired time. These models were used and compared to calculated indices included to show the predictive ability, the results show that models including the IG calculated have better predictive qualities compared to general models; in the case of investment statistical prediction they are reduced to a greater extent than in inflation, thus giving support to the importance of taking into account such information as input for predictions in the economic field.

**Keywords:** Google Trends, Inflation, Investment, Prediction.

JEL Classifications: O12, O33

# Índice

**Dedicatoria**

**Agradecimiento**

**Resumen**

**Abstract**

<b>I. Introducción</b> .....	10
<b>II. Marco teórico</b> .....	12
2.1. Antecedentes.....	12
2.2. Bases teórico científicas .....	14
<b>III. Metodología</b> .....	16
3.1. Tipo y nivel de investigación.....	16
3.2. Diseño de investigación.....	17
3.3. Población, muestra y muestreo .....	17
3.4. Criterios de selección.....	18
3.5. Operacionalización de variables .....	19
3.6. Técnicas e instrumentos de recolección de datos .....	10
3.7. Procedimientos .....	11
3.8. Plan de procesamiento y análisis de datos .....	11
3.9. Matriz de consistencia .....	12
3.10. Criterios éticos .....	12
<b>IV. Resultados y Discusión</b> .....	13
<b>V. Conclusiones</b> .....	26
<b>VI. Recomendaciones</b> .....	27
<b>VII. Referencias Bibliográficas</b> .....	28
<b>VIII. Anexos</b> .....	30
Anexo 1: Test de Causalidad en el sentido de Granger .....	30
Anexo 2: Test de Heterocedasticidad .....	30
Anexo 3: Función Impulso-Respuesta.....	31
Anexo 4: Descomposición de Varianza de la predicción VAR.....	32

## Índice de tablas

<b>Tabla 1:</b> <i>Variables endógenas analizadas (proxy)</i> .....	17
<b>Tabla 2:</b> <i>Palabras por Proxy de Inversión (por <math>\rho &gt;  0.7 </math>)</i> .....	18
<b>Tabla 3:</b> <i>Palabras por Proxy de Inflación (por <math>\rho &gt;  0.7 </math>)</i> .....	18
<b>Tabla 4:</b> <i>Operacionalización de la Variable dependiente</i> .....	19
<b>Tabla 5:</b> <i>Operacionalización de la variable independiente</i> .....	20
<b>Tabla 6:</b> <i>Matriz de consistencia</i> .....	12
<b>Tabla 7:</b> <i>Correlación (<math>\rho</math>) entre variable macroeconómica – índice</i> .....	13
<b>Tabla 8:</b> <i>Probabilidad de significancia estadística de los índices y el nivel de ajuste</i> .....	13
<b>Tabla 9:</b> <i>Palabras utilizadas que conforman los índices de mayor incidencia</i> .....	14
<b>Tabla 10:</b> <i>Especificación de los Modelos VAR</i> .....	17
<b>Tabla 11:</b> <i>Descomposición de Varianza de la predicción VAR</i> .....	19
<b>Tabla 12:</b> <i>Parámetros estimados y estadísticos de capacidad predictiva. Variable Endógena: Índice de Confianza Empresarial (ICE)</i> .....	21
<b>Tabla 13:</b> <i>Parámetros estimados y estadísticos de capacidad predictiva. Variable Endógena: IPC sin Alimentos (IPCsA)</i> .....	22
<b>Tabla 14:</b> <i>Prueba Diebold y Mariano</i> .....	23

## Índice de figuras

<b>Figura 1:</b> <i>Dispersión entre Inversión (ICE) – IG(I) Fuente: Elaboración propia</i> .....	15
<b>Figura 2:</b> <i>Dispersión entre Inflación (IPCsA) – IG(<math>\pi</math>) Fuente: Elaboración propia</i> .....	15
<b>Figura 3:</b> <i>Función Impulso-Respuesta Inversión. Respuesta de ICE ante IG(I) Fuente: Base de datos en Eviews. Elaboración propia</i> .....	18
<b>Figura 4:</b> <i>Función Impulso-Respuesta Inflación. Respuesta de IPCsA ante IG(<math>\pi</math>) Fuente: Base de datos en Eviews. Elaboración propia</i> .....	18
<b>Figura 5:</b> <i>Predicción para Inversión (ICE) Fuente: Base de datos en Eviews. Elaboración propia Notas: Variables tomadas en logaritmos</i> .....	24
<b>Figura 6:</b> <i>Predicción para Inflación (IPCsA) Fuente: Base de datos en Eviews. Elaboración propia Notas: Variables tomadas en logaritmos</i> .....	25

## I. Introducción

El estudio de los agregados económicos tiene importancia vital para los agentes económicos en cualquier país del mundo y con ello, la disponibilidad de datos que les permita observar las tendencias según se desenvuelven, puede mejorar la calidad de sus evaluaciones económicas y, a su vez, sus decisiones. Por lo cual, se hace deseable contar con estadísticas confiables que permitan monitorear constantemente la situación económica del país.

Por ejemplo, dadas las características del mercado laboral peruano, las tasas reportadas de desempleo no reflejarían la situación del mercado laboral en el Perú de manera precisa pues no incorporan los cambios que pueden estar relacionados con fenómenos sociales como la Internet. Por otro lado, con el nivel de digitalización que se experimenta actualmente, la oportunidad de conocer los intereses y preocupaciones de la población en tiempo real es ahora más factible. Aproximadamente 4.9 millones de peruanos alcanzan en promedio 21.5 horas de visitas online al mes, de los cuales el 38.8% se encuentra entre las edades de 15-24 años, el 26.3% entre 25-34 años, el 17.5% entre 35-44 años y el 11.1% entre 45-54 años (Fosk, 2011).

En ese sentido, una de las fuentes de información más estudiadas ha sido Google Trends, cuya actividad generada por las búsquedas online, permite relacionar el volumen de búsquedas de una determinada consulta en un área específica y a lo largo del tiempo, lo que resultaría en una herramienta útil para realizar un seguimiento de los cambios que ocurren en la economía casi de manera inmediata, sobre todo en épocas de crisis, como lo evidenciado por Suhoy (2009), quien confirma que una serie de categorías de Google Trends sí hubieran podido predecir, en una fecha muy cercana a la predicha por los datos oficiales, el declive económico que experimentó Israel en el 2008. Así, según consta en los antecedentes, ha sido exitosa la estrategia de incorporar como una variable a los modelos de pronósticos la intensidad en las búsquedas en Google debido a que refleja las intenciones y cuantifica las expectativas de los agentes económicos.

La presente propuesta implica el cálculo de un índice a partir de los datos obtenidos de Google Trends para diferentes variables macroeconómicas, en el supuesto de que dicha información ayudaría a monitorizar de manera más eficiente los agregados macroeconómicos peruanos. Por tanto, ¿Es el uso de información en tiempo real, brindada por Google Trends, útil para una predicción más certera de variables macroeconómicas en una economía como la Peruana?

Nuestra hipótesis es en efecto, la existencia de una mayor certeza en la predicción de los indicadores oficiales de variables macroeconómicas peruanas al incluir los índices elaborados con la información de Google Trends. Por ello, el objetivo general fue predecir en tiempo real, las variables macroeconómicas del Perú, en base a Google Trends, en otro modo, establecer que los índices (IG) tienen relevancia en la predicción de las variables macroeconómicas obtenidas de las fuentes tradicionales peruanas como el Banco Central de Reserva del Perú (BCRP) y el Instituto Nacional de Estadística e Informática (INEI).

La descripción de la relación del comportamiento de las variables macroeconómicas y los índices calculados, la identificación de los IG de mayor incidencia (a través de la correlación entre los índices y las variables macro dado el periodo de estudio), así como su interrelación ante un impulso no previsto para un pronóstico de corto plazo que brinde mayores elementos de juicio a los agentes económicos sobre la evolución de las variables, forman también, parte de los objetivos planteados en la presente investigación.

El desarrollo del tema se justificó en el aporte a la investigación académica peruana por su carácter innovador ante la clara necesidad de entender más el uso del tiempo y de la actividad de la población en línea. El único trabajo en el Perú en este ámbito es un análisis del desempleo llevado a cabo en el 2013 por Chang y Del Río. La inclusión de datos en las diversas y recientes investigaciones en torno a Google Trends ha demostrado resultados satisfactorios, sustentando su fiabilidad no solo en el ámbito económico, sino también en lo social, médico, financiero, entre otros; concluyendo en su mayoría, que los datos que brinda presentan el menor error de predicción (aún más en países donde el uso de internet es ampliamente extendido en la sociedad).

Además, en un mundo en rápida transformación como en el que vivimos, parece cada vez más crucial ser consciente de lo que está sucediendo en el presente con vistas a proponer soluciones en un futuro próximo. La justificación práctica que se deriva es el desarrollo de metodologías válidas que permitan hacer frente al tiempo empleado a recoger datos en un alcance geográfico nacional. Como ya se ha señalado, internet contiene y ofrece una cantidad enorme y continua de datos que no presenta costo alguno y se puede llevar a cabo en el tiempo deseado, lo que significa es extremadamente pequeña la línea de tiempo en la que la información es proporcionada con un margen de maniobra más amplio y que la econometría clásica todavía tiene que aprovechar adecuadamente.

