

# Indicador Oportuno del **Consumo Privado**

Síntesis metodológica



Instituto Nacional de Estadística y Geografía

## **Indicador Oportuno del Consumo Privado**

**Síntesis metodológica**



**Obras complementarias publicadas por el INEGI sobre el tema:**

Indicador Mensual del Consumo Privado en el Mercado Interior (IMCPMI). Indicador Mensual de la Actividad Industrial (IMAI). Indicador Global de la Actividad Económica (IGAE). Encuesta Mensual sobre Empresas Comerciales (EMEC). Encuesta Mensual de Opinión Empresarial (EMOE). Encuesta Nacional sobre Confianza del Consumidor (ENCO). Encuesta Nacional de Ocupación y Empleo (Nueva Edición) (ENOEN). Registro Administrativo de la Industria Automotriz de Vehículos Ligeros (RAIAVL). Registro Administrativo de la Industria Automotriz de Vehículos Ligeros (RAIAVL). Vehículos de Motor Registrados en Circulación (VMRC) Venta al público de automóviles.

**Catalogación en la fuente INEGI:**

339.4201072 Instituto Nacional de Estadística y Geografía (México).  
Indicador Oportuno del Consumo Privado : síntesis metodológica / Instituto Nacional de Estadística y Geografía.-- México : INEGI, c2022.

vii, 11 p.

1. México - Condiciones económicas. 2. México - Calidad de vida.

**Conociendo México**

**800 111 4634**

**[www.inegi.org.mx](http://www.inegi.org.mx)**

**[atencion.usuarios@inegi.org.mx](mailto:atencion.usuarios@inegi.org.mx)**



**INEGI Informa**



**@INEGI\_INFORMA**

Registro en trámite

2022, **Instituto Nacional de Estadística y Geografía**

Edificio Sede

Avenida Héroe de Nacozari Sur 2301

Fraccionamiento Jardines del Parque, 20276 Aguascalientes,

Aguascalientes, Aguascalientes, entre la calle INEGI,

Avenida del Lago y Avenida Paseo de las Garzas.

## Presentación

---

El **Instituto Nacional de Estadística y Geografía (INEGI)** presenta la publicación del Indicador Oportuno del Consumo Privado, el cual proporciona estimaciones oportunas del Indicador Mensual del Consumo Privado del Mercado Interior (IMCPMI) obtenidas a través de modelos econométricos y de *nowcasting*.

## Antecedentes

---

En México, como en otros países, la pandemia y las acciones empleadas para contenerla han provocado un impacto en la actividad económica, frenando la producción de algunos sectores económicos y modificando los patrones de consumo de los hogares. Actualmente, el Indicador Mensual del Consumo Privado en el Mercado Interior (IMCPMI) se publica alrededor de nueve semanas después del mes de referencia. Por tanto, un indicador que gane oportunidad es de gran importancia para la toma de decisiones en materia de política económica, dado que el consumo privado es la variable más relevante entre los componentes de la demanda agregada. En consecuencia, la Dirección de Cuentas de Corto Plazo y Regionales y la Dirección General Adjunta de Investigación realizaron un importante esfuerzo para construir un indicador que permita estimar el consumo de los hogares mexicanos de manera precisa y oportuna.

Dicho esfuerzo concluyó en la estimación del Indicador Oportuno del Consumo Privado (IOCP) mediante un conjunto de modelos econométricos y de aprendizaje máquina (*machine learning*). El IOCP proporciona información de los dos meses posteriores con respecto al último dato disponible del IMCPMI. De esta manera, las estimaciones se presentan 6 semanas después del cierre del mes de referencia para  $T + 1$  y 2 semanas para  $T + 2$ . Con esto, se adelantan tres semanas de oportunidad para  $T + 1$  y siete semanas para  $T + 2$ , a la publicación de los datos oficiales del IMCPMI.

La publicación consiste en una estimación puntual de la variación anual, variación mensual y los niveles del índice (base 2013=100) en cifras desestacionalizadas, junto con sus intervalos de confianza al 95 por ciento. Cabe aclarar que esta estimación no pretende sustituir al cálculo tradicional.

Para obtener la estimación del primer mes con respecto al último dato oficial publicado, se emplea un conjunto de metodologías con el fin de potencializar toda la información actualizada para realizar las estimaciones oportunas o *nowcasts*. Para el segundo mes se utiliza la metodología desarrollada por Corona et al. (2022), quienes utilizan Modelos de Factores Dinámicos (MFD) para estimar oportunamente al Indicador Global de la Actividad Económica. En este caso se adecúa dicha metodología para estimar oportunamente al IMCPMI.

El resto de este documento se divide en dos secciones, la primera para explicar el procedimiento de *nowcasting* para obtener la estimación del primer mes con respecto al último dato oficial publicado, aquí llamada  $T + 1$ , y la segunda sección para el segundo mes, denominada  $T + 2$ .

