

Un Indicador de Tendencia Artificial para el Consumo Privado con Google Trends

Juan Tenorio

Heidi Alpiste

Jakelin Remón

Arian Segil

MEF

XLII Encuentro de Economistas del BCRP

Las opiniones vertidas en este documento son de entera responsabilidad de los autores y no reflejan necesariamente las opiniones del MEF o las del BCRP

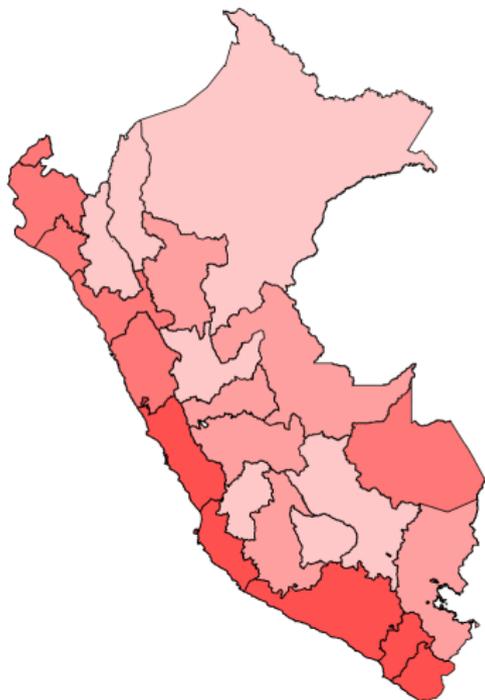
Octubre 2024

Contenido

- 1 Motivación**
 - Google Trends
 - Consumo privado
- 2 Revisión de literatura**
 - Google Trends, consumo privado y otras variables
 - Google Trends y la generación de nuevos indicadores
- 3 Base de datos de Google Trends**
 - Extracción
 - Selección
 - Categorización
- 4 Metodología**
 - Benchmark models
 - Redes neuronales artificiales
 - Redes neuronales recurrentes
 - Evaluación de proyección
- 5 Resultados**
 - Indicadores obtenidos para el consumo privado
 - Pruebas y calibración de los parámetros
 - Análisis de robustez

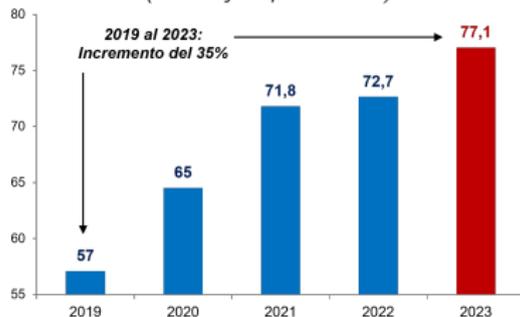
Motivación: ¿Por qué utilizar Google Trends?

Población de 6 años a más que usa internet según ámbito geográfico en 2022 (Porcentaje respecto al total)



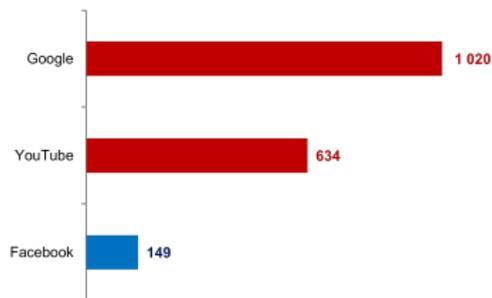
Fuente: INEI, elaboración propia.

Población de 6 años a más que usa internet (Porcentaje respecto al total)



Fuente: INEI, elaboración propia.

