Indicador Oportuno de la Actividad Económica

Síntesis metodológica

Antecedentes

En 2019, la Dirección General Adjunta de Investigación incorporó dentro de su Plan Anual de Investigación el desarrollo de un modelo de *nowcasting* para la actividad económica de México, utilizando un amplio número de series de tiempo económicas y financieras que permitieran estimar, de manera oportuna, el trimestre inmediato del Producto Interno Bruto (PIB).

La idea fundamental fue explotar la información de series de tiempo mensuales; principalmente las que proporciona el Banco de Información Económica. Para lo anterior, se propuso la utilización de regresión LASSO como método de selección de variables, determinando aquellas que resultaran frecuentemente significativas en un periodo de datos de entrenamiento. Una vez seleccionadas las variables, se estimó un Modelo de Factores Dinámicos (MFD) con la metodología de Doz *et al.* (2011), con el objetivo de utilizar los factores significativos que minimizaran el error de predicción en un periodo de datos de entrenamiento, y seleccionado la especificación con menor error de estimación para realizar el *nowcast* final.

A partir del 2020 se enfocaron los esfuerzos en estimar los *nowcasts* del Indicador Global de la Actividad Económica (IGAE) dado que esta serie es más oportuna que el PIB, principalmente por su periodicidad mensual, lo que conllevó a realizar una serie de ajustes econométricos. Dichas estimaciones se denominan Indicador Oportuno de la Actividad Económica (IOAE).

El modelo actual de *nowcasting* tiene las siguientes características:

- Estimación para los dos meses posteriores al cierre del último dato publicado oficialmente para el IGAE, así como para las actividades secundarias y terciarias.
- Selección de variables económicas y financieras con el criterio de oportunidad y alta correlación respecto a la variable a estimar.
- Incorporación de fuentes no tradicionales de información y de alta frecuencia como Google Trends y un indicador de movilidad.
- Selección de tópicos relevantes de Google Trends a través de regresión LASSO y/o regresión con validación cruzada.
- Transformación de variables de tal forma que maximice la correlación con la variable a estimar.
- Validación estadística sobre el número de factores.
- Prueba de estacionariedad para los errores idiosincráticos que validan la estimación consistente de los factores y de las cargas asociadas.
- Combinación de *nowcasts* con modelos que tienen error de estimación estadísticamente igual (Prueba Diebold-Mariano).

- Estimación de intervalos de confianza al 95% para los factores, cargas de variables y *nowcasts*.[1]
- Estimación Monte Carlo de los pesos de las variables una vez suavizado el factor por Kalman.

Finalmente, en ningún caso, el IOAE es predicción o pronóstico, sino estimaciones que se realizan con la información oportuna y de la alta frecuencia que requiere el modelo. Las estimaciones realizadas anteceden a las cifras oficiales, por ende, dichas estimaciones se actualizan conforme la información también lo hace. Es importante señalar que el IOAE no puede considerarse como estadística oficial.

1. Marco teórico

1.1. 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), Carusso (2018) o recientemente, Gálvez-Soriano (2020). Todas las variables son de libre acceso[2] y se introducen al modelo desestacionalizadas, ya sea al obtenerlas así directamente de la fuente o bien al utilizar el paquete X-13ARIMA-SEATS.

Las variables consideradas se describen en la Tabla 1.[3]

1.2. Metodologías

1.2.1. 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

$$X_i = \left(f(X_i^*) \middle| \max_{corr} [f \langle X_i^*(T-H) \rangle, y] \right), \tag{1}$$

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.

1.2.2. Regresión LASSO

[1] Los intervalos de confianza no son necesariamente simétricos debido a que su estimación considera la mediana de los intervalos inferior y superior al 95%, para los modelos que, en datos de prueba, otorgan nowcasts con errores estadísticamente igual según la prueba de Diebold-Mariano.
[2] El indicador de movilidad se genera con la metodología de Graff *et al.* (2020).
[3] El número de variables puede modificarse conforme se disminuya el error de estimación en el periodo de validación cruzada de series de tiempo, fase en la cual se seleccionan los modelos de nowcasting.

La regresión LASSO (Tibshirani, 1996) es una técnica que puede ser utilizada como método de selección de variables cuando se tienen una gran cantidad de estas, dado que las técnicas tradicionales basadas en criterios de información quedan rebasadas al incrementarse exponencialmente la cantidad de covariables. Minimizar la suma de cuadrados residuales sujeto a que la suma del valor absoluto de los coeficientes sea menor que una constante, permite alcanzar dos objetivos: i) dejar fuera todas las variables irrelevantes y retener aquellas que sí lo son y ii) estimar los coeficientes con la misma velocidad y distribución con la que se hubieran estimado de conocer desde un principio cuáles eran las variables relevantes. 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.

