



Seminario Anual de Cuentas Nacionales en América Latina y el Caribe:
Hacia el SCN 2025

EL MACHINE LEARNING EN LAS CUENTAS
NACIONALES. UN INDICADOR OPORTUNO DEL
CONSUMO PRIVADO PARA MÉXICO



Dirección General de Estadísticas Económicas
Dirección General Adjunta de Cuentas Nacionales
Dirección de Cuentas de Corto Plazo y Regionales

Septiembre 2024

MOTIVACIÓN

Pandemia por SARS-CoV-2 (COVID-19) en 2020

Necesidad o Pretensión

Predicción o Pronóstico

Las estimaciones se realizan con **información oportuna** de la alta frecuencia y que requiere de **modelos**



Las estimaciones realizadas **anticipan a las cifras oficiales**

- ❑ Estimación Oportuna del PIB Trimestral
- ❑ Indicador Mensual Oportuno de la Actividad Manufacturera
- ❑ Indicador Oportuno de la Actividad Económica (mensual)
- ❑ **Indicador Oportuno del Consumo Privado (Mensual)**

ANTECEDENTES

En 2023, el Consumo Privado en México, representa el 51.2% del PIB.

El Indicador Mensual del Consumo Privado se publica nueve semanas después de concluido el mes de referencia.

Interés de los usuarios especializados en un indicador más oportuno del consumo privado y de la inversión.



OBJETIVO

Ofrecer una visión anticipada sobre la evolución del gasto de bienes y servicios del consumo privado, para **dos meses posteriores** al dato del Indicador Mensual del Consumo Privado publicado.

Mediante el uso de:

- Información estadística disponible.
- Técnicas de *machine learning*.
- Modelos econométricos.



PROCESO DE PRODUCCIÓN DE LA INFORMACIÓN

Se divide en dos procedimientos o fases:

DGA de Cuentas Nacionales

Estimación del mes 1 (T+1):

- ◆ Se cuenta con información estadística adecuada.
- ◆ Estimación más precisa.
- ◆ Combinación de diferentes modelos de *machine learning*.

DGA de Investigación

Estimación del mes 2 (T+2):

- ◆ Se cuenta con información estadística mínima.
- ◆ Estimación menos precisa.
- ◆ Combinación de modelos de Factores Dinámicos y de regresión con errores ARMA*.

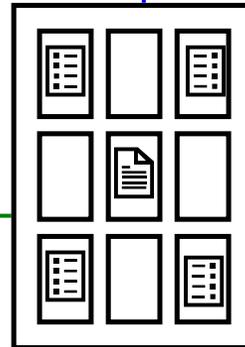
*Errores ARMA: errores autorregresivos y medias móviles.

FUENTES DE INFORMACIÓN PARA T+1

Tradicional
(mensuales)



No tradicional
(mensuales)



Tradicional
(diarias)