## II. Marco teórico

### 2.1. Antecedentes

La literatura referente a Google Trends ha sido relativamente abundante en los últimos años, el estudio pionero fue el de Ginsberg et al. (2009), en el campo de la medicina. Su investigación asoció las búsquedas de 45 términos referentes a la gripe para predecir brotes dos semanas antes que los informes del CDC (Centro para el Control y Prevención de Enfermedades). Posteriormente, fue usado para evaluar el cambio en la capacidad predictiva al incluir información de búsquedas en Google; Choi y Varian (2009a), Song et al. (2013) y Carrière-Swallow y Labbé (2010) realizaron trabajos econométricos que incluyen al índice de Google Trends para predecir las ventas mensuales de las industrias en Hong Kong (ventas minoristas, ventas de automóviles, casas y viajes), la demanda por cuartos de hoteles en Charlotte, Estados Unidos y ventas de autos en Chile, respectivamente. Todos ellos encontraron que los modelos que incorporan dicho índice generan predicciones más acertadas en términos de ajuste y eficiencia, comparada con modelos que no la incluyen.

Así mismo, se ha demostrado la efectividad del uso de los datos de Google Trends para mejorar predicciones en variables micro y macroeconómicas. Al incluir un índice de Google Trends, Choi y Varian (2009b) demostraron que se mejora el ajuste de modelos que pretenden predecir la ocurrencia de solicitudes de beneficios por desempleo en Estados Unidos. Della Penna y Huang (2009) y Schmidt y Vosen (2009) construyeron un índice de confianza del consumidor para predecir el consumo privado en Estados Unidos; ambos trabajos evidencian que dicho índice es más preciso que los ya existentes (The University of Michigan Consumer Sentiment Index (CSI) y The Conference Board Consumer Confidence Index (CCI)). Guzmán (2011), por su parte, demuestra que Google Trends puede ser utilizado para predecir el nivel de inflación.

En la misma línea, Suhoy (2009) evalúa si una serie de categorías de Google Trends - contratación de personal, electrodomésticos, viajes, bienes raíces, alimentos y bebidas y belleza y cuidado personal- podrían haber predicho la caída en el PBI de Israel que ocurrió en 2008 y en efecto, concluyen que estas series sí hubieran podido predecir el declive económico que experimentó Israel en una fecha muy cercana a la predicha por los datos oficiales. Además, el índice de Google Trends para la categoría "contratación de personal" sirve como un indicador líder de la demanda por trabajo en ese país.

Así también, las investigaciones de D'Amuri y Marcucci (2009) y Bersier (2010) mostraron que al incluir el índice de Google, las predicciones del desempleo en Alemania y Estados Unidos mejoran, mientras que Chang y Del Río (2013) mejoraron la predicción contemporánea y un período hacia adelante del índice de empleo de Lima para empresas de 100 y más trabajadores (IE100) incluyendo un índice de Google; empero, no permite anticipar la senda futura para más de un periodo.

En Colombia, Mejía et al. (2013) construyeron indicadores sectoriales adelantados de actividad (ISAAC) con el propósito de pronosticar series de tiempo de PBI sectorial para cada una de las nueve ramas de actividad colombiana y un indicador agregado del PBI total, el ISAAC+. Su ejercicio permite conocer en cuáles sectores los índices de Google Trends tienen la mayor relevancia estadística para anticipar de manera adecuada el PBI.

La línea metodológica usada no es muy amplia y en su mayoría se ha hecho uso de modelos econométricos autorregresivos y sus derivados, con la aplicación de distintos test para las variables empleadas según ha sido el caso de análisis. Así, Varian y Choi (2009a) utilizaron un modelo autorregresivo estacional para comprobar la hipótesis de que el volumen de búsqueda de las categorías que corresponden a ventas minoritarias, ventas de automóviles, casas y viajes podría estar correlacionado con el nivel de actividad económica en esas industrias. Ajustaron el modelo simple de predicción con una variable dummy para suavizar el efecto de quiebres en la serie e hicieron uso del Error Medio Absoluto para verificar si la información de Google Trends ayudaba en la predicción de las series.

Siguiendo lo explicado anteriormente, la metodología de modelo autorregresivo integrado de media móvil (ARIMA) fue usado por D'Amuri y Marcucci (2009) para predecir el desempleo en EE.UU.; en contraste con Askitas y Zimmerman (2009) quienes realizaron un trabajo similar en Alemania utilizando modelos de corrección de errores con cuatro palabras clave relacionadas a la búsqueda de empleo. En Perú, Chiang y Del Río (2013) utilizaron modelos ARMA y ARDL haciendo predicciones dentro y fuera de la muestra con el índice de Google de desempleo (IGD) calculado previamente; la ausencia del sesgo del IGD se evaluó mediante un modelo de corrección de errores.

Por otro lado, Della Penna y Huang (2009) construyen un índice de confianza del consumidor sobre la base de búsquedas (SBI, por sus siglas en inglés). Demuestran que el SBI anticipa (predice) cambios en el CSI y en el CCI, como en predecir el crecimiento en el gasto de consumo personal y en ventas minoristas. Mientras que Schmidt y Vosen (2009) comparan

cómo varía el poder predictivo del modelo simple al agregar cada indicador de confianza del consumidor, primero con CSI, luego con el CCI y tercero con el índice Google.

Song et al. (2013) por su parte, utilizan cuatro modelos econométricos: modelo de rezagos distribuidos (ARDL), modelo autorregresivo integrado de media móvil (ARIMA), modelo de suavización espacial (ES) y modelo de parámetro variante en el tiempo (TVP) para predecir la demanda de hoteles en Charlotte, EE.UU. En tanto que Suhoj (2009) encuentra con probabilidades bayesianas dentro de la muestra que las series de categorías de Google Trends: contratación de personal, electrodomésticos, viajes, bienes raíces, alimentos y bebidas y belleza y cuidado personal, sí pudieron haber predicho la caída del PBI de Israel en el 2008 antes que los datos oficiales.

## 2.2. Bases teórico científicas

Las predicciones realizadas sugieren que la capacidad predictiva aumenta más en el modelo con el índice de Google en relación a otros modelos y por tal, en base a lo elaborado por otros autores, los índices se construyen empleando dos metodologías teóricas modificadas a criterio de los investigadores.

- A. Selección de palabras claves: listado de palabras o frases que intuitivamente tienen una relación con la variable macroeconómica a estudiar. Se eligen las que generan la mayor correlación con el índice vigente.
- B. Mensualización de cada serie semanal: Debido a que la frecuencia de las variables endógenas es mensual y los datos de Google Trends tienen frecuencia semanal, se convertirán las series de frecuencia semanal a mensual a través de promedios simples.
- C. Indexación de las series: Primero se calculan el promedio y la desviación estándar para cada palabra clave "n". Luego, se obtienen los ponderadores " $\alpha_n$ " a partir de la división entre la desviación estándar de una palabra clave y la suma de las desviaciones estándar de las "n" palabras clave (1). Finalmente, los índices se obtienen a partir de la suma producto de las observaciones de las palabras clave en el mes t y sus ponderadores (2).

$$\alpha_n = \frac{\sigma_n}{\sum_{n=1}^N \sigma_n} \quad (1)$$

$$IG_t = \sum_{n=1}^N x_{tn} \alpha_n \quad (2)$$

La segunda metodología refiere al índice complejo ponderado de Laspeyres que combina magnitudes simples para conformar lo generalmente llamado índice sintético de Laspeyres. Éste se conformará por índices ya calculados por Google que están en función de la magnitud compleja  $H$ , siendo  $H$  el número de consultas en el instante “ $t$ ” por palabra “ $n$ ” (series mensualizadas por un promedio simple).

A. Indexación ponderada: Se tiene en cuenta el número de links resultantes de las consultas de la palabra clave “ $n$ ” para calcular su importancia relativa a la que denominaremos “ $w_n$ ”. Los índices Laspeyres se obtienen a partir de la suma producto de los ponderadores y los índices calculados por Google para cada palabra clave en el mes  $t$ .

$$L_t(H) = \sum_{n=1}^N w_n I_{nt}(H) ; n = 1, 2, \dots, N ; t = 0, \dots, T \quad (3)$$

Por otro lado, el modelo planteado respecto a la metodología ARMA sigue la forma:

$$y_t = \beta_0 + \sum_{i=1}^p \beta_i y_{t-i} + \epsilon_t + \sum_{i=1}^q \alpha_i \epsilon_{t-i} \quad (4)$$

Donde  $y_t$  es la variable endógena para el periodo  $t$  y  $\epsilon_t$  es el error. Se elegirá el mejor modelo según los siguientes criterios: significancia económica y estadística de las variables y ajuste. (Trujillo, 2010; Gujarati, 2010; Johnston, 2011)

Después de la elección del mejor modelo ARMA, se replicará la misma especificación en un modelo aumentado que incluirá al Índice de Google, así:

$$y_t = \beta_0 + \sum_{i=1}^p \beta_i y_{t-i} + \epsilon_t + \sum_{i=1}^q \alpha_i \epsilon_{t-i} + \sum_{i=t}^T \gamma_i I G_t \quad (5)$$

Así también, los estadísticos de predicción utilizados son la Raíz del Error Cuadrático Medio (RECM) y el Error Absoluto Medio (EAM). (Castro y Rivas-Llosa, 2005)

$$\sqrt{ECM} = \sqrt{\frac{1}{H} \sum_{t=T+1}^{T+H} \hat{e}_t^2}$$

$$EAM = \frac{1}{H} \sum_{t=T+1}^{T+H} |\hat{e}_t|$$

Finalmente, se emplea la modelación VAR y la estimación de la función de impulso-respuesta traza el efecto corriente y valores futuros ante un shock aleatorio en alguna de las variables así como el impacto en periodos posteriores. A la par con ello, se descompone la varianza para medir el error de predicción de la función impulso-respuesta en los periodos evaluados.