## Índice

---

<b>1. IOCP: obteniendo estimaciones para <math>T + 1</math></b>	<b>1</b>
1.1 Datos	1
1.2 Métodos de estimación	1
1.2.1 Regresión lineal con penalización	1
1.2.2 Factores dinámicos	2
1.2.3 Modelo de factores dinámicos jerárquicos	3
1.2.4 Regresión MIDAS	4
<b>2. IOCP: obteniendo estimaciones para <math>T + 2</math></b>	<b>5</b>
2.1 Fuentes de información	5
2.2 Transformaciones previas	5
2.3 PLS como método de selección de variables	5
2.4 Modelo de factores dinámicos	5
2.5 Modelo de <i>nowcasting</i>	6
2.6 Consideraciones metodológicas	6
<b>Referencias</b>	<b>7</b>
<b>Anexos</b>	<b>9</b>
A. Variables utilizadas para la estimación del IOCP	10
B. Sistemas de pago de bajo valor de Banxico	11

# 1. IOCP: obteniendo estimaciones para $T + 1$

---

## 1.1 Datos

Las variables utilizadas son seleccionadas a partir de su relación con el IMCPMI con base en la teoría económica y de manera estadística según el criterio de oportunidad, correlación y causalidad en sentido de Granger. Se incorporaron fuentes tradicionales de información tanto de frecuencia mensual como diaria (ver Cuadro 1). Todas las variables se introducen al modelo en cifras originales.

## 1.2 Métodos de estimación

Para generar la estimación del IOCP se aplicaron distintos métodos de estimación: tres modelos de regresión con errores ARMA (uno con un factor dinámico de dos etapas con selección *Lasso*, otro con un componente principal con selección *Lasso*, y uno más con un factor dinámico con selección por mínimos cuadrados parciales); dos modelos de regresión lineal con penalización (*Elastic Net* y *Ridge*); un modelo de factores dinámicos jerárquicos; y por último, dos modelos de regresión de muestreo de datos mixtos con variables de alta frecuencia (uno a partir del monto total de tarjetas de crédito y débito, y otro mediante un factor estático que selecciona entre las 30 variables que componen el monto total de tarjetas de crédito y débito). Dependiendo del método de estimación, se realizan las transformaciones pertinentes de acuerdo con los requerimientos de cada metodología.

### 1.2.1 Regresión lineal con penalización

Las estrategias de penalización (o regularización) son adecuadas cuando se tienen muchas variables predictoras. Los métodos de regresión lineal con regularización incorporan penalizaciones en los coeficientes obtenidos por mínimos cuadrados ordinarios, con el objetivo de evitar sobreajuste, reducir varianza y minimizar la influencia de las variables menos relevantes.

La regresión por mínimos cuadrados estima los parámetros  $\beta_0, \beta_1, \dots, \beta_N$ , donde  $N$  indica el número de variables independientes, que minimizan la suma del cuadrado de los errores:

$$SSE = \sum_{i=1}^n \left( y_i - \beta_0 - \sum_{j=1}^N \beta_j x_{ij} \right)^2, \quad (1)$$

con  $n$  el número de observaciones.

**Regresión Ridge.** La regresión Ridge (Hoerl, 1970) añade una penalización sobre la suma de los parámetros de la regresión al cuadrado. Los coeficientes de la regresión Ridge  $\hat{\beta}^R$  son aquellos que minimizan:

$$\sum_{i=1}^n (y_i - \hat{y}_i)^2 - \lambda \sum_{j=1}^N \beta_j^2, \quad (2)$$

donde  $\lambda \geq 0$  es un parámetro de ajuste que se determina de forma separada. El término de penalización  $\lambda \sum \beta_j^2$  es pequeño cuando los coeficientes son cercanos a cero, por lo que podría decirse que contrae los coeficientes que se obtienen por mínimos cuadrados. Cuando  $\lambda = 0$ , el término de penalización no tiene efecto y la regresión Ridge se convierte en una estimación por mínimos cuadrados. Sin embargo, si  $\lambda$  tiende a infinito, el impacto de la penalización crece y los crecientes estimados se aproximan a cero. A diferencia de los mínimos cuadrados que solo generan una única estimación, la regresión Ridge produce una estimación  $\hat{\beta}_\lambda^R$  para cada  $\lambda$ , por lo que la selección de un buen valor de  $\lambda$  es crucial. La regresión Ridge contrae los parámetros estimados y los hace más cercanos a cero, pero el modelo no hace que ningunos de ellos sea exactamente cero en ningún valor de  $\lambda$ .

**Regresión Lasso.** Una alternativa a la regresión Ridge es Lasso (*Least absolute shrinkage and selection operator*), introducido por Tibshirani (1996). Los coeficientes del modelo Lasso  $\hat{\beta}_\lambda^L$  son aquellos que minimizan:

$$\sum_{i=1}^n (y_i - \hat{y}_i)^2 - \lambda \sum_{j=1}^N |\beta_j|, \quad (3)$$

en donde el término de penalización es lo único que difiere con la regresión *Ridge*. Se dice que *Lasso* utiliza una penalización  $L_1$  y *Ridge* una penalización  $L_2$ .

La penalización  $L_1$  tiene el efecto de igualar a cero algunos de los coeficientes estimados para algunos valores de  $\lambda$ . Por lo que, este modelo aplica simultáneamente regularización y selección de variables. De igual forma que en la regresión *Ridge*, la elección de un valor óptimo de  $\lambda$  es crucial para la obtención de un modelo adecuado. Al conseguir que algunos coeficientes sean exactamente cero, *Lasso* tiene una ventaja ante *Ridge* en escenarios donde no todos los predictores son importantes para el modelo y cuando se desea excluir el efecto de las variables menos influyentes.