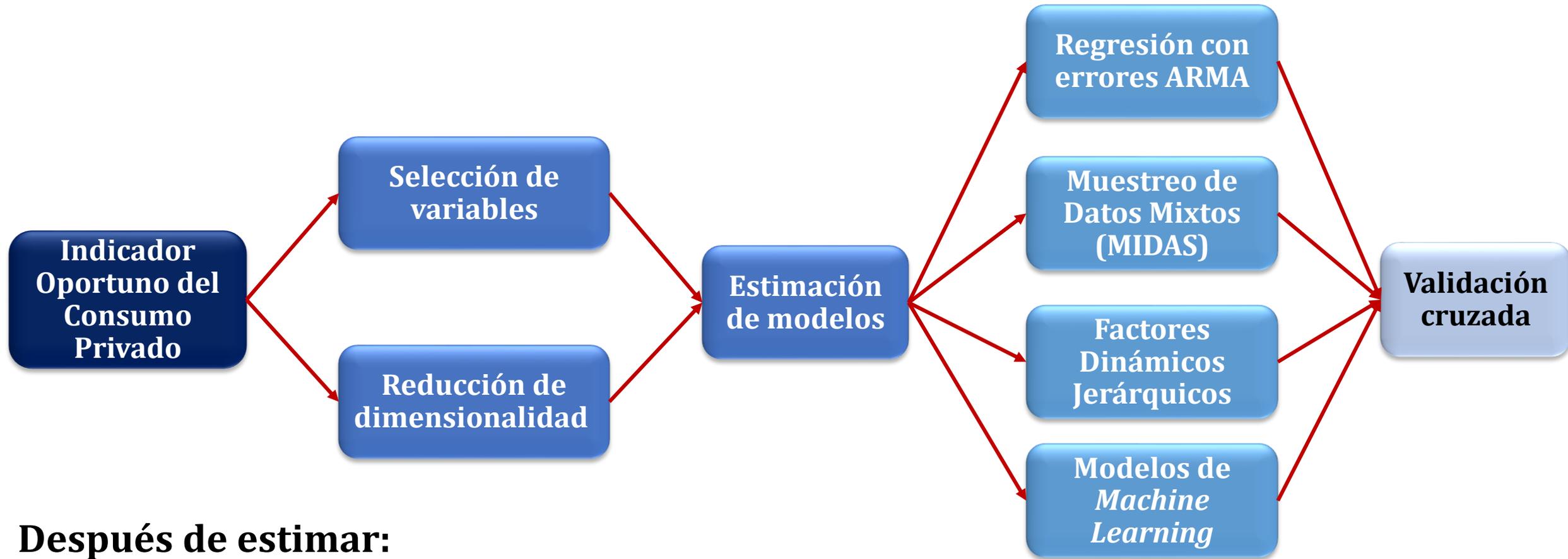


INFORMACIÓN SELECCIONADA PARA T+1

Criterio de selección:

- √ Que tengan una **mayor oportunidad que la usada en el IMCP.**
- √ Que estén altamente **correlacionadas y causalmente vinculadas** con el IMCP, para evaluar su capacidad predictiva.
- √ Son provenientes de fuentes tradicionales, en **frecuencia mensual y diaria.**
- √ Todas las variables se utilizan en **cifras desestacionalizadas.**

METODOLOGÍA DEL PROCEDIMIENTO PARA T+1



Después de estimar:

1. Se obtienen **ocho pronósticos** para el Consumo privado total.
2. Se obtiene el **promedio o mediana** de los ocho modelos.
3. Se obtienen los **intervalos de confianza** al 95%.

SELECCIÓN DE VARIABLES CON MACHINE LEARNING

En el análisis de grandes conjuntos de datos, las técnicas de *machine learning* se emplean para seleccionar solo las variables más relevantes, para evitar sobreajuste.

RIDGE

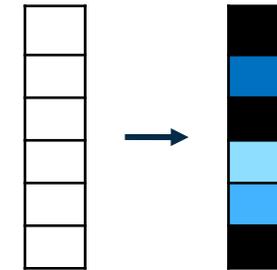
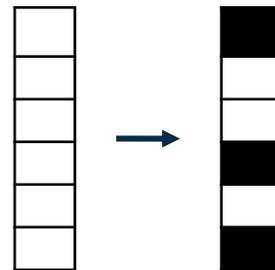
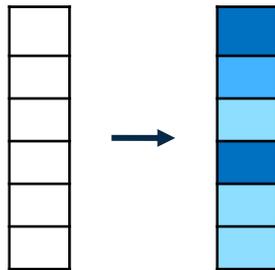
- Se enfoca en la regularización, penalizando la suma de los cuadrados de los coeficientes de regresión.
- Conservar todas las variables asignando pesos diferentes según su importancia.

LASSO

- Aplica una penalización sobre la suma de los valores absolutos de los coeficientes de regresión.
- Permite eliminar variables con poca importancia.