### **III. Metodología**

#### **3.1. Tipo y nivel de investigación**

La investigación como un conjunto de procesos sistemáticos, críticos y empíricos aplicados al estudio de un problema o fenómeno, presenta de acuerdo con Fernández, Hernández y Baptista (2014), lo siguiente:

**Enfoque:** La investigación tiene un enfoque mixto: cualitativa – cuantitativa. Ello dado que presenta características, proceso y bondades cuantitativas como medición de un fenómeno, uso de la estadística, prueba de hipótesis, es deductiva, probatoria, replicable, predicción; y cualitativas como planteamientos más abiertos que van enfocándose, profundidad de significados que se extraen de los datos, inductivo, analiza realidades subjetivas, presenta amplitud y riqueza interpretativa.

**Tipo:** Investigación aplicada, ya que el propósito fundamental de este tipo de investigación es la de resolver un problema.

**Nivel:** investigación exploratoria, descriptiva, correlacional, explicativa. Los autores explican que las investigaciones en un campo específico del conocimiento pueden incluir diferentes niveles en su desarrollo, por lo que es posible iniciar como una investigación exploratoria, ser luego descriptiva y correlacional, para terminar como explicativa.

*Exploratoria*, porque se indagó un hecho con poca literatura e indagación desde una perspectiva innovadora (inclusión de datos de Google Trends), lo cual prepara el terreno del conocimiento para futuros estudios con ampliaciones.

*Descriptiva*, considera al fenómeno en su conjunto, se enfoca en comprender el comportamiento de las variables.

*Correlacional*, dado que se asociaron conceptos causa efecto, para predecir y cuantificar una variable a través de otra.

*Explicativo*, porque se va más allá de la descripción de los sucesos, se responde a las causas de éstos, los por qué de sus relaciones y en qué condiciones.

### 3.2. Diseño de investigación

Es no experimental porque se estudiaron a las variables conforme a su realidad observable y no se ejecutó sobre ellas manipulación alguna. Es longitudinal ya que los datos del estudio han sido observados y estudiados en varios periodos de tiempo (explicado en el punto 3.3. Población, muestra y muestreo). Es prospectivo dado que el carácter último de la investigación expresado en su objetivo general, fue la predicción de las variables macroeconómicas en tiempo real, por lo cual se realizaron predicciones a corto plazo del comportamiento de las mismas; y además, evaluación impulso-respuesta.

### 3.3. Población, muestra y muestreo

La población fueron las bases de datos generales de las variables endógenas Inversión e Inflación y las bases de datos de las palabras asociadas a la definición de dichas variables.

La muestra fue determinada bajo muestreo no probabilístico por conveniencia, siendo conformada por datos mensuales de los periodos 2008-2013 y 2008-2014 de las variables macroeconómicas Inversión, Inflación (variables endógenas) según las siguientes proxy:

**Tabla 1:**

*Variables endógenas analizadas (proxy)*

INVERSIÓN	INFLACIÓN
	IPC Lima
Formación de Capital	IPC Alimentos y Energía
Inversión en Maquinaria	IPC Subyacente Lima
Inversión Nacional en Soles	IPC No Subyacente Lima
Índice de Confianza Empresarial	IPC Sin Alimentos y Energía Lima
	IPC Importado

*Fuente: Elaboración Propia*

Y en el caso de las variables exógenas los índices elaborados a partir de la información extraída de Google Trends en el mismo periodo y de las siguientes palabras:

**Tabla 2:**

*Palabras por Proxy de Inversión (por  $\rho > |0.7|$ )*

<b>INVERSIÓN</b>	<b><math>\rho &gt; 0.7</math></b>	<b><math>\rho &lt; -0.7</math></b>
	«Precios»	«Turismo»
Formación de Capital	«TC»	«Investment (USA)»
	«Precios»	«Turismo»
Inversión en Maquinaria	«TC»	«Investment (USA)»
	«Precios»	«Bienes Raíces»
Inversión Nacional en Soles	«Liquidez»	«Investment (USA)»
		«Crisis Financiera»
Índice de Confianza Empresarial		«Crisis Económica»

*Fuente: Elaboración Propia*

**Tabla 3:**

*Palabras por Proxy de Inflación (por  $\rho > |0.7|$ )*

<b>INFLACIÓN</b>	<b><math>\rho &gt; 0.7</math></b>	<b><math>\rho &lt; -0.7</math></b>
	«Inflación Perú»	
IPC Lima (2009 = 100)	«Dólar»	«Créditos»
IPC Alimentos y Energía	«Arroz»	«Exportaciones»
IPC Subyacente Lima (2009 = 100)	«Carne»	«Inflación»
IPC No Subyacente Lima (2009=100)	«Pollo»	«Balanza Comercial»
IPC Sin Alimentos y Energía Lima (2009=100)	«Tipo de Cambio»	«IPC»
	«Dinero»	
	«Inflación Perú»	
IPC Importado (2009 = 100)	«Dólar»	
	«Pollo»	

*Fuente: Elaboración Propia*

### **3.4. Criterios de selección**

Para las variables endógenas: El periodo de tiempo seleccionado se justifica por que la información de Google Trends para las variables trabajadas en el caso peruano está disponible desde esa fecha.

Para las variables exógenas: Palabras correlacionadas a un nivel superior o igual a 0.7 (Tablas 2 y 3); y Correlación con sentido congruente con teoría económica.

### 3.5. Operacionalización de variables

**Tabla 4:**

*Operacionalización de la Variable dependiente*

Variable	Definición Conceptual	Dimensiones	Indicadores
<b>Variables Macroeconómicas</b>	Es la representación de un concepto macroeconómico que ayudan a explicar y entender el comportamiento de la economía de un país (Blanchard et al, 2012)	<b>Inversión</b>	Formación de Capital
			Inversión en Maquinaria
			Inversión Nacional en Soles
			Índice de Confianza Empresarial
			IPC Lima (2009 = 100)
		<b>Inflación</b>	IPC Alimentos y Energía
			IPC Subyacente Lima (2009 = 100)
			IPC No Subyacente Lima (2009=100)
			IPC Sin Alimentos y Energía Lima (2009=100)
			IPC Importado (2009 = 100)

*Fuente: Elaboración Propia*

**Tabla 5:***Operacionalización de la variable independiente*

<b>Variable</b>	<b>Definición Conceptual</b>	<b>Dimensiones</b>	<b>Indicadores</b>
		<b>Índice de Google de Inversión</b> <i>IG(I)</i>	Crisis Financiera Crisis Económica
<b>Datos en tiempo real</b>	Cantidad de información digital acumulada con una mayor disponibilidad de datos operativos activos mediante tiempo real (Malvicino y Yoguel, 2015)	<b>Índice de Google de Inflación</b> <i>IG(<math>\pi</math>)</i>	Créditos Exportaciones Inflación Balanza comercial IPC

*Fuente: Elaboración Propia*

### 3.6. Técnicas e instrumentos de recolección de datos

En base al enfoque macroeconómico desarrollado, se requiere abarcar alcances nacionales, por lo que resulta común entonces, el uso fuentes secundarias por el extenso y arduo trabajo que implicaría obtener los datos de manera primaria.

Los datos provenientes de la fuente Google Trends almacenan las búsquedas en internet de los individuos, lo cual implica que el análisis está básicamente enfocado a la población que hace uso de internet. En el Perú, éste no es universal y es fundamentalmente la región de la costa la que está más inmersa en tal proceso de globalización (por diferencias culturales, entorno socioeconómico, etc.). Debido al desarrollo de las sociedades, existe hoy, una conducta normada de buscarlo todo en internet, lo que implica que los datos que extraemos están sesgados por quienes hacen búsquedas de tipo académico, entretenimiento u otros fines, que si bien para otras líneas de investigación ello es sumamente importante, en nuestro caso no lo es tanto.

Por el tipo de data (secundaria) no se presenta el problema de reactividad ya que las búsquedas en internet son una aproximación más precisa de los verdaderos gustos, preferencias, deseos, convicciones e ideales que cada individuo refleja y por tanto, no hay resistencia a proporcionar información o propensión a modificar su comportamiento para dar una mejor impresión. Respecto de la confiabilidad de este proceso de recolección de datos provenientes de Google Trends, el respaldo se tiene de las investigaciones en distintas partes del mundo, aplicado a diferentes áreas (económica, financiera, social, médica, entre otras) y por personalidades de reconocimiento a nivel mundial.

En su mayoría, ciertas características u orientación tanto del investigador como del entrevistado tienden a tener efectos sobre lo investigado. En el caso de Google Trends no se tiene este tipo de efectos pues no hay un vínculo directo con el entrevistado; de hecho, no se pueden hacer distinciones negativas de género, estatus, estilos o efectos raciales, ni influir en la motivación y seriedad de las respuestas; tan similar como que el participante se presenta voluntariamente al hacer uso del buscador, sin patrones y roles preponderantes ni saturación de por medio.