**Regresión *Elastic Net*.** Para superar algunas de las limitaciones de *Lasso* y al mismo tiempo aprovechar las bondades de *Ridge*, se desarrolló el modelo *Elastic Net* (ENET) (Lever *et al.*; 2016; Zou y Hastie, 2005), el cual resulta de la combinación de las penalizaciones  $L_1$  y  $L_2$ . En este caso, los coeficientes  $\hat{\beta}_\lambda^{Enet}$  son aquellos que minimizan:

$$\sum_{i=1}^n (y_i - \hat{y}_i)^2 - \lambda_1 \sum_{j=1}^N |\beta_j| - \lambda_2 \sum_{j=1}^N \beta_j^2. \quad (4)$$

Si se sustituye  $\alpha = \frac{\lambda_1}{\lambda_1 + \lambda_2}$  y  $\lambda = \lambda_1 + \lambda_2$ , se obtiene la expresión equivalente que es más utilizada para fines computacionales,

$$\sum_{i=1}^n (y_i - \hat{y}_i)^2 - \lambda (\alpha \sum_{j=1}^N |\beta_j| - (1 - \alpha) \sum_{j=1}^N \beta_j^2). \quad (5)$$

El grado en que influye cada una de las penalizaciones está controlada por el parámetro  $\alpha$  que está comprendido en el intervalo  $[0,1]$ . Cuando  $\alpha = 0$ , es una regresión *Ridge* y cuando  $\alpha = 1$  es una regresión *Lasso*.

Si se mantienen fijos los parámetros  $\alpha$  y  $\lambda$ , la estimación de  $\beta$  para las regresiones *Ridge*, *Lasso* y *Elastic Net* se puede realizar mediante el método de máxima verosimilitud penalizada a través de técnicas de optimización convexa (Friedman *et al.* 2009).

## 1.2.2 Factores dinámicos

El Modelo de Factores Dinámicos expresa al vector de  $X_t$  de variables observadas, como un conjunto de factores no observados o latentes y un componente idiosincrático de media cero, en donde tanto los factores latentes y el término idiosincrático están generalmente correlacionados serialmente (Stock y Watson, 2016).

Sea  $X_t = (x_{1,t}, x_{2,t}, \dots, x_{N,t})$ , con  $0 \leq i \leq n$ , el conjunto de  $N$  variables predictoras observadas, que satisfacen el supuesto de estacionariedad débil. La especificación general del modelo de factores dinámicos es:

$$X_t = \Lambda f_t + \epsilon_t, \quad (6)$$

$$f_t = \sum_{i=1}^p A_i f_{t-i} + B u_t, \quad u_t \sim i.i.d. N(0, I_q). \quad (7)$$

A  $\Lambda$  se le conoce como matriz de cargas y a su  $i$ -ésimo elemento se le conoce como la carga de la serie  $X_{it}$ . La ecuación (6) impone una estructura de proceso VAR(p) para los factores  $f_t$ . Tanto  $\epsilon_t$  como  $u_t$  son normales y se asume que no están correlacionados para ningún rezago, es decir,  $E[\epsilon_t u_{t-k}] = 0$ , para todo  $k$ . En general, el componente idiosincrático puede estar correlacionado serialmente, por lo que el modelo descrito por (6) y (7) está incompletamente especificado. Para que esto no sea así es deseable especificar un modelo paramétrico para  $\epsilon_t$ .

**Factores dinámicos de dos etapas.** En el primer paso, los estimadores preliminares de los factores y los estimadores de los parámetros del modelo se calculan a partir de un análisis de componentes principales. El análisis de componentes principales (PCA) es una técnica que permite reducir  $N$  variables predictoras en una cantidad menor de variables construidas como combinaciones lineales de las originales.

Sea  $X_t = (x_{1,t}, x_{2,t}, \dots, x_{N,t})$ , con  $1 \leq t \leq T$  el conjunto de  $N$  variables predictoras, las cuales se supone que han sido estandarizadas. Se consideran  $N$  combinaciones lineales de  $X_t$  como:

$$\begin{aligned} PC_{1,t} &= w_1 X_t = \sum_{k=1}^N w_{1,N} x_{k,t} \\ &\vdots \\ PC_{N,t} &= w_p X_t = \sum_{k=1}^N w_{k,N} x_{k,t} \end{aligned} \quad (8)$$

lo que resulta en  $N$  variables a las que se les conoce como componentes principales de  $X$ . A los coeficientes  $w_k$  ( $1 \leq k \leq N$ ) se les conoce como cargas (o *loadings*) porque muestran que tanto contribuye la variable  $x_{it}$  a explicar a cada uno de los componentes. Las cargas óptimas satisfacen tres restricciones: cada uno de los componentes principales tiene la mayor varianza posible, la correlación entre cada uno de los componentes es igual a cero y cada una de las cargas tiene norma igual a uno.