ELASTIC NET

- Supera algunas de las limitaciones de Lasso y, al mismo tiempo, aprovecha las bondades de Ridge, dando como resultado una combinación de ambas regresiones.
- Permite eliminar algunas variables, pero conservar otras, asignando diferentes pesos según su importancia.



REDUCCIÓN DE DIMENSIONALIDAD CON MACHINE LEARNING

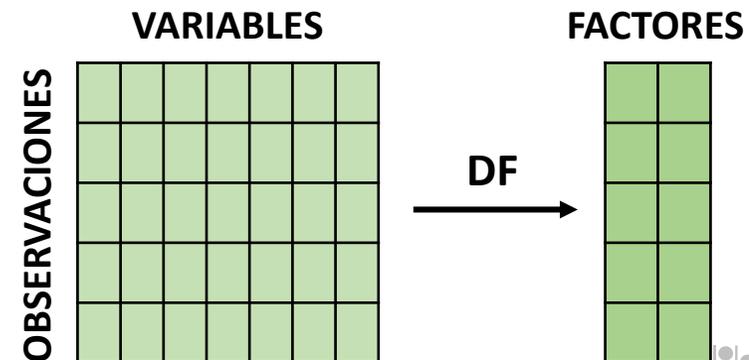
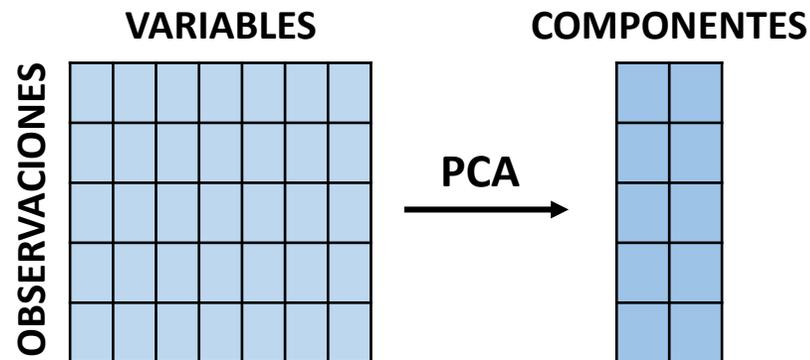
Busca simplificar los datos reduciendo el número de variables en dos o tres componentes o factores combinados, mientras se conserva la mayor cantidad posible de información relevante.

COMPONENTES PRINCIPALES (PCA)

- Transforma un conjunto de variables correlacionadas en un conjunto más pequeño de variables no correlacionadas, capturando la mayor cantidad de varianza posible en los datos.

FACTORES DINÁMICOS (DF)

- Captura las dinámicas temporales y las interdependencias en los datos de series temporales, reduciendo la dimensionalidad.
- Ayudan a descubrir tendencias a largo plazo entre las variables.



MODELOS UTILIZADOS

1

Regresión con errores ARMA

Combina la regresión lineal con modelos ARMA para capturar la estructura de los términos de error.

2

Muestreo de Datos Mixtos

Permite estimar ecuaciones con variables de diferentes frecuencias en un mismo modelo sin tener que homogeneizarlas.

3

Factores Dinámicos Jerárquicos

Descompone las series temporales en factores que capturan la variabilidad a distintos niveles jerárquicos, para comprender las interacciones dentro y entre estos niveles.