Por otro lado, la información recogida por el INEI, así como del Banco Central de Reserva del Perú (entidad de prestigio mundial) reúne una cantidad considerable de variables posibles para seleccionar las variables endógenas. Estas entidades no son en efecto, extrañas o desconocidas para el entrevistado, lo que puede influir en la motivación de las respuestas, pero, se apela a la seriedad de quienes responden estas encuestas (funcionarios del sistema financiero,

analistas económicos, ejecutivos de empresas no financieras, población en general) y al hecho de que dicha información está controlada y monitoreada, lo que realza la fiabilidad de la fuentes por prestigio e innumerables usos en investigación económica y social, minimizando los efectos que el investigador primario pudiese tener sobre los entrevistados y los efectos de éstos últimos sobre los datos.

### **3.7. Procedimientos**

La descarga de datos es a través de una herramienta online adjunta a Google Trends, denominada Google Insights, que permite la exportación de los datos en formato CSV (compatible con Microsoft Excel) y no presenta efectos del instrumento por no tener que responder a preguntas en particular. Así también, se extrae información de la base de datos correspondiente a las fuentes tradicionales de las cuentas nacionales peruanas: BCRP e INEI.

### **3.8. Plan de procesamiento y análisis de datos**

Para el análisis de las variables, la elección se dio tras evaluar distintas series de datos de posibles variables endógenas para cada una de las variables macroeconómicas Inversión e Inflación (Tabla 2). Sobre la base de los datos de Google Trends, se construyeron índices por ambas metodologías presentadas (Chiang y Del Río - Laspeyres) utilizando un bloque de palabras referidas a cada variable real de análisis (Tabla 3), para luego ser introducidos en los modelos econométricos especificados para la predicción (ARMA), llegar a una única elección, es decir, un índice por variable macroeconómica con las nomenclaturas IG seguido de la inicial de la variable en estudio. Con este índice se procede a realizar un análisis econométrico en relación con los datos de las variables ya validados.

Se evaluaron estadísticos de predicción como el error cuadrático medio (RECM) y el error absoluto medio (EAM). Asimismo, se empleó la prueba Diebold y Mariano para dar más robustez a los resultados, así como el test de Causalidad de Granger (Anexo 1) y el de heterocedasticidad de White (Anexo 2).

### 3.9. Matriz de consistencia

**Tabla 6:**

*Matriz de consistencia*

<b>Problema principal</b>	<b>Objetivo general</b>	<b>Hipótesis principal</b>	<b>Variable independiente</b>
¿Es el uso de información en tiempo real, brindada por Google Trends, útil para una predicción más certera de variables macroeconómicas en una economía como la peruana?	Predecir en tiempo real, las variables macroeconómicas del Perú, en base a Google Trends	Existe una mayor certeza en la predicción de los indicadores oficiales de variables macroeconómicas peruanas al incluir los índices elaborados con la información de Google Trends	Datos en tiempo real  <b>Variable dependiente</b>  Variables macroeconómicas

*Fuente: Elaboración Propia*

### 3.10. Criterios éticos

Como criterios éticos se tuvieron la Fiabilidad y la Validez, en la recolección de los datos, lo cual garantizó resultados merecedores de crédito y confianza.

La Fiabilidad refiere la posibilidad de ser un estudio replicable, asegurando resultados verdaderos, pero independientes de las circunstancias de la investigación.

La Validez por su parte, concierne a la correcta interpretación de los resultados a partir de un bagaje teórica que sustentan los hallazgos o dan pie a futuras investigaciones.

#### IV. Resultados y Discusión

A fin de determinar qué índices de los calculados tienen mayor incidencia en las variables macroeconómicas peruanas Inversión e Inflación (representadas por el Índice de Confianza Empresarial (ICE) y el Índice de Precios al Consumidor sin alimentos (IPCsA)), se procedió a correlacionar las series y especificar los modelos ARMA (1,0) y ARMA (1,1) para Inversión e Inflación, respectivamente y así evaluar el nivel de ajuste y la probabilidad de significancia estadística de los índices. De esto se obtuvo que los índices calculados por la metodología usada por Chang y Del Río (2013) resultaron ser los de mayor incidencia en las variables (Tablas 7 y 8). A partir de este punto denominaremos como IG (I) e IG ( $\pi$ ) a los índices seleccionados para cada variable real Inversión e Inflación.

**Tabla 7:**

*Correlación ( $\rho$ ) entre variable macroeconómica – índice*

Variables Macroeconómicas	Índice Google Chang y Del Rio (IGR)		Índice Google Ponderador Laspeyres (IGL)	
	<i>Positivo</i>	<i>Negativo</i>	<i>Positivo</i>	<i>Negativo</i>
	<i>Inversión</i> <sup>1/</sup>		<b>-0.6789***</b>	
<i>Inflación</i> <sup>1/</sup>	<b>0.8794***</b>	<b>-0.8712***</b>	<b>0.9674***</b>	<b>-0.8298***</b>

*Fuente: Elaboración Propia*

**Tabla 8:**

*Probabilidad de significancia estadística de los índices y el nivel de ajuste*

Variables Macroeconómicas	Índice Google		$R^2$ *
	Chang y Del Rio (IGR)	Ponderador Laspeyres (IGL)	
	<i>Negativo</i>	<i>Positivo</i>	
<i>Inversión</i>	<b>0.0035<sup>1</sup></b>		0.8360***
<i>Inflación</i>	<b>0.0076<sup>1</sup></b>	0.2465 <sup>1</sup>	0.9980***

*Fuente: Elaboración Propia*

*Notas: \*  $R^2$  obtenido de la regresión ARMA para toda la muestra con el IG significativo.*

Estos resultados pueden explicarse probablemente por las palabras empleadas en la conformación de los índices, ya que su búsqueda acapara expectativas de los agentes económicos sobre acontecimientos en su entorno. Las palabras del IG (I) que demuestran un interés sobre posibles shocks económicos y financieros, se relacionan con las 10 dimensiones de la actividad económica (país-sector-empresa) que componen el ICE, lo que influiría en la

toma de decisiones de inversión empresarial. El IG ( $\pi$ ), por su parte, al conformarse por palabras no vinculadas a bienes de consumo, reflejaría el comportamiento de los agentes sobre el IPCsA que es una variable medida a través de una serie temporal sin influencia del consumo intertemporal de la canasta básica (Tabla 9).

**Tabla 9:**

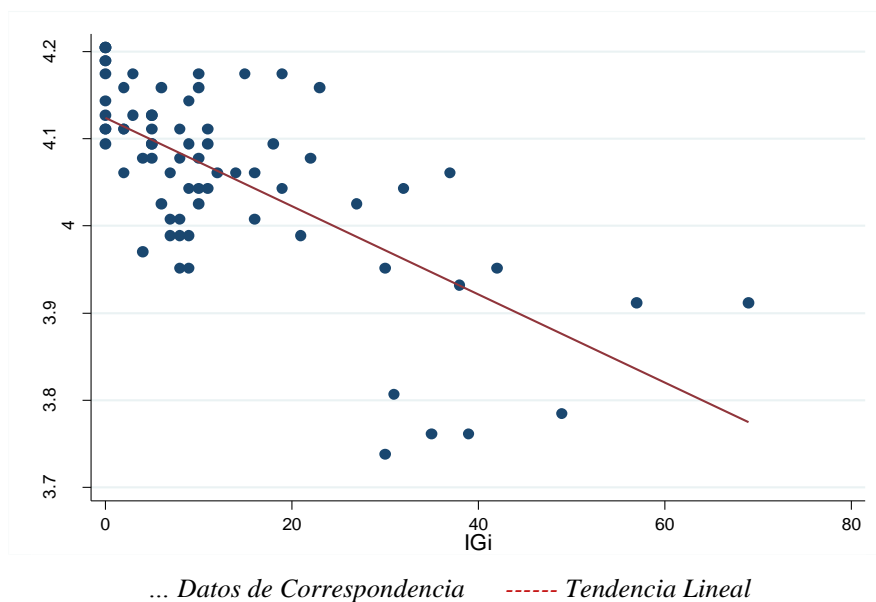
*Palabras utilizadas que conforman los índices de mayor incidencia*

<i>Inversión</i>		<b>IG(I)</b>	
<i>(índice de confianza empresarial)</i>		« Crisis Financiera »	
		« Crisis Económica »	
<i>Inflación</i>		<b>IG(<math>\pi</math>)</b>	
<i>(IPC sin Alimentos)</i>		« Créditos »	« Balanza Comercial »
		« Exportaciones »	« IPC »
		« Inflación »	

*Fuente: Elaboración Propia*

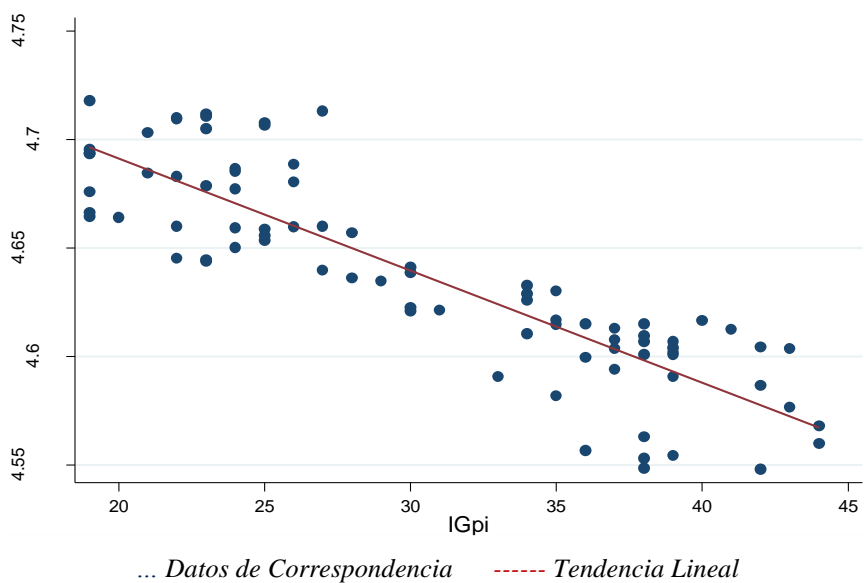
Estos resultados, sin embargo, pueden presentar algunas consideraciones que impliquen menor robustez, como la parcial subjetividad en la elección de palabras o el descarte de otras posibles variables proxy de inversión e inflación. Empero, bajo un enfoque de correlación y significancia estadística, no se invalidan los resultados hallados, además de no ser disonantes con otras investigaciones en cualquier campo (económico, social, financiero, etc.). La correlación mostrada entre sus variables de estudio y los índices de búsqueda en Google es muy significativa, como en Suhoy (2006); Askitas y Zimmerman (2009), Dzienlinski (2010), Redondo (2013), entre otros.