Otra variante de este mismo planteamiento realiza una selección de variables por medio de una regresión mínimos cuadrados parciales (PLSR por sus siglas en inglés), donde la solución de mínimos cuadrados parciales está dada por:

$$\beta = (X^T X)^{-1} X^T Y, \quad (9)$$

el problema a menudo es que  $X^T X$  es no invertible o singular, ya sea porque el número de variables (columnas) de  $X$  excede el número de objetos (filas), o debido a que una o más de las variables independientes son una combinación lineal de las otras variables, es decir, multicolinealidad perfecta.

Métodos de regresión multivariada como PCA y PLS evitan los problemas que se pueden encontrar en  $X^T X$  descomponiendo  $X$  en puntuaciones ortogonales  $T$  y cargas  $P$ :

$$X = TP.$$

PLS tiene como objetivo incorporar información sobre  $X$  e  $Y$  en la definición de las puntuaciones y las cargas, de tal manera que describen tanto como sea posible la covarianza entre  $X$  e  $Y$ .

En el segundo paso, se tiene en cuenta la heterocedasticidad de los componentes idiosincráticos y la dinámica de los factores comunes. Los valores verdaderos de los parámetros ahora se reemplazan por sus estimaciones, y la dinámica de los factores se estima a partir de las estimaciones preliminares asociadas de los factores. Una vez obtenidos los factores, se les puede incorporar en un modelo para  $y_t$  de la siguiente manera:

$$y_t = \beta_0 + \beta f_t + u_t, \quad (10)$$

$$y_t = \beta f_t + \frac{\theta(L)}{\varphi(L)} z_t. \quad (11)$$

Al modelar con cifras originales, se propone un modelo autorregresivo con medias móviles (ARMA) (Stock y Watson, 2016).

### 1.2.3 Modelo de factores dinámicos jerárquicos

Si un panel de datos se puede organizar en bloques utilizando información *a priori*, entonces la variación entre bloques y dentro de los mismos en los datos puede ser capturada por el marco del modelo de factor dinámico jerárquico. Cada bloque se puede dividir en subbloques para llegar a un modelo multinivel. Su característica principal es que las ecuaciones de transición para los factores en cada nivel tienen intersecciones variables en el tiempo que dependen de los factores en el siguiente nivel superior. Fueron considerados por Gammerman y Mignon (1993)

$$X_{ibt} = \lambda_{gib}(L) g_{bt} + e_{x_{ibt}}, \quad (12)$$

$$g_{bt} = \Lambda_{ib}(L) f_t + e_{g_{bt}}, \quad (13)$$

$$\Phi(L) f_t = u_t, \quad (14)$$

donde las variables  $X_{ibt}$  y  $X_{jbt}$  dentro de un bloque  $b$  están correlacionadas debido a los factores comunes  $f_t$  o los choques específicos del bloque  $e_{g_{bt}}$ , pero las correlaciones entre bloques solo son posibles a través de  $f_t$ . Algunas

de las  $X_{it}$  pueden no pertenecer a un bloque y podrán verse afectados los factores comunes directamente. Además, se puede permitir que los componentes idiosincráticos sigan procesos autorregresivos estacionarios:

$$\Phi_{X_{ib}}(L)e_{X_{ibt}} = \varepsilon_{X_{ibt}} \quad (15)$$

$$\Phi_{g_b}(L)e_{g_{bt}} = \varepsilon_{g_{bt}} \quad (16)$$

### 1.2.4 Regresión MIDAS

La regresión MIDAS (Ghysels *et al.* 2004, 2007) es una técnica de estimación que permite que los datos muestreados a diferentes frecuencias se utilicen en la misma regresión. Esta regresión permite la estimación de ecuaciones que explican una variable dependiente de baja frecuencia (como el consumo privado mensual) por variables de alta frecuencia y sus rezagos (como los montos diarios operados a través de tarjetas de crédito y débito), de una manera parsimoniosa pero flexible.

Sea  $Y_t^M$  la variable mensual del consumo y  $X_{m,t}^D$  el monto operado en tarjetas, donde  $M$  representa la frecuencia mensual,  $D$  la frecuencia diaria y  $m$  el número de transacciones diarias en el mes. Una predicción del consumo en  $h$  periodos en el futuro ha de seguir la forma:

$$Y_{t+h}^{M,h} = \mu^h + \sum_{j=0}^{p_Y^M-1} \rho_{j+1}^h Y_{t-j}^M + \beta^h \sum_{j=0}^{q_X^D-1} \sum_{i=0}^{m-1} w_{i+j*m}^{\theta^h} X_{m-1,t-j}^D + \varepsilon_{t+h}^h \quad (17)$$

Este modelo tiene una constante, el tradicional término AR con  $p_Y^M$  rezagos mensuales de la variable dependiente  $Y_t^M$  y un término que incorpora  $q_X^D$  veces  $m$  rezagos diarios para la variable independiente. El término que multiplica a la variable diaria  $w_{i+j*m}^{\theta^h}$  merece una atención especial. Este término es el esquema de ponderación que reducirá el número de parámetros a estimar y conducirá a un modelo más parsimonioso en lugar de tener que estimar un coeficiente para cada rezago de alta frecuencia.