Websites más visitadas en Perú hasta Ago-24 (Millones de visitas)

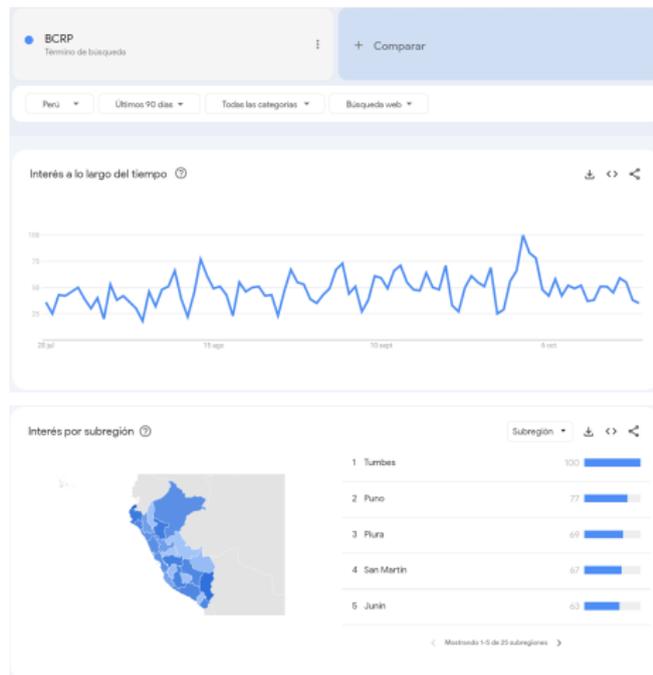


Fuente: Semrush, elaboración propia.

Motivación: ¿Cómo lo hace Google Trends?

- Google Trends permite medir la popularidad de un término a lo largo del tiempo, con un índice entre 0 y 100.
- La construcción de los índices se puede resumir en los siguientes pasos:
 - 1 Se divide el número de búsquedas del término entre el total en una región y/o período seleccionado.
 - 2 Los datos se indexan en una escala de 0 a 100, dividiendo cada valor por el mayor resultado y se multiplica por 100.

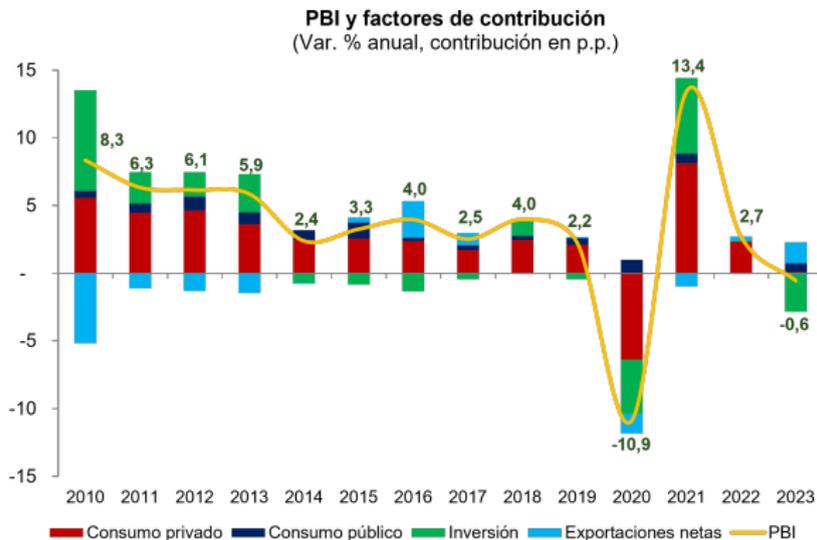
Término de Búsqueda "BCRP" en Google Trends



Fuente: Google Trends.

Motivación: ¿Por qué elegir el consumo privado?

- Entre 2010 y 2023, el PBI registró un crecimiento promedio de 3,6%, donde el consumo privado ha contribuido con 2,6 p.p. Cabe destacar que, en términos generales, el consumo privado representa el 64% del PBI, consolidando su importancia en la actividad económica.



Fuente: BCRP, elaboración propia.