En ese sentido, se hizo una inspección visual para describir la relación de comportamiento de nuestras variables y se obtuvo una relación inversa entre ambas variables reales y sus índices (resultados asociadas a diversos factores no necesariamente causales). Por ejemplo, el ICE tuvo su peor tramo de caída en el periodo 2008-2009 y en paralelo, el IG (I) reflejó un tramo de valores muy altos (Figura 1). Así también, el IPCsA, en todo el periodo de estudio reflejó un comportamiento creciente poco variante que conjuga con una tendencia decreciente del IG ( $\pi$ ) (Figura 2).



**Figura 1:** *Dispersión entre Inversión (ICE) – IG (I)*

*Fuente: Elaboración propia*



**Figura 2:** *Dispersión entre Inflación (IPCsa) – IG ( $\pi$ )*

*Fuente: Elaboración propia*

Una explicación a la relación en la trayectoria de las variables y sus índices es lo suscitado en la historia económica de un país que influyen en el accionar de sus agentes económicos como son los periodos de crisis, alcances de precios históricos de commodities, cifras alarmantes en las bolsas de valores del mundo, entre otros que se reflejan en un comportamiento de preocupación.

Así, una de las posibles vías de enlace es que a medida que los agentes se preocupan por el estado económico del país (búsqueda de frases: «Crisis Financiera», «Crisis Económica»)

tendrán expectativas de una economía en peligro (para la inversión) y consecuentemente esto se traduciría en retrasos o cancelaciones de decisiones de inversión. Así mismo, consultar palabras claves como «Créditos», «Inflación», «Balanza Comercial», etc. también reflejaría expectativas de malestar económico, incurriéndose en mayor consumo presente por temer a cambios futuros no esperados; esto eleva la demanda interna y provoca caída de precios por conductas impacientes.

Cabe recalcar que esas posibles vías de enlace son válidas por respaldar la conducta de los agentes bajo expectativas, pese a no poder ser generalizadas ya que los motivos de búsqueda en internet son variados y no solo preocupación por el contexto macroeconómico del país. La relación de comportamiento entre estas variables concuerdan con lo encontrado en Mejía et al. (2013), Kato y Suárez (2007), Figueroa (s.f.), Chang y Del Río (2013); éste último observó para Perú, que a medida que el índice de desempleo incrementa, hay mayores búsquedas de empleo en internet.

Por otro lado, para establecer la dinámica de interrelaciones entre las variables macroeconómicas y los índices calculados, se procedió a realizar un análisis VAR. Tanto para Inversión como para Inflación, se encontró una bondad de ajuste satisfactoria y más alta, que para los índices (Tabla 10). Mientras que la significancia global de los coeficientes concernientes a los rezagos se hallaron estadísticamente distintos de cero (sobre la base del test F).

**Tabla 10:***Especificación de los Modelos VAR*

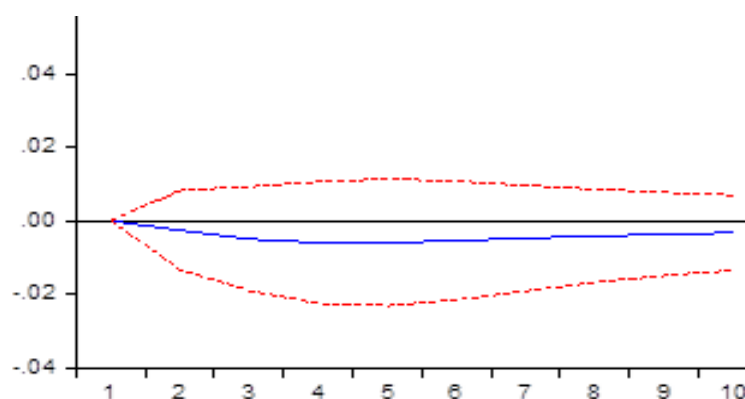
<i>Inversión</i>			<i>Inflación</i>		
	ICE	IG(I)		IPCsA	IG( $\pi$ )
	0.926651	35.24813		0.844262	236.4111
	-0.1324	-22.8456	IPCsA(-1)	-0.14594	-207.175
ICE(-1)	[ 6.99886]	[ 1.54288]		[ 5.78493]	[ 1.14112]
			.	.	.
	-0.077914	-64.86552	.	.	.
	-0.13095	-22.5951	.	.	.
ICE(-2)	[-0.59500]	[-2.87078]	IPCsA(-12)	0.054224	-609.3775
				-0.12	-170.35
	-0.000347	0.862528		[ 0.45187]	[-3.57722]
	-0.00073	-0.1257			
IG(I)(-1)	[-0.47595]	[ 6.86189]	IG( $\pi$ )(-1)	7.87E-05	0.479918
				-8.90E-05	-0.1264
	-4.76E-05	-0.223994		[ 0.88347]	[ 3.79695]
	-0.00073	-0.12604	.	.	.
IG(I)(-2)	[-0.06522]	[-1.77722]	.	.	.
			.	.	.
	0.616968	125.173	IG( $\pi$ )(-12)	-0.000121	-0.021854
	-0.30986	-53.4658		-8.40E-05	-0.11933
C	[ 1.99114]	[ 2.34118]		[-1.43603]	[-0.18313]
			C	0.060571	-78.64412
R-squared	0.818596	0.706016		-0.09254	-131.364
Adj. R-squared	0.807433	0.687924		[ 0.65456]	[-0.59867]
			R-squared	0.998914	0.940565
			Adj. R-squared	0.998359	0.910215

*Fuente: Base de datos en Eviews*

Las especificaciones de los modelos VAR no deberían sorprender en cuanto a un menor ajuste de los índices en comparación con las variables. A medida que la economía peruana esté más inmersa en el internet, la diferencia entre estas bondades de ajuste disminuiría.

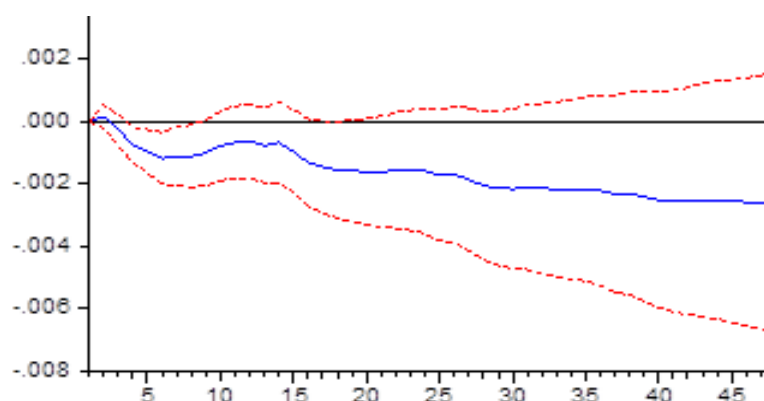
Por otro lado, al generar un impacto de un shock aleatorio (innovación) en alguna de las variables del sistema, la dinámica de los efectos de las perturbaciones muestran reacciones más marcadas en los índices que en las variables reales (Anexo 3).

La Inversión ante un impulso en el IG (I) tiene un ligero choque en los períodos iniciales que tiende a desaparecer a partir del sexto período; mientras que para la Inflación, un shock en el IG( $\pi$ ) tiene efectos más fuertes, significativos y altamente persistentes que se estabilizan a partir del periodo 27 aproximadamente (Figura 3 y 4).



**Figura 3:** *Función Impulso-Respuesta Inversión. Respuesta de ICE ante IG (I)*

*Fuente: Base de datos en Eviews. Elaboración propia*



**Figura 4:** *Función Impulso-Respuesta Inflación. Respuesta de IPCsA ante IG ( $\pi$ )*

*Fuente: Base de datos en Eviews. Elaboración propia*

De forma complementaria, se presentan los resultados de la descomposición de la varianza, en la que se obtuvo que el error de predicción de las variables reales en un periodo adelante, es en su totalidad debido a innovaciones en ellas mismas. Sin embargo, a medida que aumenta el horizonte temporal de predicción, la participación de los índices va creciendo. Se apreció que el IG ( $\pi$ ) tiene mayor participación sobre el error de predicción que el IG (I) (Tabla 11).