1.2.3. Modelo de factores dinámicos

El MFD subyacente considerado es:

$$X = FP' + \varepsilon, \tag{2}$$

$$\Phi(\mathrm{L})F = \eta, \tag{3}$$

$$\Omega(\mathrm{L})\varepsilon = a, \tag{4}$$

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, $\Phi(\mathrm{L})$ es la matriz que contiene los coeficientes autorregresivos de los factores y η es la matriz de disturbios del factor. Para el componente idiosincrático ε, se pueden hacer la misma analogía con relación a las matrices $\Omega(\mathrm{L})$ y a. Se asume que F puede ser no estacionario, pero ε se supone estacionario, restringiendo que las matrices que contienen $\Phi(\mathrm{L})$ 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:

$$\tilde{F} = X\tilde{P}N^{-1}. \tag{5}$$

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 una de sus importancias 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 (5) reemplazando $\hat{F}$ por $\tilde{F}$, por lo que la contribución de las variables sobre el factor es $\hat{P}^*$.

1.2.4. 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:

$$y = \alpha + \beta\hat{F} + u, \tag{6a}$$

$$\phi_p(L)u = \Theta_{\mathrm{q}}(L)e. \tag{6b}$$

Nótese que la expresión (6) asume que el error u puede tener una estructura autorregresiva y de medias móviles (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, y la estimación de los parámetros de la expresión (6) 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:

$$y_{T+h} = \hat{\alpha} + \hat{\beta}\hat{F}_{T+h} + \hat{u}_{T+h}, \tag{7a}$$

$$\hat{\phi}_p(L)\hat{u}_{T+h} = \widehat{\Theta}_{\mathrm{q}}(L)\hat{e}_{T+h}, \tag{7b}$$

para $h = 1{,}2$. 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 IOAE se obtiene mediante la combinación de *nowcasts* con errores de estimación en datos de prueba estadísticamente iguales según la prueba de Diebold-Mariano. Se considera la mediana de estos, así como el modelo tradicional propuesto por Giannone *et al.* (2008), ponderando los pesos de acuerdo también, a su error de estimación.

1.3. 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 1,000 muestras.

Para mayores detalles sobre la metodología se recomienda revisar a Corona *et al.* (2020)

Referencias

Bai, J. (2003) Inferential theory for factor models of large dimensions. *Econometrica* 71(1):135–171.

Bai, J. y Ng, S. (2004) A PANIC attack on unit roots and cointegration. *Econometrica* 72(4):1127–1177.

Bai, J., y Ng, S. (2013). Principal components estimation and identification of static factors. *Journal of Econometrics*, *176*(1), 18-29.

Caruso, A. (2018). Nowcasting with the help of foreign indicators: The case of Mexico. *Economic Modelling*, 69:160-168.

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.

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. Working Paper. https://arxiv.org/abs/2101.10383

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.

Galvez-Soriano, Oscar. (2020). Nowcasting Mexico's quarterly GDP using factor models and bridge equations. *Estudios Economicos*, 35(2):213-265.

Giannone, D., Reichlin, L., y Small, D. (2008). Nowcasting: The real-time informational content of macroeconomic data. *Journal of Monetary Economics*, *55*(4), 665-676.

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. *Working paper, https://arxiv.org/pdf/2009.01826.pdf*

Onatski, A. (2010). Determining the number of factors from empirical distribution of eigenvalues. *The Review of Economics and Statistics*, *92*(4), 1004-1016.

Tibshirani, R. (1996). Regression shrinkage and selection via the LASSO. *Journal of the Royal Statistical Society: Series B (Methodological)*, *58*(1), 267-288.

Anexos

Tabla 1. Variables del MFD

Variable	Descripción	Fuente
ANTAD	Ventas totales	ANTAD
Confianza-comercio	Momento adecuado para invertir del sector comercial	INEGI
Confianza-construcción	Momento adecuado para invertir del sector comercial	INEGI
Confianza-manufacturas	Momento adecuado para invertir de las manufacturas	INEGI
Confianza-servicios	Momento adecuado para invertir del sector servicios	INEGI
Combustibles-SENER	Demanda de combustibles	Secretaría de Energía
Producción industrial	Índice de producción industrial	INEGI
IMSS	Asegurados permanentes y eventuales del Seguro Social	Instituto Mexicano del Seguro Social
BMV	Índice de precios y cotizaciones de la Bolsa Mexicana de Valores	Banco de México
Producción industrial-EUA	Índice de producción industrial de los Estados Unidos	Bureau of Economic Analysis
Ventas al por menor	Ingresos por suministros de bienes y servicios. Comercio al por menor.	INEGI
Importaciones	Importaciones totales	INEGI
M4	Agregado monetario M4	Banco de México
Movilidad Twitter	Índice de movilidad de Twitter	INEGI, CentroGeo, INFOTEC
Ocupación hotelera	Ocupación hotelera en corredores y agrupamientos	Secretaría de Turismo
Pedidos manufactureros	Indicador de pedidos manufactureros	INEGI

Producción de vehículos	Producción de vehículos automotores	INEGI
Remesas	Remesas familiares	Banco de México
SP 500	Índice Standard & Poor's	Yahoo! Financiero
Empleo de manufacturas	Personal ocupado (serie de tenencia)	INEGI
Tipo de cambio	Tipo de cambio nominal promedio	Banco de México
Desocupación	Tasa de desocupación en áreas urbanas	INEGI
TIIE 28	Tasa de interés interbancaria de equilibrio a 28 días	Banxico
Exportaciones	Exportaciones totales	INEGI
Tópico-Google Trends	Tópicos de Google Trends	Google