4

Modelos de *Machine Learning*

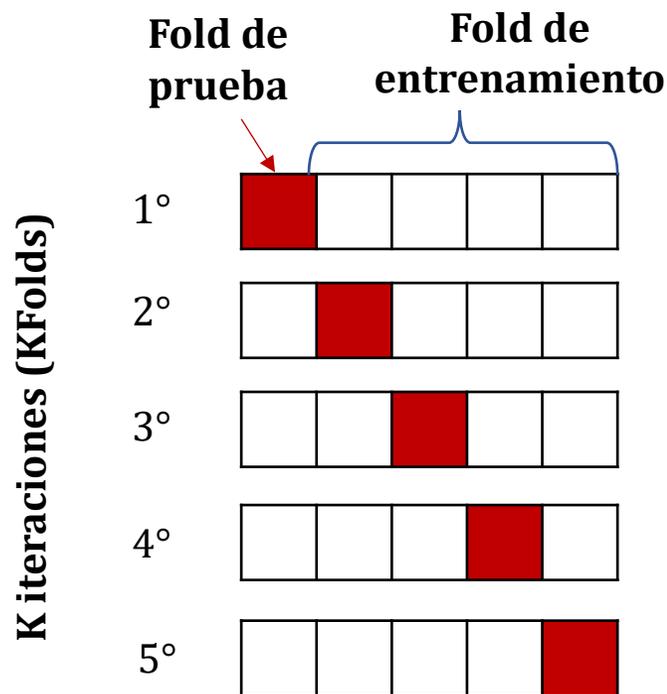
- **Ridge.** Mantiene todas las variables en el modelo.
- **Lasso.** Eliminar algunas variables de poca importancia.
- **Elastic Net.** Permite eliminar algunas variables y conserva otras asignando diferentes pesos según su importancia.

VALIDACIÓN CRUZADA EN MACHINE LEARNING

En lugar de usar diferentes conjuntos de datos para entrenamiento y validación

¡En la validación cruzada usaremos todos los datos!

Asegura que un modelo de predicción funcione bien.

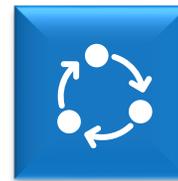


KFold Cross Validation (CV)



División

- Divide los datos en k subconjuntos diferentes.



Entrenamiento y Prueba

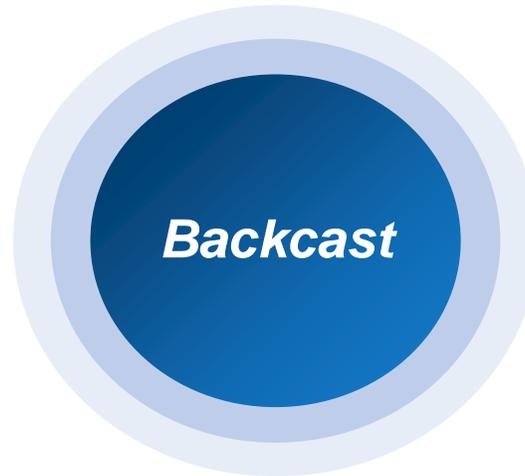
- Entrena el modelo en k-1 fold.
- Evalúa el modelo en el fold restante.
- Repite el proceso k veces.

El desempeño final del modelo será simplemente el **promedio** de los desempeños.

RESULTADOS PUBLICADOS

La **primera** publicación del **Indicador Oportuno del Consumo Privado** ofreció estimaciones correspondientes a los meses de **octubre y noviembre de 2022**.