Por otro lado, el error de predicción de los índices, a diferencia de las variables reales, ya no es representado totalmente por la variable principal, sino que es compartido tanto por la variable real como por su índice desde el primer período de predicción (Anexo 4).

**Tabla 11:**

*Descomposición de Varianza de la predicción VAR*

		<i>Respuesta de ICE ante IG(I)</i>			
		Periodo	S.E.**	VA4*	IG4RN*
<b><i>Inversión</i></b> <b><i>(índice de confianza</i></b> <b><i>empresarial)</i></b>	1		0.0470	100	0
	2		0.0650	99.84	0.15
	3		0.0756	99.46	0.53
	4		0.0826	99.03	0.96
	5		0.0876	98.67	1.32
	6		0.0910	98.43	1.56
	7		0.0941	98.26	1.7
	8		0.0963	98.16	1.83
	9		0.0981	98.08	1.91
	10		0.0994	98.03	1.96
		<i>Respuesta de IPCsA ante IG(<math>\pi</math>)</i>			
		Periodo	S.E.**	VA6*	IGRN*
<b><i>Inflación</i></b> <b><i>(IPC sin Alimentos)</i></b>	1		0.0015	100	0
	2		0.0019	99.29	0.70
	3		0.0022	97.99	2.00
	4		0.0027	91.16	8.83
	5		0.0032	84.78	15.21
	6		0.0036	77.30	22.69
	7		0.0038	71.67	28.32
	8		0.0041	68.01	31.98
	9		0.0043	65.41	34.58
	10		0.0045	65.19	34.80

*Fuente: Base de datos en Eviews. Elaboración propia*

*Notas: \* \*\* Aproximación a dos y cuatro decimales*

Este comportamiento, para inversión, se asume porque los agentes inversionistas, al generar un impacto en las búsquedas producto de un clima de incertidumbre en la economía, ocasionan una respuesta débil en la inversión, lo que nos lleva a pensar que el internet aún no toma suficiente participación en las herramientas de toma de decisión de los inversionistas. Las bandas de desviación estándar son amplias, es así que el grado de predicción del IG (I) es menor a corto y largo plazo.

Para el caso de inflación, un cambio en la conducta del agente reflejado en un impulso del IG ( $\pi$ ) tiene un fuerte impacto en la variable debido a que el mecanismo de respuesta es más inmediato. Esta variable muestra un mayor grado de predicción a Corto Plazo que posiblemente se sustente porque las palabras clave tienen resultados parecidos que generan una tendencia de acumulación por la interrelación de enlaces vinculados a ellas.

Comúnmente los trabajos con análisis VAR coinciden en que su simplicidad limita la aplicabilidad práctica en el análisis de política. Por lo que los alcances deductivos en el análisis VAR no pueden ser generalizados. Sin embargo, los resultados no se alejan de la realidad en un contexto como el nuestro y responden con el objetivo de la investigación.

La literatura escasa sobre análisis VAR entre componentes económicos y consultas en internet dificulta el contraste de los resultados aquí obtenidos, a pesar de ello, autores como McCandless, G., Gabrielli, F. y Murphy T. (2001) y Aponte et.al. (2004) realizan trabajos relacionados con las variables estudiadas y muestran interrelaciones significativas a la par con nuestros hallazgos.

Finalmente, para establecer que los datos obtenidos en tiempo real (IG) tienen relevancia para la predicción de las variables macroeconómicas, se procedió a dividir la muestra y aislar una parte para la evaluación de capacidad predictiva (muestra de trabajo y muestra de validación). A partir de ello, en los estadísticos obtenidos, se observó que los modelos que incluyen los IG calculados presentan mejores cualidades predictivas en comparación con los modelos generales (Tabla 12 y 13).

En ambas especificaciones ARMA (inflación e inversión) se obtuvieron coeficientes significativos tanto para el modelo general como para el modelo aumentado. Con niveles de observaciones distintas, el modelo aumentado posee un mayor R<sup>2</sup> ajustado, aunque para la inflación, al incluir el índice, el cambio es muy pequeño en comparación con la inversión (0.02% vs. 2.23%). Así también, los estadísticos de predicción se reducen al incluir el índice tanto para el método de predicción dinámico como para el estático. En el caso de inversión esta

reducción es de 18,12% y 17,65% respectivamente, mientras que para la inflación, esta reducción fue de 11,085% y 9,075%.

Satisfactoriamente los valores de los estadísticos de predicción son cercanos a cero, esto indica que la predicción es cercana a sus valores históricos. Así también, a través de un análisis comparativo entre los métodos de predicción, se observó una mayor diferencia en los modelos para Inversión. En este caso el pasar del método dinámico a estático, reduce los estadísticos de predicción en promedio de 0.045641 a 0.0171345. Sin embargo, en inflación, el cambiar de método los incrementa ligeramente de 0.001356 a 0.001440. Es decir, el uso de valores para las predicciones posteriores, sea información histórica o valores proyectados, sí alteran los resultados, aunque no de forma significativa.

**Tabla 12:**

*Parámetros estimados y estadísticos de capacidad predictiva. Variable Endógena: Índice de Confianza Empresarial (ICE)*

<b>Variables Exógenas</b>	<b>Modelo General</b>	<b>Modelo Aumentado</b>
$AR(1)$	0.9042 (0.0535)	0.8861 (0.0605)
$C$	4.0248 (0.0653)	4.0580 (0.0529)
$IGI_t$		-0.0183 (0.0006)
<i>Observaciones</i>	65	65
$R^2_{ajustado}$	0.8165	0.8347
$RECM_{dinamico}$	0.058987	0.048938
$EAM_{dinamico}$	0.052399	0.042344
$RECM_{estatico}$	0.022518	0.018897
$EAM_{estatico}$	0.019027	0.015372

*Fuente: Base de datos en Eviews. Elaboración propia*

*Notas: Nivel de significancia menor a 5%. - Las variables se encuentran en logaritmos a excepción del Índice calculado*

**Tabla 13:**

*Parámetros estimados y estadísticos de capacidad predictiva. Variable Endógena: IPC sin Alimentos (IPC<sub>sA</sub>)*

<b>Variables Exógenas</b>	<b>Modelo General</b>	<b>Modelo Aumentado</b>
<i>AR</i> (1)	0.3809 (0.1752)	0.4934 (0.2014)
<i>AR</i> (2)	0.6094 (0.1744)	0.4973 (0.2005)
<i>MA</i> (1)	0.8465 (0.1249)	0.7880 (0.1524)
<i>C</i>	4.9789 (0.3817)	4.9666 (0.3631)
<i>IGI<sub>t</sub></i>		-0.011 (4.97E-05)
<i>Observaciones</i>	76	76
<i>R<sub>ajustado</sub><sup>2</sup></i>	0.9975	0.9977
<i>RECM<sub>dinamico</sub></i>	0.001529	0.001377
<i>EAM<sub>dinamico</sub></i>	0.001476	0.001334
<i>RECM<sub>estatico</sub></i>	0.001735	0.001521
<i>EAM<sub>estatico</sub></i>	0.001443	0.001359

*Fuente: Base de datos en Eviews. Elaboración propia*

*Notas: Nivel de significancia menor a 5%. - Las variables se encuentran en logaritmos a excepción del Índice calculado*

En el caso de la prueba de Diebold y Mariano, con un nivel de confianza de 98%, hay suficiente evidencia estadística para no rechazar que el modelo aumentado es mejor que el general en términos de predicción para la inversión, es decir que los errores del modelo general son mayores a los del aumentado. Sin embargo para la inflación, el nivel de confianza se reduce a un 89% lo que indica una mayor probabilidad (11%) de no rechazar que los modelos sean similares en términos de predicción esto nos lleva pensar que ambos modelos son buenos predictores (Tabla 14).

**Tabla 14:***Prueba Diebold y Mariano*

	<b>ICE</b>	<b>IPCSA</b>
<b>Observaciones</b>	06	06
<b>EstadísticoDM</b>	2.31	1.61
<b>Pvalue</b>	0.02	0.11

*Fuente: Base de datos en Eviews. Elaboración propia*

*Notas: Ho=modelos 1 y 2 son similares en términos de predicción*

*Ha=modelo 2 predice mejor que modelo 1*

Una plausible explicación al mejor ajuste del modelo aumentado puede deberse a que la variable incluida (IG) utiliza expectativas en su mecanismo generador, a pesar de no considerarse como variable explicativa. Es decir, su comportamiento se ajusta al camino evolutivo de la variable real siendo relevantes en la especificación del modelo de predicción ya que tienen la idea intuitiva de “mirar hacia adelante”.

Además, la mejora de la calidad de la predicción vista en la reducción de los estadísticos de predicción RECM y EAM para el caso del ICE, es porque el IG (I) (variable incluida) contiene elementos de búsqueda sobre eventos muy enlazados con el componente temporal de la Inversión; mientras que los elementos de búsqueda del IG ( $\pi$ ) se enlazan parcialmente al componente temporal del IPCSA debido a sus factores fluctuantes. Sin embargo al ser estos estadísticos de predicción muy cercanos a cero, se torna complicado reducirlos en mayor proporción, lo que indica una buena capacidad predictiva del modelo.