Contenido

- 1 Motivación**
 - Google Trends
 - Consumo privado
- 2 Revisión de literatura**
 - Google Trends, consumo privado y otras variables
 - Google Trends y la generación de nuevos indicadores
- 3 Base de datos de Google Trends**
 - Extracción
 - Selección
 - Categorización
- 4 Metodología**
 - Benchmark models
 - Redes neuronales artificiales
 - Redes neuronales recurrentes
 - Evaluación de proyección
- 5 Resultados**
 - Indicadores obtenidos para el consumo privado
 - Pruebas y calibración de los parámetros
 - Análisis de robustez

Revisión de literatura I

Google Trends y el consumo privado:

- Vosen y Schmidt (2011) elaboran un indicador de consumo privado en US basado en Google Trends, comparando su poder predictivo con indicadores tradicionales (MCSI y CCI).
- Woo y Owen (2019) incorporan datos de Google Trends y noticias para aumentar la precisión de las proyecciones de consumo privado en US.
- Song y Coupé (2023) aprovechan los datos de Baidu para predecir el consumo privado en China, mejorando los pronósticos tanto sectoriales como agregados.

Google Trends y otras variables económicas:

- Swallow y Labbé (2010) desarrollan un indicador de compra de automóviles en Chile basado en datos de Google, mejorando la predicción de ventas.
- Chang y Del Río (2013) usan Google Trends para crear un índice que anticipa el empleo agregado en Perú.
- Lolic, Matosec y Soric (2024) construyen un índice para medir las ventas minoristas en US utilizando Google Trends.

Revisión de literatura II

Google Trends con otras variables para generar nuevos indicadores:

- Blanco (2014) emplea Google Trends junto a indicadores tradicionales para anticipar la venta de autos y construir un índice de consumo en Argentina.
- Zahedi y Rounaghi (2015) integran Google Trends y redes neuronales para predecir los precios bursátiles en Irán, mejorando las estimaciones.
- Monge *et al.* (2024) combinan técnicas de análisis de texto y análisis factorial con datos de Google Trends para desarrollar un indicador que pronostique con mayor precisión el PBI de China.
- Tenorio y Pérez (2024) utilizan Google Trends junto con otros indicadores adelantados para proyectar el PBI de Perú, logrando una mayor precisión predictiva.

Contenido

- 1 **Motivación**
 - Google Trends
 - Consumo privado
- 2 **Revisión de literatura**
 - Google Trends, consumo privado y otras variables
 - Google Trends y la generación de nuevos indicadores
- 3 **Base de datos de Google Trends**
 - Extracción
 - Selección
 - Categorización
- 4 **Metodología**
 - Benchmark models
 - Redes neuronales artificiales
 - Redes neuronales recurrentes
 - Evaluación de proyección
- 5 **Resultados**
 - Indicadores obtenidos para el consumo privado
 - Pruebas y calibración de los parámetros
 - Análisis de robustez

Base de datos: extracción de datos

- El recojo de información empezó con la selección de 130 términos de búsqueda, identificados a partir de la revisión bibliográfica y determinantes del consumo.

Términos de búsqueda utilizados para la creación de los indicadores



Fuente: Elaboración propia.

Base de datos: selección de términos

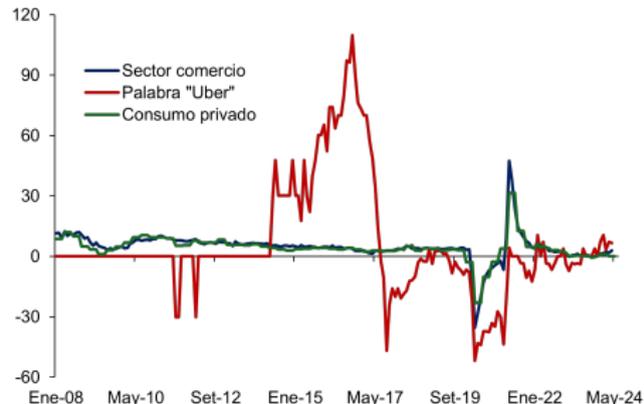
- Entre las desventajas de utilizar términos de búsqueda se encuentra que algunas palabras pueden anticipar el comportamiento del consumo privado y/o del comercio y servicios, mientras que otras no lo hacen.

Búsqueda de "spa" vs. Consumo privado vs. Índice del sector comercio ✓
(Var. % anual)



Fuente: Google Trends, INEI, elaboración propia.