Adelanta en seis semanas

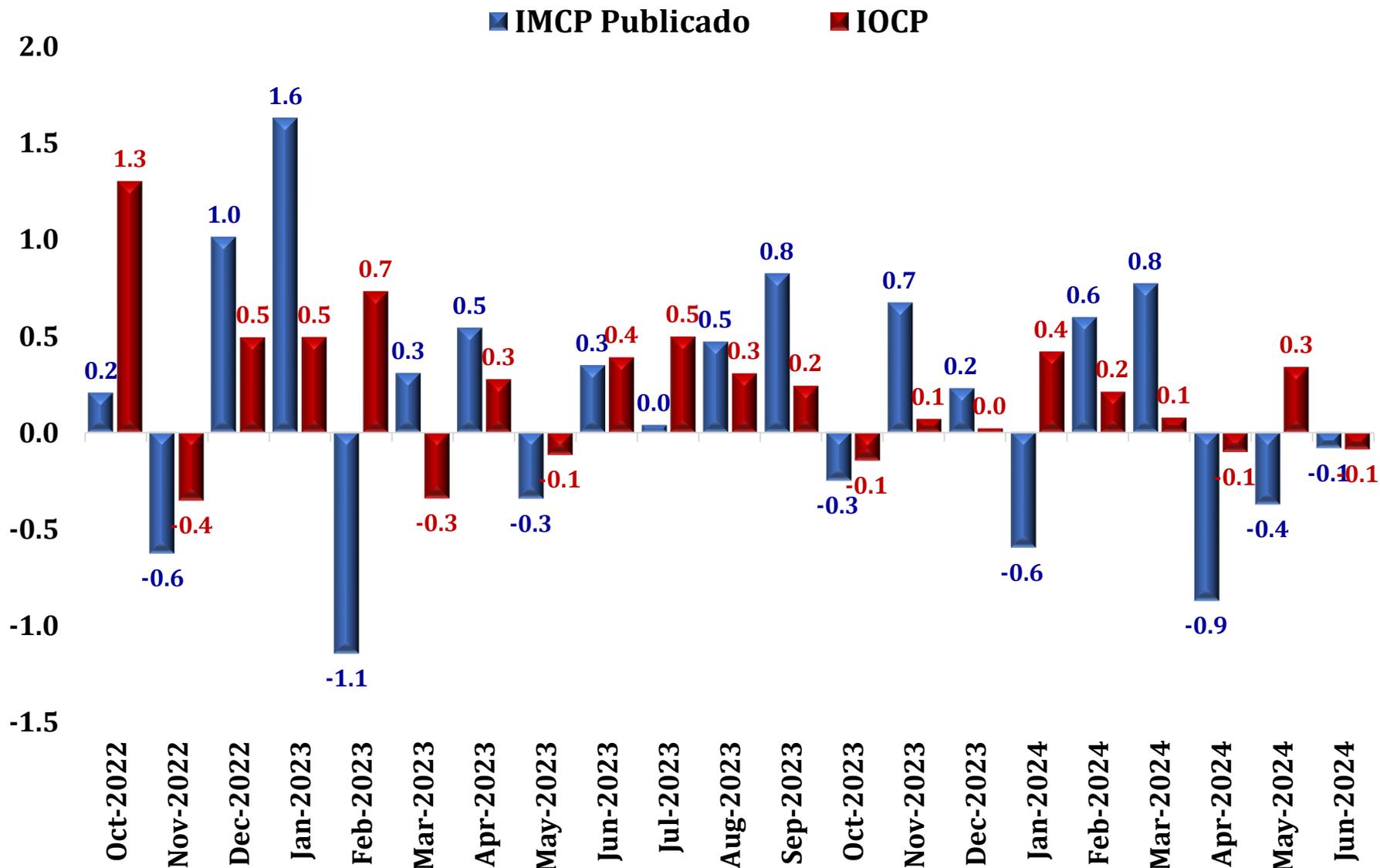


15 de diciembre de 2022

Antes de esta publicación, se realizaron **ejercicios de estimación** dentro de la muestra, abarcando el periodo de **enero de 2019 a septiembre de 2022**.

Comparativo entre el IOCP y el IMCP

(Variaciones mensuales en cifras desestacionalizadas)