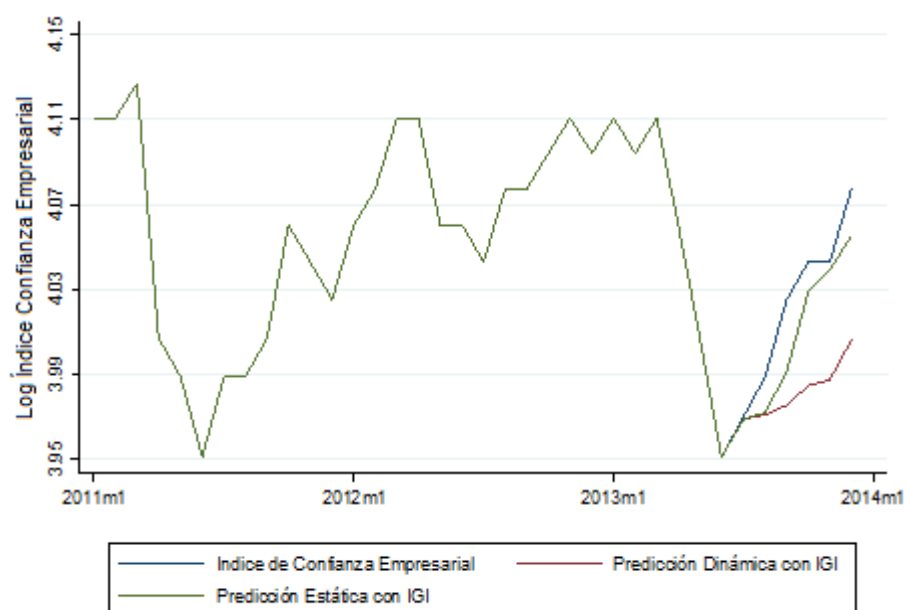
Ya que el error de predicción se acumula en el tiempo, evidentemente queda claro que éste es más volátil en el caso de las predicciones dinámicas que las de naturaleza estática. Porque en esta última, a cada paso se corrige el error que podría haberse presentado en la predicción anterior, utilizando los verdaderos valores de la variable endógena como insumo informativo para la predicción. Empero bajo determinadas características de la variable endógena (muy volátil) podría pasar el caso particular que las predicciones dinámicas sean más ajustadas que las estáticas como se evidencia para el caso de IPCSA en contraste al ICE que sí sigue el caso general.

A pesar que los resultados obtenidos cumplen los requisitos de diseño para una buena predicción, se debe tomar en cuenta que mientras se divide la muestra quedan pocas observaciones, lo que es una limitación para la aplicación de este modelo y puede ser causa de la reducción de los estadísticos de predicción así como de no poder anticipar la senda futura de

periodos prolongados. Así mismo un  $R^2$  alto puede llevar a pensar en problemas de multicolinealidad, pero no es necesariamente malo puesto que nuestro objetivo es la predicción y no estimar ecuaciones de comportamiento.

En la literatura económica se ha demostrado la efectividad del uso de los datos de Google Trends para mejorar predicciones en variables microeconómicas y macroeconómicas. Es así que al igual que este estudio, autores como Varian y Choi (2009b), Schmidt y Vosen (2009), Suhoy (2009), D'Amuri y Marcucci (2009), Bersier (2010), Chang y Del Río (2013) y Mejía et al. (2013) también concuerdan con la mayor relevancia de los modelos que incluyen variables conformadas por datos de Google Trends.

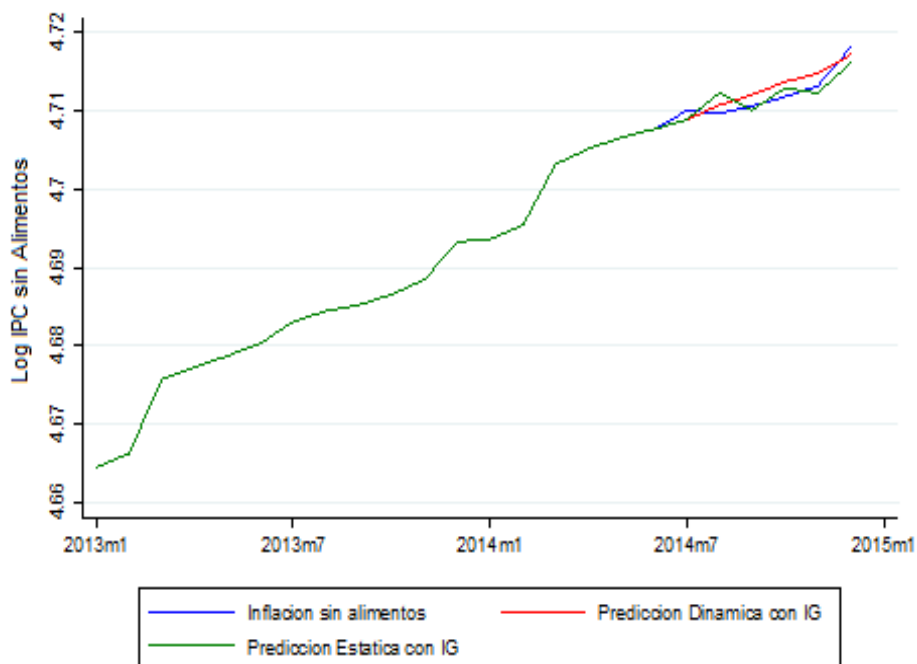
Por último se obtuvo un pronóstico que brinda mayores elementos de juicio a los agentes económicos sobre la evolución futura de las variables Inversión e Inflación.



**Figura 5:** Predicción para Inversión (ICE)

Fuente: Base de datos en Eviews. Elaboración propia

Notas: Variables tomadas en logaritmos



**Figura 6:** *Predicción para Inflación (IPC<sub>sA</sub>)*

*Fuente:* Base de datos en Eviews. *Elaboración propia*

*Notas:* Variables tomadas en logaritmos

Cabe mencionar que las especificaciones ARMA descritas al inicio de este capítulo y utilizadas para el pronóstico mostrado, podrían tener deficiencias puesto que estas especificaciones solo poseen componentes históricos (rezagos), sin considerar otras posibles variables explicativas. Además, una restricción del estudio es la frecuencia temporal de las series otorgadas por las fuentes de datos oficiales lo que impide una mejor predicción con mayor número de observaciones y componentes temporales de mayor frecuencia.

Sin embargo, al incluir los IG, las predicciones ponen en advertencia el sesgo que podrían presentar las predicciones oficiales por no considerar el uso masivo del internet ya que mediante el comportamiento espontáneo de los agentes se presentan interesantes propiedades de señalización.

## V. Conclusiones

Los modelos que incluyen los IG calculados presentan mejores cualidades predictivas en comparación con los modelos generales; en el caso de inversión los estadísticos de predicción se ven reducidos en mayor medida que en la Inflación. Entonces, los elementos de búsqueda de los IG se enlazan fuertemente al componente temporal de las series macroeconómicas. Además, analíticamente el modelo permitió encontrar nuevas relaciones y planteamientos económicos propios, así como interrogantes sobre la situación coyuntural que se produjo en el período de estudio.

Se determinó que los índices calculados IG por la metodología usada por Chang y Del Río (2013) son los de mayor incidencia en las variables de estudio por su mayor correlación y significancia estadística. Las palabras seleccionadas en la conformación de los IG mostraron su importancia conceptual en las percepciones de los agentes frente a las variables inversión e inflación. En ese sentido se obtuvo una relación inversa entre las variables y los IG.

Se comprobó que la inversión ante un impulso en el IG (I) tiene un ligero choque mientras que para la inflación, un shock en el IG ( $\pi$ ) tiene efectos más fuertes, significativos y altamente persistentes. Por otro lado, el grado de predicción del IG (I) es menor a Corto Plazo y el de IG ( $\pi$ ) es mayor. Además, la dinámica de los efectos de las perturbaciones tiene reacciones más marcadas en los índices que en las variables reales.

A través de un pronóstico de corto plazo se llega a mayores elementos de juicio para los agentes económicos sobre la evolución de las variables macroeconómicas y se ha demostrado la importancia de tomar en cuenta este tipo de información como insumo en el ámbito económico.

## **VI. Recomendaciones**

Se recomienda para próximas extensiones considerar un sistema de ponderación para diferenciar los motivos de búsqueda y poder extraer la parte vinculada netamente a las expectativas guiadas por preocupaciones económicas.

De igual forma se recomienda la incorporación de nuevas palabras clave seleccionadas por metodologías como análisis factorial u otras, sumadas a su carácter de correlación con las series macroeconómicas de fuentes oficiales.

Del mismo modo, sería interesante realizar un análisis más específico dentro de cada sector económico, para investigar donde se evidencian más los efectos de búsquedas en internet, en relación con movimientos en la actividad económica.