Para evitar la proliferación de parámetros, la regresión MIDAS utiliza métodos de ponderación de los coeficientes rezagados para imponer restricciones en la ponderación de la agregación temporal, lo que permite reducir el número de parámetros a estimar independientemente del número de rezagos utilizados en el modelo. Como señala Ghysels *et al.* (2004, 2007) existen varios esquemas de ponderación que son útiles para reducir el número de parámetros a estimar. Entre estos se incluye: la función de probabilidad Beta normalizada, el polinomio de retardo de Almon, el polinomio de retardo de Almon exponencial normalizado, las funciones escalonadas y el MIDAS no restringido.

**Regresión MIDAS con selección de variables de frecuencia mixta.** El sistema de pagos de bajo valor, específicamente los montos operados a través de tarjetas de crédito y débito se componen de 30 rubros (véase Cuadro 2) mismos que al realizarse una suma se obtiene el total de monto operado a través de tarjetas de débito y crédito. El segundo modelo MIDAS aumentado con selección de variables, consiste en una reducción de dimensiones a partir de los 30 rubros que integran el monto total a través de la metodología *Lasso*, con el objeto de hallar el número mínimo de dimensiones capaces de expresar el máximo de información contenida en los montos operados a través de tarjetas. Posteriormente, con las variables que se han seleccionado a través de *Lasso*, se prosigue a la construcción de un factor estático, a partir del objetivo de encontrar un conjunto de factores  $F$  y un conjunto de parámetros  $\lambda$  que mejor expliquen:

$$F = \lambda f + \varepsilon.$$

Con base en lo anterior es posible formalizar el modelo de la siguiente manera. Sea  $Y_t^M$  la variable mensual del consumo y  $F_{m,t}^D$  el factor de los montos operados en tarjetas, donde  $M$  representa la frecuencia mensual,  $D$  la frecuencia diaria y  $m$  el número de observaciones diarias en el mes. Una predicción del consumo en  $h$  periodos en el futuro ha de seguir la forma:

$$Y_{t+h}^{M,h} = \mu^h + \sum_{j=0}^{p_Y^M-1} \rho_{j+1}^h Y_{t-j}^M + \beta^h \sum_{j=0}^{q_F^D-1} \sum_{i=0}^{m-1} w_{i+j*m}^{\theta^h} F_{m-1,t-j}^D + \varepsilon_{t+h}^h \quad (18)$$

## 2. IOCP: obteniendo estimaciones para $T + 2$

Una vez estimado  $T + 1$ , se considera como dato observado y se realiza el *nowcasting* para  $T + 2$ , siguiendo la metodología de Corona *et al.* (2022) pero con las modificaciones pertinentes en vías de tener estimaciones certeras y oportunas para el IMCPMI.

### 2.1 Fuentes de información

Las variables utilizadas en el modelo de *nowcasting* son seleccionadas con el criterio de oportunidad y correlación; es decir, que las variables se actualicen al menos antes que la variable a estimar y que, además, estén correlacionadas con esta última. También se busca que estas variables hayan sido utilizadas o consideradas por modelos previos en la literatura para el caso de México, como Corona *et al.* (2017), Caruso (2018) o recientemente, Gálvez-Soriano (2020). Todas las variables son de libre acceso<sup>1</sup> y se introducen al modelo desestacionalizadas, ya sea al obtenerlas así directamente de la fuente o bien al utilizar el paquete X-13ARIMA-SEATS. Es importante comentar que, en este contexto, las variables utilizadas para  $T + 1$  y  $T + 2$  tienden a ser las mismas, aunque no todas las que están actualizadas para  $T + 1$  lo están para  $T + 2$  (ver Cuadro 1).

### 2.2 Transformaciones previas

Si  $X^* = (X_1^*, \dots, X_N^*)'$  es la matriz  $T \times N$  de variables del MFD donde cada  $X_i^*$  es un vector de dimensión  $T$ , se busca una transformación que satisfaga la siguiente condición:

$$X_i = \left( f(X_i^*) \Big|_{\text{corr}} \max[f(X_i^*(T - H)), y] \right), \quad (19)$$

donde  $y = (y_1, \dots, y_{T-H})'$  es la serie de tiempo a la cual se le realizarán los *nowcasts* y  $H$  son los periodos para los cuales se realizarán dichas estimaciones.

En este caso las funciones consideradas son i) ninguna transformación, ii) cambio porcentual mensual y iii) cambio porcentual anual. Para el caso de las series de *Google Trends* se consideran i) ninguna transformación y ii) rezago de la variable.

### 2.3 PLS como método de selección de variables

Conforme a lo explicado en la subsección 1.2.2., se utiliza PLS como método de selección de variables, básicamente con la idea de encontrar las componentes principales, estimadas con un vector de variables  $W_t$  que covarían con  $y_t$ . En este sentido, las cargas asociadas a la combinación lineal  $Z_t = BW_t$ , se seleccionan aquellas variables donde los elementos de  $B$  sean distintos de 0 a través de técnicas de *Bootstrap*. Este procedimiento se utiliza para seleccionar, dentro de una gran cantidad de tópicos de *Google Trends*, aquellos que resultan relevantes para la variable objetivo en un periodo de datos de prueba, por ejemplo, los últimos 36 meses observados.