Búsqueda de "Uber" vs. Consumo privado vs. Índice del sector comercio ✗
(Var. % anual)



Base de datos: categorías de búsqueda

- Se establecieron categorías en función a la finalidad del consumo privado.

Consumo privado por finalidad y categorías del ITAConsumo

Consumo privado por finalidad del INEI	Categoría del ITAConsumo	Término
Alimentos y bebidas no alcohólicas / Bebidas alcohólicas, tabaco y estupefacientes	Alimentos	Pizza Hut, restaurantes
Restaurantes y hoteles	Turismo	Despegar.com, Cusco, Avianca, SKY, Marriot, terminal terrestre, paquetes turísticos, hoteles
Salud	Cuidado personal	Smart Fit, Soho Color, spa
Transporte	Transporte	Toyota, Hyundai, Mazda, Indriver, autos, alquiler autos
Recreación y cultura	Recreación	Cinemark, discotecas, cines, teatro

Fuente: INEI, elaboración propia.

Contenido

- 1 Motivación**
 - Google Trends
 - Consumo privado
- 2 Revisión de literatura**
 - Google Trends, consumo privado y otras variables
 - Google Trends y la generación de nuevos indicadores
- 3 Base de datos de Google Trends**
 - Extracción
 - Selección
 - Categorización
- 4 Metodología**
 - Benchmark models
 - Redes neuronales artificiales
 - Redes neuronales recurrentes
 - Evaluación de proyección
- 5 Resultados**
 - Indicadores obtenidos para el consumo privado
 - Pruebas y calibración de los parámetros
 - Análisis de robustez

Metodología

- **Modelo benchmark:** técnica de reducción de dimensionalidad, donde se transforman los datos en un conjunto no correlacionado, llamado componente principal, que explica la mayor parte de la varianza de los datos observados:

$$Z = XW \quad (1)$$

donde X son los datos, W la matriz de autovectores, y Z la matriz de PC.

- **Modelo de factores dinámicos:** se estima un modelo DFM canónico siguiendo a Banbura, Marta y Modugno Michele (2014). Este modelo se puede describir de la siguiente manera:

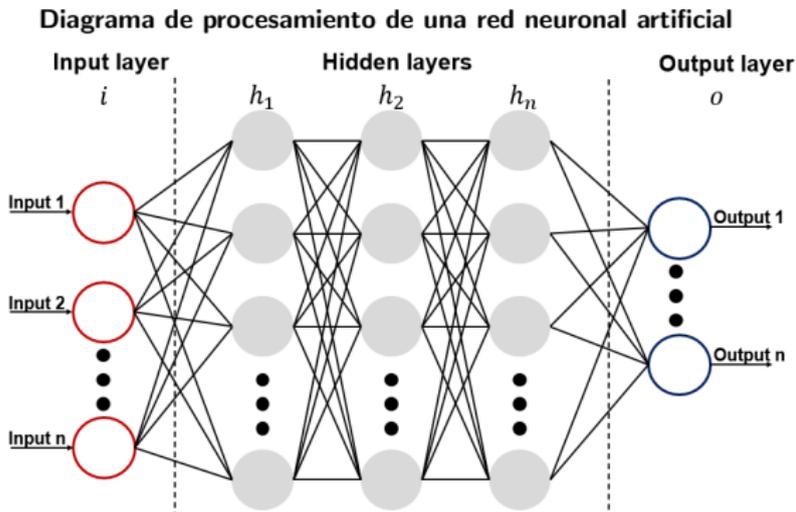
$$x_t = C_0 f_t + e_t \quad e_t \sim N(0, R) \quad (2)$$

$$f_t = \sum_{j=1}^p A_j f_{t-j} + u_t \quad u_t \sim N(0, Q_0) \quad (3)$$

donde la ecuación 2 es de medida y la 3 de transición, permitiendo al factor no observable f_t comportarse como un VAR.

Redes neuronales artificiales

- Una RNA consiste en múltiples capas de neuronas conectadas entre sí, donde cada neurona aplica una función de activación a una combinación lineal de las entradas recibidas.