## VII. Referencias Bibliográficas

- Aponte, E.; Duque, H.; Arroyo, J. y Flores, J. (2004). *Inflación y Desempleo: ejercicio econométrico para Cali-colombia*. Colombia.
- Askatas, N. y Zimmermann, K. (2009). *Google Econometrics and Unemployment Forecasting, Applied Economics Quarterly*. 55(2). Pp. 107-120, Alemania.
- Bersier, F. (2010). *Towards Better Policy and Practice using Real-Time Data*. Inglaterra: Oxford Internet Institute.
- Carriere-Swallow, Y. y Labbé, F. (2010). *Nowcasting with Google Trends in an Emerging Market*. Banco Central de Chile.
- Castro, J. y Rivas-Llosa, R. (2005). *Econometría Aplicada*. (2da ed.). Lima: Centro de Investigación de la Universidad del Pacífico.
- Chang, J., y Del Río, A. (2013). *Google Trends: Predicción del nivel de empleo agregado en Perú usando datos en tiempo real, 2005-2011*. Lima: BCRP.
- Choi, H. y Varian, H. (2009a). *Predicting the Present with Google Trends*. Google Inc.
- Choi, H. y Varian, H. (2009b). *Predicting Initial Claims for Unemployment Benefits*. Google Inc.
- Choi, H. y Varian, H. (2011). *Predicting the Present with Google Trends*. Google Inc.
- Della Penna, N. y Huang, H. (2009). *Constructing a Consumer Confidence Index for the US Using Web Search Volume*. Working Paper 2009-26. Canadá : Universidad de Alberta.
- D'Amuri, F. y Marcucci, J. (2009). *Google it! Forecasting the US unemployment rate with a Google job search index*. Working paper 2009-32. Inglaterra: Institute for social and Economic Research.
- Dzienlinski, M. (2010) *Measuring Economic Uncertainty and its Impact on the Stock Market*. NCCR FINRISK.
- Fosk, A. (2011). *Estado de Internet con un Enfoque en el Perú*. comScore Inc.
- Figueroa, B. (s.f.). *Midiendo Incertidumbre Política Económica con Google y su Impacto en el Mercado Accionario: Evidencia Chilena*. Pontificia Universidad Católica de Chile.
- Ginsberg, J., Mohebbi, M., Patel, R., Brammer, L., Smolinski, M. y Brilliant, L. (2009). *Detecting influenza epidemics using search query data*. 457. Febrero 2009.
- Gujarati, D. y Porter, D. (2010). *Econometría*. McGraw Hill.
- Guzmán, G. (2011). *Internet Search Behavior as an Economic Forecasting Tool: The Case of Inflation Expectations*.
- Johnston, J. (2011). *Métodos de Econometría*. Barcelona: Vicens-Vives.
- Kato, E. y Suárez, M. (2007). *Búsquedas en Google, Incertidumbre e Inversión en México*. Universidad Autónoma de Querétaro.
- McCandless, G., Gabrielli, F., y Murphy, T. (2001). *Modelos Económicos de predicción macroeconómica en la Argentina*. Banco Central de la República Argentina.

- Mejía, L., Moncalve, D., Parra, Y., Pulido, S. y Reyes, A. (2013). *Indicadores ISAAC: Siguiendo la Actividad Sectorial a partir de Google Trends*. Colombia.
- Redondo, J. (2013). *Uso de Google Trends para predecir el Nivel y la Estructura del Desempleo en España*. Universidad Politécnica de Valencia: España.
- Schmidt, T. y Vosen, S. (2009). *Forecasting Private Consumption: Survey-based Indicators vs. Google Trends*. RUHR Economic Papers: Germany.
- Song, H., Pan, B. y Yang, Y. (2013). *Predicting Hotel Demand Using Destination Marketing Organizations' Web Traffic Data*. Journal of Travel Research
- Suhoy, T. (2009). *Query Indices and a 2008 Downturn*. Jerusalem: Bank of Israel.
- Trujillo, G. (2010). Cap.11 Econometría Dinámica en *Econometría con Eviews* (pp.91-109). Perú: Universidad Nacional de Cajamarca.

## VIII. Anexos

### Anexo 1: Test de Causalidad en el sentido de Granger

#### Tabla:

*Test de Causalidad de IPC sin alimentos – IG ( $\pi$ )*

Null Hypothesis:	Obs	F-Statistic	Prob.
IPCSA does not Granger Cause IG_PI	72	3.76351	0.0005
IG_PI does not Granger Cause IPCSA		2.65215	0.0085

#### Tabla:

*Test de Causalidad de Formación de Capital – IGI*

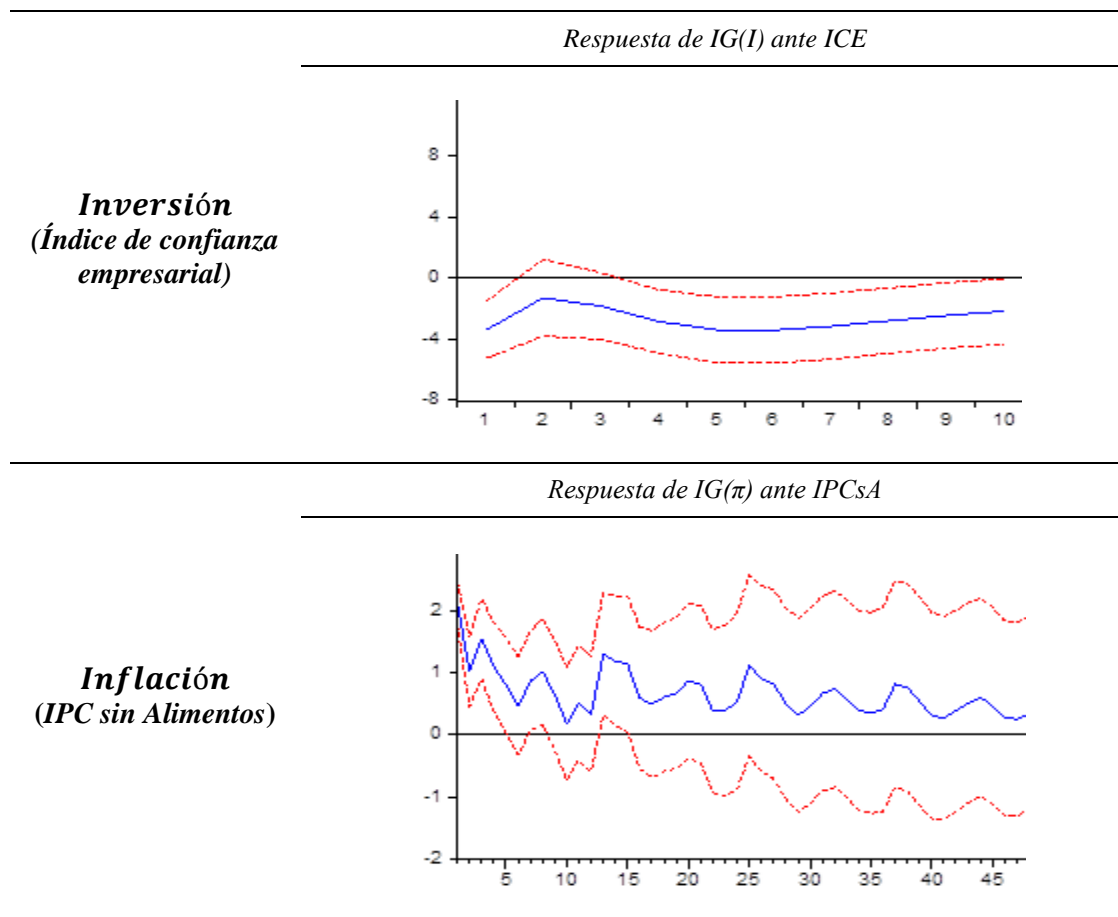
Null Hypothesis:	Obs	F-Statistic	Prob.
ICE does not Granger Cause IGI	70	5.36658	0.0070
IGI does not Granger Cause ICE		0.26142	0.7708

### Anexo 2: Test de Heterocedasticidad

#### Tabla:

*Test de Heterocedasticidad de Formación de Capital – IGI*

<i>Metodología de Johansen</i>	
<b>Variables: Formación de Capital – IGI</b>	
<b>Estadístico Traza</b>	
<i>No existe vector de cointegracion</i>	0.0119
<i>Hasta 1 vector de cointegracion</i>	0.0308
<b>Estadístico Valor Propio máximo</b>	
<i>No existe vector de cointegracion</i>	0.0376
<i>Hasta 1 vector de cointegracion</i>	0.0308
<i>¿Cointegran?</i>	SI
<i>Notas: Nivel de significancia al 0.05 haciendo uso de los p-values Mackinnon-Haug-Michelis (1999)</i>	

**Anexo 3: Función Impulso-Respuesta****Tabla:***Función Impulso-Respuesta*

**Anexo 4:** Descomposición de Varianza de la predicción VAR**Tabla:***Descomposición de Varianza de la predicción*

		<i>Respuesta de IG(I) ante ICE</i>			
		Period	S.E.**	VA4*	IG4RN*
<b><i>Inversión</i></b> <i>(Índice de confianza empresarial)</i>		1	8.1183	17.57	82.42
		2	10.3901	12.24	87.75
		3	11.1913	13.17	86.82
		4	11.6818	17.95	82.04
		5	12.1855	24.14	75.85
		6	12.6594	29.56	70.43
		7	13.0492	33.60	66.39
		8	13.3505	36.45	63.54
		9	13.5803	38.48	61.51
		10	13.7568	39.95	60.04
		<i>Respuesta de IG(<math>\pi</math>) ante IPCsA</i>			
		Period	S.E.**	VA6*	IGRN*
<b><i>Inflación</i></b> <i>(IPC sin Alimentos)</i>		1	2.1744	4.81	95.18
		2	2.4047	4.24	95.75
		3	2.8847	4.36	95.63
		4	3.1573	8.13	91.86
		5	3.2723	8.46	91.53
		6	3.3050	8.29	91.70
		7	3.4633	10.19	89.80
		8	3.6228	9.95	90.04
		9	3.7361	12.25	87.74
		10	3.7423	12.32	87.67
<b>Notas:</b> * ** Aproximación a dos y cuatro decimales					