### 2.4 Modelo de factores dinámicos

El MFD subyacente considerado, difiere un poco al especificado en la subsección 1.2.2., dado que aquí hacemos énfasis también en la dinámica del componente idiosincrático. En consecuencia, y siguiendo la notación de esta sección tenemos que:

$$X = FP' + \varepsilon, \quad (20)$$

$$\Phi(L)F = \eta, \quad (21)$$

$$\Omega(L)\varepsilon = \alpha, \quad (22)$$

donde  $F$  es la matriz de los  $r (< N)$  factores dinámicos,  $P$  es la matriz de cargas o las contribuciones de los factores sobre las variables,  $\Phi(L)$  es la matriz que contiene los coeficientes autorregresivos de los factores y  $\eta$  es la matriz de disturbios del factor. Para el componente idiosincrático  $\varepsilon$ , se pueden hacer la misma analogía con relación a las

<sup>1</sup> El indicador de movilidad se genera con la metodología de Graff et al. (2020).

matrices  $\Omega(L)$  y  $a$ . Se asume que  $F$  puede ser no estacionario, pero  $\varepsilon$  se supone estacionario, restringiendo que las matrices que contienen  $\Phi(L)$  son diagonales. Note que  $L$  es el operador de rezagos.

Para estimar  $F$  y  $P$  se asumen algunas restricciones de identificabilidad (ver Bai y Ng, 2013) de tal forma que  $\tilde{P}$  es  $\sqrt{N}$  los primeros  $r$  vectores propios de la matriz  $X'X$  por lo que los factores estimados por componentes principales se determinan como:

$$\tilde{F} = X\tilde{P}N^{-1}. \quad (23)$$

Finalmente, se modela la dinámica de  $\tilde{F}$  ajustando un vector autorregresivo diagonal para obtener las matrices y condiciones iniciales que requiere el filtro de Kalman y así, utilizando su suavizamiento, se obtiene una versión  $\hat{F}$  que contempla la dinámica temporal del factor. Nótese que el suavizamiento de Kalman permite estimar el factor dinámico aun cuando no toda la información esté disponible al tiempo  $T$ , por ello su importancia en el contexto del *nowcasting*. Nótese que  $\tilde{P}^* = \tilde{P}N^{-1}$  puede verse como la contribución de las variables sobre el factor. En este trabajo se estima  $\hat{P}$  a través de técnicas Monte Carlo, la cual satisface la expresión (6) reemplazando  $\hat{F}$  por  $\tilde{F}$ , por lo que la contribución de las variables sobre el factor es  $\hat{P}^*$ .

## 2.5 Modelo de *nowcasting*

Una vez que se han estimado los factores se incorporan en el modelo de *nowcasting*, el cual escribimos de la siguiente manera:

$$y = \alpha + \beta\hat{F} + u, \quad (24a)$$

$$\phi_p(L)u = \theta_q(L)e. \quad (24b)$$

Nótese que las expresiones (24) asumen que el error  $u$  puede tener una estructura ARMA de tal manera que los órdenes  $p$  y  $q$  se determinan al minimizar el error de predicción en datos de prueba, actualizando los *nowcasts* un paso hacia adelante. La medida de error seleccionada es la Media Absoluta del Error (MAE) y el número de datos de prueba suelen ser los últimos 36 meses, aunque esto no es obligatorio; la estimación de los parámetros de las expresiones (24) se obtiene por máxima verosimilitud.

En consecuencia, una vez determinados los parámetros  $p$  y  $q$ , los *nowcasts*, o estimaciones oportunas, se realizan con la siguiente expresión:

$$y_{T+h} = \hat{a} + \hat{\beta}\hat{F}_{T+h} + \hat{u}_{T+h}, \quad (25a)$$

$$\hat{\phi}_p(L)\hat{u}_{T+h} = \hat{\theta}_q(L)\hat{e}_{T+h}, \quad (25b)$$

para  $h = 2$ . Un caso particular es cuando  $p$  y  $q$  son iguales a 0, en esta situación, el modelo es equivalente al que proponen Giannone *et al.* (2008), obra seminal en lo que se refiere a la literatura de *nowcasting*.

Finalmente, el IOCP se obtiene mediante la combinación de *nowcasts* con errores de estimación en datos de prueba estadísticamente iguales según la prueba de Diebold-Mariano. Como método de combinación, se considera la mediana de los *nowcasts* de los modelos con error estadísticamente igual al mejor modelo.

## 2.6 Consideraciones metodológicas

Algunas consideraciones estrictamente metodológicas son:

- Se desea, por interpretabilidad, que el número de factores sea  $\hat{r} = 1$ . Para ello se verifica este supuesto con el criterio de Onatski (2010).
- Se verifica que los errores idiosincráticos sean estacionarios con la prueba PANIC (Bai y Ng, 2004).
- Se verifica que los residuos,  $\hat{e}$ , no estén autocorrelacionados mediante la prueba de Ljung-Box.
- Se estiman los intervalos de confianza de los factores y las cargas según Bai (2003).
- La estimación Monte Carlo de los pesos del factor sobre las variables, una vez suavizado el factor por Kalman, se realiza utilizando 1000 muestras.