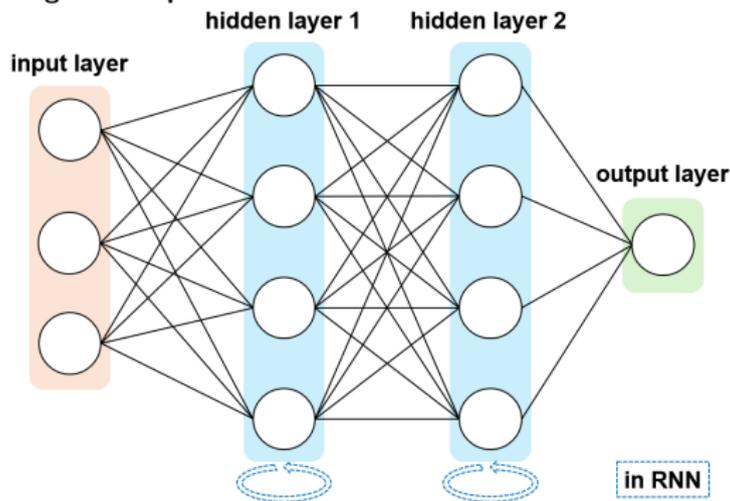


Fuente: Elaboración propia

Redes neuronales recurrentes

- Las RNN cuentan con conexiones recurrentes que permiten que la salida de una neurona se retroalimente como entrada para la misma neurona en pasos temporales subsiguientes. Esta estructura permite a las RNN "recordar" información previa en la secuencia de datos.

Diagrama de procesamiento de una red neuronal recurrente



Fuente: Elaboración propia

Evaluación de proyección

- Disponemos finalmente de un conjunto total de **130 términos que abarcan desde enero de 2008 hasta octubre de 2024**. La evaluación y selección de los términos óptimos se llevarán a cabo de manera independiente para la creación del indicador.
- **Estrategia de evaluación de proyecciones:** El método que evaluará la precisión en la proyección de cada modelo se realizará por medio del MSE y el RMSE, este último con la ecuación:

$$RMSE = \sqrt{\frac{1}{T} \sum_{t=1}^T (y_t - \hat{y}_t)^2} \quad (4)$$

- Posterior a esta evaluación, emplearemos el método de Diebold y Mariano (1995) y Giacomini y White (2006) para determinar si la incorporación de categorías de Google mejora la proyección del consumo privado.

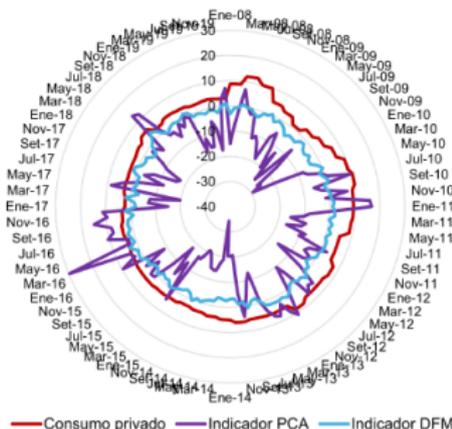
Contenido

- 1 Motivación**
 - Google Trends
 - Consumo privado
- 2 Revisión de literatura**
 - Google Trends, consumo privado y otras variables
 - Google Trends y la generación de nuevos indicadores
- 3 Base de datos de Google Trends**
 - Extracción
 - Selección
 - Categorización
- 4 Metodología**
 - Benchmark models
 - Redes neuronales artificiales
 - Redes neuronales recurrentes
 - Evaluación de proyección
- 5 Resultados**
 - Indicadores obtenidos para el consumo privado
 - Pruebas y calibración de los parámetros
 - Análisis de robustez