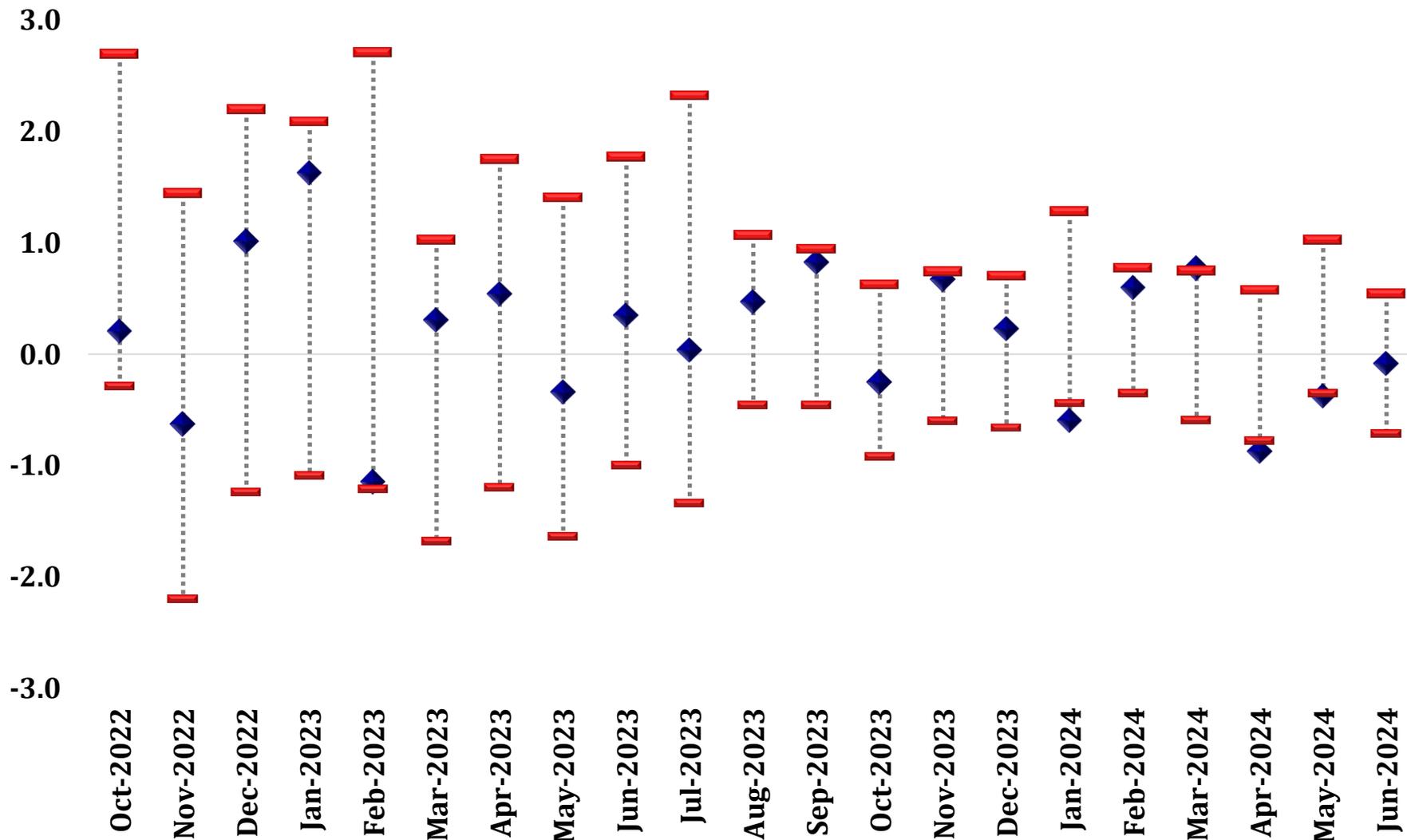


El promedio de las diferencias de los 21 meses publicados es de 0.22

Intervalos de Confianza del IOCP vs IMCP

(Variaciones mensuales en cifras desestacionalizadas)

◆ IMCP Publicado — Límite inferior — Límite superior



Los intervalos de confianza al 95% capturan 81% al valor observado, es decir, 17 de los 21 meses estimados.

METODOLOGÍA DEL PROCEDIMIENTO PARA T+2

Misma metodología del Indicador Oportuno de la Actividad Económica (IOAE).