## Referencias

---

- [1] Bai, J. (2003) Inferential theory for factor models of large dimensions. *Econometrica* 71(1),135–171.
- [2] Bai, J. y Ng, S. (2002). Determining the Number of Factors in Approximate Factor Models. *Econometrica*, 70(1), 191-221.
- [3] Bai, J. y Ng, S. (2004) A PANIC attack on unit roots and cointegration. *Econometrica* 72(4),1127–1177.
- [4] Bai, J. y Ng, S. (2008). Forecasting economic time series using targeted predictors. *Journal of Econometrics*, 146(2), 304-317.
- [5] Bai, J. y Ng, S. (2008). Large Dimensional Factor Analysis. *Foundations and Trends in Econometrics*, 3(2), 89-163.
- [6] Bai, J., y Ng, S. (2013). Principal components estimation and identification of static factors. *Journal of Econometrics*, 176(1), 18-29.
- [7] Caruso, A. (2018). Nowcasting with the help of foreign indicators: The case of Mexico. *Economic Modelling*, 69,160-168.
- [8] Corona, F., González-Farías, G. y López-Pérez J. (2020). A nowcasting approach to generate timely estimates of Mexican economic activity: An application to the period of COVID-19. arXiv:2101.10383.
- [9] Corona, F., González-Farías, G. y López-Pérez J. (2022). Timely Estimates of the Mexican Economic Activity, *Journal of Official Statistics*, 38(3), 733-765.
- [10] Corona, F., González-Farías, G., y Orraca, P. (2017). A dynamic factor model for the Mexican economy: are common trends useful when predicting economic activity? *Latin American Economic Review*, 26(1), 1-35.
- [11] Doz, C., Giannone, D., y Reichlin, L. (2011). A two-step estimator for large approximate dynamic factor models based on Kalman filtering. *Journal of Econometrics*, 164(1), 188-205.
- [12] Friedman, J., Hastie, T. y Tibshirani, R. (2009). Regularization Paths for Generalized Linear Models via Coordinate Descent. *Journal of Statistical Software*. 33(1), 1-22.
- [13] Gálvez-Soriano, Oscar. (2020). Nowcasting Mexico's quarterly GDP using factor models and bridge equations. *Estudios Económicos*, 35(2):213-265.
- [14] Gammerman, D. y Mignon, H. S. (1993). Dynamic Hierarchical Models. *Journal of Royal Statistical Society Series B*55, 629-642.
- [15] Geweke, J. (1977). The Dynamic Factor Analysis of Economic Time Series Models Latent Variables. *Socioeconomic Models*, 365-83.
- [16] Ghysels, E., Santa-Clara, P., y Valkanov, R. (2004). The MIDAS touch: Mixed data sampling regression models. *UCLA: Finance*.
- [17] Ghysels, E., Sinko, A., y Valkanov, R. (2007). MIDAS Regressions: Further Results and New Directions. *Econometric Reviews*, 26(1), 53-90.
- [18] Giannone, D., Reichlin, L. y Small, D. (2008). Nowcasting: the real-time informational content of macroeconomic data. *Journal of Monetary Economics*.
- [19] Graff, M., Moctezuma, D., Miranda-Jiménez, S., y Téllez E.S. (2020). A Python Library for Exploratory Data Analysis and Knowledge Discovery on Twitter Data. *arXiv* working paper, <https://arxiv.org/pdf/2009.01826.pdf>

- [20] Hoerl, A. (1970). Ridge Regression: Biased Estimation for Nonorthogonal Problems. *Technometrics*, 12(1), 55-67.
- [21] Khun, M. y Johnson, K. (2013). *Applied Predictive Modeling*. Springer, New York.
- [22] Lever, J., Krzywinski, M. y Altman, N. (2016). Regularization. *Nature Methods* 13, 803-804. <https://doi.org/10.1038/nmeth.4014>.
- [23] Onatski, A. (2010). Determining the number of factors from empirical distribution of eigenvalues. *The Review of Economics and Statistics*, 92(4), 1004-1016.
- [24] Stock, J. y Watson, M. (2002). Forecasting Using Principal Components From a Large Number of Predictors. *Journal of the American Statistical Association*, 460(97).
- [25] Stock, J. y Watson, M. (2016). Factor Models and Structural Vector Autoregressions in Macroeconomics, in Taylor, J. y Uhlig, H. (eds.) *Handbook of Macroeconomics Volume 2*, Elsevier.
- [26] Tibshirani, R. (1996). Regression shrinkage and selection via the Lasso. *Journal of the Royal Statistical Society*, 58(1), 267-288.
- [27] Zou, H. y Hastie, T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society, Series B*, 67(2), 301-320.