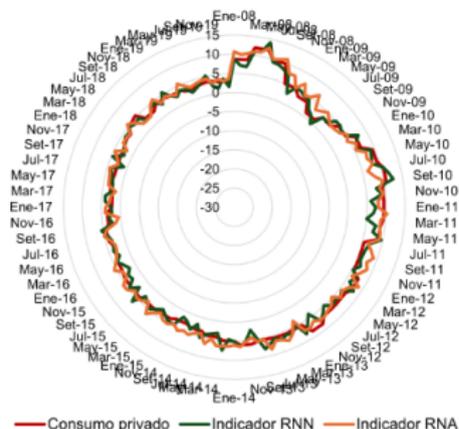
Resultados

- Se obtienen los indicadores y, posteriormente, se aplica la Validación Cruzada (CV) para optimizar los parámetros de cada modelo. Se observa que las redes neuronales logran capturar un patrón consistente que se manifiesta de manera robusta a lo largo de los diferentes conjuntos de datos del consumo privado.

Consumo privado vs Indicadores por metodología



(a) Consumo Privado vs Benchmark

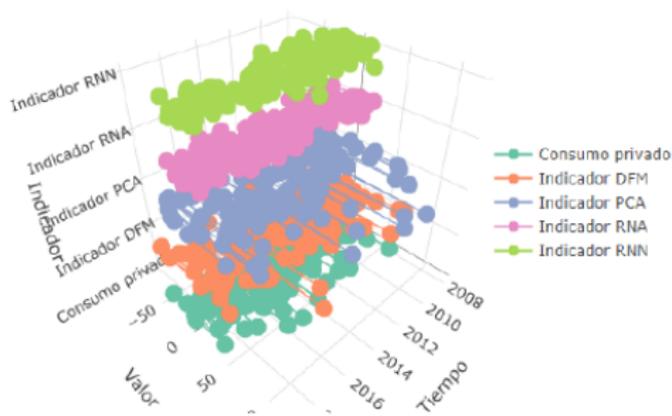


(b) Consumo Privado vs Indicadores de Redes

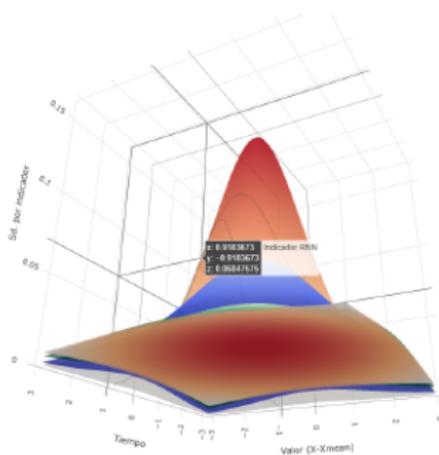
Fuente: Elaboración propia.

Análisis de volatilidad

Análisis de la volatilidad por cada indicador



(a) Varianza condicional $CVAR(\epsilon)$



(b) Varianza $Var(\epsilon)$

Fuente: Elaboración propia.

Pruebas y calibración de los parámetros

- Se divide el conjunto de datos en **tres partes clave: entrenamiento, validación y prueba**. Inicialmente, el modelo se entrena utilizando los datos del conjunto de entrenamiento (*in-sample*), los cuales minimicen el RMSE.
- Para determinar el conjunto de parámetros óptimos por método, la muestra se dividirá en tres segmentos: i) entrenamiento (Ene-2008 a Ago-2014) con 5 ventanas (folds), ii) Validación (Set-2009 a May-2022), y iii) Testeo (Jun-2022 a Oct-2024)

Training dataset					Validation set		Testing set	
2008m1-2014m08					2014m09-2022m5		2022m06-2024m10	
↔					↔		↔	
Fold 1	Fold 2	Fold 3	Fold 4	Fold 5				

Fuente: Elaboración propia.

- Este enfoque permitirá optimizar cada parámetro, garantizando que cualquier metodología no solo se ajuste adecuadamente a los datos de entrenamiento, sino que también generalice una memoria a largo plazo con nuevos datos y prevenga el sobreajuste. Esto mejorará la estabilidad y la precisión de las proyecciones.