Se siguen los siguientes pasos:

1. **Considera el IOCP para T+1** como observado del Consumo privado.
2. **Desestacionalización** de las series de tiempo: X13ARIMA-SEATS.
3. **Transformación óptima de covariables:** Máxima correlación respecto al Consumo privado.
4. **Selección de tópicos de Google Trends:** Mínimos Cuadrados Parciales.
5. **Estimación de Modelos de factores dinámicos:** Componentes principales más Suavizamiento de Kalman
6. **Regresión con errores ARMA(p, q) con validación cruzada:** Selección de parámetros p y q con menor error con datos fuera de muestra.
7. **Selección de modelos:** Pruebas de Diebold-Mariano.
8. **Combinación de modelos:** Mediana de modelos con igual error de estimación.

Nota: Se estiman modelos en niveles (no estacionarios), variaciones % mensuales y anuales; así mismo se considera la especificación que genera menor error de estimación.

DIFUSIÓN



Calendario de difusión de información estadística y geográfica y de Interés Nacional del INEGI 2024

	IIN	Tipo de cifras	Periodicidad ¹	Ene.	Feb.	Mar.	Abr.	May.	Jun.	Jul.	Ago.	Sep.	Oct.	Nov.	Dic.
UNIDAD DE MEDIDA Y ACTUALIZACIÓN															
Unidad de Medida y Actualización (UMA), 2024. ³	No	NA	Anual	9											

INFORMACIÓN EXPERIMENTAL		
Bienestar Autorreportado (BIARE).		
Indicadores de Bienestar Autorreportado de la Población Urbana. ³	No	D
Cubo de Datos Geospaciales de México.		
Índice de Clasificaciones de Agua Superficial desde el Espacio (ICASE) Landsat, 2023.	No	
Índice de Vegetación de Diferencia Normalizada (NDVI), 1984-2023.	No	
Geomediana Landsat, 2023.	No	
Indicador Oportuno de la Actividad Económica (IOAE). ³	No	C
Indicador Oportuno del Consumo Privado (IOCP). ³	No	C

INEGI English Otros idiomas Contacto +A

Temas Programas de información Sistemas de Consulta Infraestructura Acerca del INEGI

Buscar... Enviar

Inicio / Investigación / Estadísticas experimentales

Investigación

Datos resultantes de modelos estadísticos que no constituyen estadísticas oficiales del INEGI y que tienen como propósito ampliar la oferta de información.

Estadísticas experimentales

Esta sección recoge proyectos en desarrollo que cuentan con aspectos innovadores, ya sea en las fuentes de información, métodos estadísticos y el ámbito de estudio o la forma de difundir los resultados.

- Estadísticas experimentales
- Modelos experimentales
- Análisis geoespacial experimental
- Registro de datos primarios
- Grupos de investigación

 [Índice Nacional de Competitividad](#)

 [Indicador Mensual Oportuno de la Actividad Manufacturera](#)

 [Indicador Oportuno de la Actividad Económica](#)

 [Indicador Oportuno del Consumo Privado](#)

CONCLUSIONES



Estimación del IOCP **para dos meses posteriores** al último dato publicado del IMCP.



Establecer **diferentes metodologías** para T+1 y T+2 potencializa el uso de la información oportuna disponible.



Las técnicas de *machine learning* facilitan el **procesamiento y la extracción de información de grandes conjuntos de datos**, superando las limitaciones de los métodos tradicionales.



Se emplean **modelos de *machine learning*** para robustecer la precisión de los pronósticos sin dejar de lado los modelos estadísticos tradicionales.



La **validación cruzada** evalúa la precisión de los modelos de *machine learning*, proporcionando una estimación más confiable.



La **combinación de modelos** contribuye a reducir los errores de estimación.

ARTÍCULO DE INVESTIGACIÓN

Revista Internacional de Estadística y Geografía

Nowcasting del Consumo Privado en México

https://rde.inegi.org.mx/wp-content/uploads/2023/pdf/RDE41/RDE41_art04.pdf



GRACIAS

Conociendo
México

800 111 46 34
www.inegi.org.mx
atencion.usuarios@inegi.org.mx

    **INEGI Informa**