## Anexos

## A. Variables utilizadas para la estimación del IOCP

Cuadro 1

Variable	Descripción	Fuente
<b>Variables tradicionales</b>		
ANTAD	Ventas totales	ANTAD
BMV	Índice de precios y cotizaciones de la Bolsa Mexicana de Valores	Banxico
Cartera de crédito vigente	Cartera vigente otorgada al sector privado no bancario destinada al consumo	Banxico
Confianza-comercio	Momento adecuado para invertir del sector comercial	INEGI
Confianza-construcción	Momento adecuado para invertir del sector construcción	INEGI
Confianza-manufacturas	Momento adecuado para invertir de las manufacturas	INEGI
Confianza-servicios	Momento adecuado para invertir del sector servicios	INEGI
Cuartos hotel	Ocupación de cuartos en hoteles y moteles	SECTUR
Demanda diesel	Demanda nacional de Diesel	SENER
Demanda gasolina	Demanda nacional de gasolina	SENER
Demanda turbosina	Demanda nacional de turbosina	SENER
Desempleo Estados Unidos	Tasa de desempleo abierto, U3	Bureau of Labor Statistics
Desocupación	Tasa de desocupación en áreas urbanas	INEGI
Empleados construcción EUA	Total de empleados, miles, construcción	Bureau of Labor Statistics
Empleados manufacturas EUA	Total de empleados, miles, manufacturas	Bureau of Labor Statistics
Empleados servicios EUA	Total de empleados, miles, servicios	Bureau of Labor Statistics
Empleados totales EUA	Total de empleados, miles	Bureau of Labor Statistics
Empleo de manufacturas	Personal ocupado (serie de tenencia)	INEGI
ENOE PEA	Población Económicamente Activa (ENOE)	INEGI
ENOE PO	Población Ocupada (ENOE)	Banxico
Exportaciones	Exportaciones totales	INEGI
FLASH*	Información interna de carácter confidencial del INEGI	Directa
Google Trends	Tópicos relevantes de Google Trends	Google
IATM	Indicador Agregado de Tendencia para la Manufactura	INEGI
ICC	Indicador de Confianza del consumidor	INEGI
IMP bienes consumo	Importaciones de bienes de consumo de la BCMM	Banxico
IMP bienes intermedios	Importaciones de bienes intermedios de la BCMM	Banxico
Importaciones	Importaciones totales	INEGI
IMSS	Asegurados permanentes y eventuales del Seguro Social	Instituto Mexicano del Seguro Social
INPC	Índice Nacional de Precios al Consumidor	INEGI
ITCR	Índice de tipo de cambio real bilateral con EUA	Banxico
M4	Agregado monetario M4	Banxico
Ocupación hotelera	Ocupación hotelera en corredores y agrupamientos	Secretaría de Turismo
Pasajeros avión	Número de pasajeros transportados por la aviación comercial	ASA
Pedidos manufactureros	Indicador de pedidos manufactureros	INEGI
Precio del petróleo	Precio de la mezcla mexicana de petróleo	Banxico
Producción de vehículos	Producción de vehículos automotores	INEGI
Producción industrial	Índice de producción industrial	INEGI
Producción industrial-EUA	Índice de producción industrial de los Estados Unidos	Bureau of Economic Analysis
Producción manufacturera-EUA	Índice de producción manufacturera de los Estados Unidos	Bureau of Economic Analysis
Remesas	Remesas familiares	Banxico

(Continúa)

Variable	Descripción	Fuente
<b>Variables tradicionales</b>		
SP 500	Índice Standard & Poor's	Yahoo! Financiero
SPEI	Transferencias realizadas en el Sistema de Pagos Electrónicos Interbancarios	Banxico
TIIIE 28	Tasa de interés interbancaria de equilibrio a 28 días	Banxico
Tipo de cambio	Tipo de cambio nominal promedio	Banxico
Ventas al por menor	Ingresos por suministros de bienes y servicios Comercio al por menor	INEGI
Ventas CFE	Ventas de la CFE-CRE	Directa
Ventas vehículos	Ventas de vehículos las Estadísticas de Vehículos de Motor Registrados en Circulación	INEGI
<b>Variables no tradicionales</b>		
Google Trends	Tópicos relevantes de Google Trends	Google
Movilidad google	Índice de movilidad de Google	Google
Movilidad Twitter	Índice de movilidad de Twitter	Elaboración propia con datos de Twitter
<b>Variables tradicionales de alta frecuencia</b>		
Monto tarjetas bancarias	Transacciones realizadas con tarjetas bancarias	Banxico
Monto tarjetas crédito	Monto real operado a través de tarjetas de crédito	Banxico
Monto tarjetas débito	Monto real operado a través de tarjetas de débito	Banxico

*\*Esta variable no se incluye como insumo en los tabulados debido a su carácter confidencial.*

## B. Sistemas de pago de bajo valor de Banxico

Cuadro 2

<b>Monto operado a través de tarjetas en</b>		
1. Agencias de viajes	11. Gasolineras	21. Peaje
2. Agregadores	12. Gobierno	22. Refacciones y ferretería
3. Aseguradoras	13. Grandes superficies	23. Renta de autos
4. Beneficencia	14. Guarderías	24. Restaurantes
5. Colegios y universidades	15. Hospitales	25. Salones de belleza
6. Comida rápida	16. Hoteles	26. Supermercados
7. Educación básica	17. Misceláneos	27. Telecomunicaciones
8. Entretenimiento	18. Médicos y dentistas	28. Transporte aéreo
9. Estacionamientos	19. No definido	29. Transporte terrestre de pasajeros
10. Farmacias	20. Otros	30. Ventas al detalle (Retail)