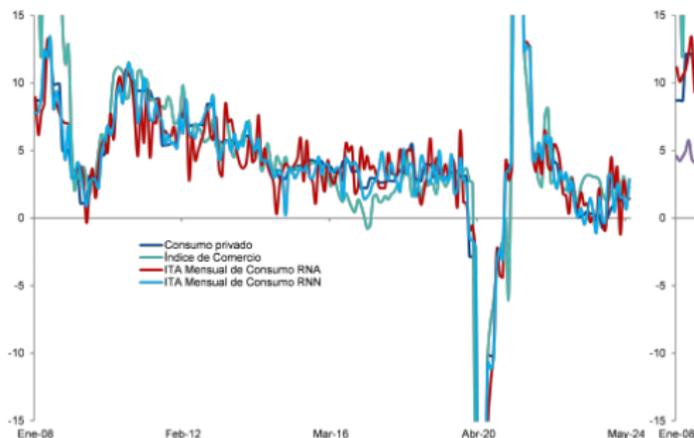
Priors y rango de los hiperparámetros

Modelo	Hiperparámetro	Rango	Valor optimizado
PCA	Número de componentes	2 a 12	6
DFM	Número de factores	2 a 10	4
	Longitud de la serie	0.01 to 2.5	1.2
RNA	Número de capas ocultas	2 to 64	32
	Número de neuronas	6 to 256	64
RNN	Número de capas ocultas	2 to 48	24
	Número de neuronas	6 to 256	32
	Tamaño del lote (Batch Size)	2.5 to 6.5	2.1

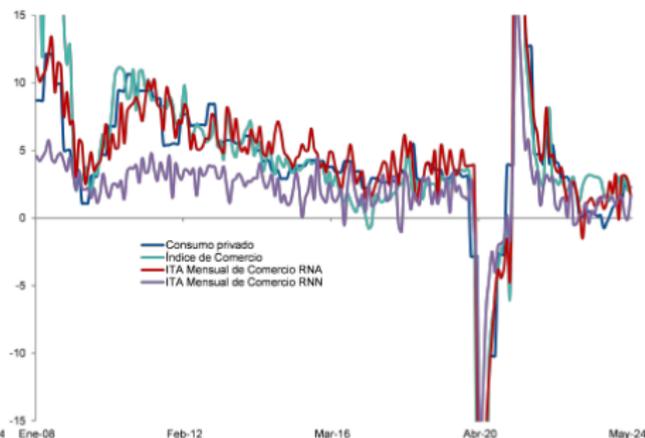
Fuente: Elaboración propia.

ITAC mensual de consumo privado

Consumo privado vs Indicadores mensuales de redes neuronales optimizados



(a) ITAC mensual con hiperparámetros óptimos

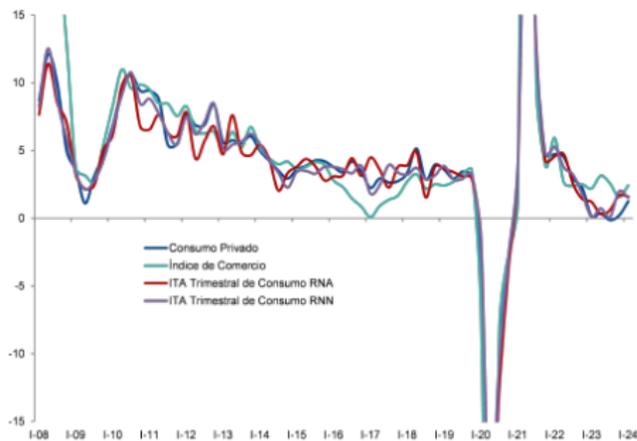


(b) ITAC mensual con priors

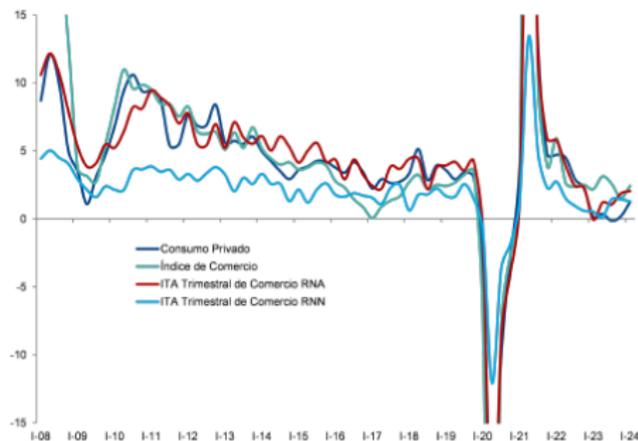
Fuente: Elaboración propia.

ITAC trimestral de consumo privado

Consumo privado vs Indicadores trimestrales de redes neuronales optimizados



(c) ITAC trimestral con hiperparámetros óptimos



(d) ITAC trimestral con priors

Fuente: Elaboración propia.

Análisis de robustez

- Una ecuación con indicadores líderes del consumo privado (*Stepwise Least Squares*) que determine el componente comercio y servicios.
- Una ecuación de estado-espacio que de forma recurrente se va agregando cada categoría del ITAC en un orden generado por BSTS de Scott y Varian (2013).

$$\begin{aligned} \text{ComercioySS}_t &= \beta_1 \text{Empleo}_t + \beta_2 \text{CreditConst}_t + \beta_3 \text{CreditHip} + \beta_4 \text{IPC} + \epsilon_t \\ \text{Consumo}_t &= \beta_1 X_t + \beta_2 \text{VarGoogle}_{it} + e_t \end{aligned} \quad (5)$$

Evaluación de los estadísticos de predicción

Modelos	Estimado	MSE	RMSE	p-value (DM)	p-value (GW)
Alimentos	0.614	2.35	1.40	0.029	0.002
Transporte	0.241	3.61	6.45	0.057	0.017
Turismo	0.795	2.02	0.98	0.005	0.032
Recreación	0.713	4.51	5.62	0.026	0.014
Cuidado personal	0.727	3.56	6.61	0.049	0.059
Total	0.791	1.69	1.12	0.001	0.002

Fuente: Elaboración propia.

Conclusiones

- El uso de términos de búsqueda en Google puede mejorar los indicadores de alta frecuencia al detectar patrones complejos en el consumo privado mediante redes neuronales recurrentes (RNN). **La combinación de variables de sentimiento y RNN permite capturar dinámicas no lineales.**
- **Los modelos RNN permiten hacer predicciones con datos incompletos**, lo que es esencial en contextos de incertidumbre y fragmentación de información.
- **Adaptar el uso de Google Trends, según categorías específicas de consumo, mejora la capacidad predictiva** en comparación con enfoques homogéneos.
- Los datos de Google Trends pueden anticipar eventos económicos importantes, **ofreciendo a los responsables de políticas una herramienta clave para medir el consumo y la actividad económica.**
- **La integración de Google Trends en modelos predictivos debe ajustarse según el objetivo del análisis**, ya sea para proyecciones en tiempo real o para prever tendencias futuras de consumo privado.

Agenda pendiente

- Además, una ventaja de este tipo de indicadores es que también nos pueden brindar información adelantada sobre el incremento de la criminalidad y sentimiento de inseguridad en el país, como el indicador de inseguridad ciudadana, construido a través de Google Trends.

Participación de palabras buscadas en Google Trends en el indicador (Porcentaje)

Palabra	Pesos
Extorsión	40
Estafas	21
Homicidio	14
Violación	10
Cámaras de seguridad	9
Policía	5
Comisaría	1

Indicador mensual de inseguridad ciudadana¹ (Unidad estandarizada)



¹ La correlación lineal (Pearson) y la no lineal (Spearman), se mide sin considerar el año 2020 y 2021.
Fuente: Google Trends. SIDPOL-Ministerio del Interior. estimaciones MEF.

Un Indicador de Tendencia Artificial para el Consumo Privado con Google Trends

Juan Tenorio

Heidi Alpiste

Jakelin Remón

Arian Segil

MEF

XLII Encuentro de Economistas del BCRP

Las opiniones vertidas en este documento son de entera responsabilidad de los autores y no reflejan necesariamente las opiniones del MEF o las del BCRP

Octubre 2024