

# Medición de la Actividad Económica a partir de Grandes Datos (MAGDA)

*IMCO y OPI con el apoyo de la Fundación Friedrich Naumann\**

*Ciudad de México a 18 de enero de 2017*

## Objetivo

El propósito de este proyecto es generar una medición de los niveles y el crecimiento de la actividad económica en las zonas metropolitanas y que tenga las siguientes características:

- *Uso de grandes datos.* La actividad económica produce bases de datos vivas y grandes (*big data*) que pueden aprovecharse para estimar la actividad económica.
- *Desagregación.* Más allá de los estados, MAGDA se enfocará en evaluar las ciudades y zonas metropolitanas más importantes del país.
- *Frecuencia.* MAGDA utilizará datos con frecuencia trimestral, por lo que podrá ser publicado oportunamente.
- *Accesibilidad.* Los insumos no dependen de las cuentas nacionales. Se utilizarán datos transaccionales y satelitales en la elaboración, que están disponibles para el público.
- *Participación.* Además de publicarse regularmente, MAGDA hace uso de tecnologías y metodologías modernas, para ser accesible al público en general. En un repositorio de control de versiones<sup>1</sup>, se harán accesibles las fuentes de los datos así como el código utilizado para procesarlo, de principio a fin. Los usuarios interesados podrán replicar el proceso, hacer modificaciones para uso propio, o incluso proponer ajustes a la metodología.

## Descripción de los datos

Se utilizaron datos que corresponden a tres variables principales: producto o actividad económica (Producto Interno Bruto Estatal –PIBE– e Indicador Trimestral de la Actividad Económica Estatal –ITAEE–), mediciones de luminosidad nocturna (fuente abierta) y transacciones en cajeros automáticos de la Comisión Nacional Bancaria y de Valores –CNBV–. Los primeros, PIBE e ITAEE, son generados por el INEGI para las entidades federativas y constituyen la guía de calibración de datos para MAGDA. Se relaciona el PIBE y el ITAEE con las variables de luminosidad y CNBV, que se calculan para las entidades federativas y las nuevas desagregaciones.

A continuación se describen las bases de agregación que se utilizan en el proceso de modelado de MAGDA.

## Agregación sociopolítica

Se utilizan cuatro niveles de agregación  $\mathcal{L}$ ,  $\mathcal{M}$ ,  $\mathcal{C}$ ,  $\mathcal{E}$  que corresponden a localidades urbanas, municipios, ciudades o áreas metropolitanas y estados.

La forma en que se utiliza cada nivel es la siguiente:

---

\*El Instituto Mexicano para la Competitividad (IMCO) es una asociación civil que propone políticas públicas y acciones viables e influir en su ejecución para lograr un México próspero e incluyente ([www.imco.org.mx](http://www.imco.org.mx)). OPI es una empresa líder de mercado en la investigación de grandes datos ([www.opi.la](http://www.opi.la)). Fundación Friedrich Naumann es una organización no gubernamental dedicada a la revaloración del Liberalismo como filosofía (<http://www.la.fnst.org/>)

<sup>1</sup>[www.github.com](http://www.github.com)

- *Estados*: los índices existentes más usados se publican a nivel estatal y nacional. A medida que generamos un nuevo índice, hay que compararlo con los indicadores existentes para validarlo, y éso se hizo con la agregación a nivel estatal.
- *Ciudades, o zonas metropolitanas*: MAGDA tiene como objetivo describir la actividad económica en las ciudades, que son el núcleo del crecimiento del país. Es posible obtener datos de los niveles correspondientes, ya sea agregándolos o heredándolos. Por ejemplo, los niveles de luminosidad o de área se obtienen a nivel municipal, y después se suman para tener los niveles de las ciudades. Por otro lado, siguiendo las mediciones oficiales para los estados, se hacen modelos que relacionan los comportamientos del PIBE y de luminosidad, mismos que se heredan a las ciudades correspondientes.
- *Municipios*: los municipios constituyen la división administrativa básica en este proyecto que cubre el territorio nacional. Tanto los datos de luminosidad como los de la CNBV se pueden obtener en este nivel de agregación.
- *Localidades urbanas*: si los municipios constituyen la base de MAGDA, se utilizan estas localidades para definir el soporte del mismo. Específicamente para las mediciones de luminosidad que se explican más adelante, ésto implica fijar cada uno de los anteriores estados, zonas metropolitanas, municipios a sus correspondientes localidades, donde se concentra la mayoría de la luminosidad y de su actividad económica.

Se utilizó el Marco Geoestadístico Nacional (INEGI, 2014) para identificar las regiones correspondientes. A partir de los municipios se agregan tanto en ciudades como en estados para obtener las variables en cada nivel. La agregación de municipios en ciudades sigue la metodología que usa IMCO para su Índice de Competitividad Urbana<sup>2</sup>.

## Producto y actividad económica

El PIBE es el Producto Interno Bruto por Entidad Federativa; es el indicador oficial de actividad económica de los estados y se publica anualmente por el INEGI. En diferentes periodos que duran alrededor de una década, el PIBE conlleva distintas metodologías que incorporan reglamentación novedosa del momento. El periodo más reciente, de 2008 a 2014, sigue los lineamientos del Sistema de Clasificación Industrial de América del Norte. Se puede obtener más información en el sitio web del INEGI.

Sin embargo el cálculo del PIBE conlleva un tiempo de rezago que incluye la recaudación de las cuentas de las diferentes secretarías y la revisión de las metodologías con estándares internacionales. Para una publicación más reciente y frecuente se cuenta con el Índice Trimestral de Actividad Económica Estatal, ITAEE. La serie de este índice se extiende desde 2003 hasta 2016 y tiene periodicidad trimestral, como su nombre lo indica. De acuerdo a la documentación del ITAEE debe considerarse como un adelanto del PIBE pues incorpora parte de la metodología correspondiente, aunque no coincide del todo con los cálculos anuales debido principalmente a la calendarización de la actividad primaria<sup>3</sup>.

Se utilizan los datos del PIBE en un periodo de fijo en  $t_0 = 2014$  y que se denota como  $\pi_{e,t_0}$  para cada estado  $e \in \mathcal{E}$ . Las series del ITAEE se escribe como  $\iota_{e,t}$  donde  $t \in \mathcal{T}$  se extiende desde 2011 a 2016. De acuerdo a estos índices se estimaran los niveles y crecimientos de los estados que después se extiende a las zonas metropolitanas  $c \in \mathcal{C}$ .

## Luminosidad nocturna

Para medir la luminosidad nocturna se usaron imágenes del sitio de la Administración Americana Oceánica y Atmosférica de los Estados Unidos (NOAA por sus siglas en inglés). Los archivos tienen fotos satelitales del globo terráqueo e información específica para seleccionar el territorio de interés de la República Mexicana. Las imágenes fueron sometidas previamente a procesos en los cuales se limpian de efectos que distorsionan la

<sup>2</sup><http://imco.org.mx/competitividad/indice-de-competitividad-urbana-2016/>

<sup>3</sup>Fuente: [http://www.inegi.org.mx/est/contenidos/proyectos/scn/c\\_anuales/pib\\_ef/presentacion.aspx](http://www.inegi.org.mx/est/contenidos/proyectos/scn/c_anuales/pib_ef/presentacion.aspx)

información. Por ejemplo, la presencia de nubes en alguna región obstruye la luz cuando se toma la foto satelital y la haría parecer menos luminosa. Los archivos comprenden periodos de un mes, de donde se puede distinguir cuáles de las zonas fueron distorsionadas por nubes y remover dicho efecto.

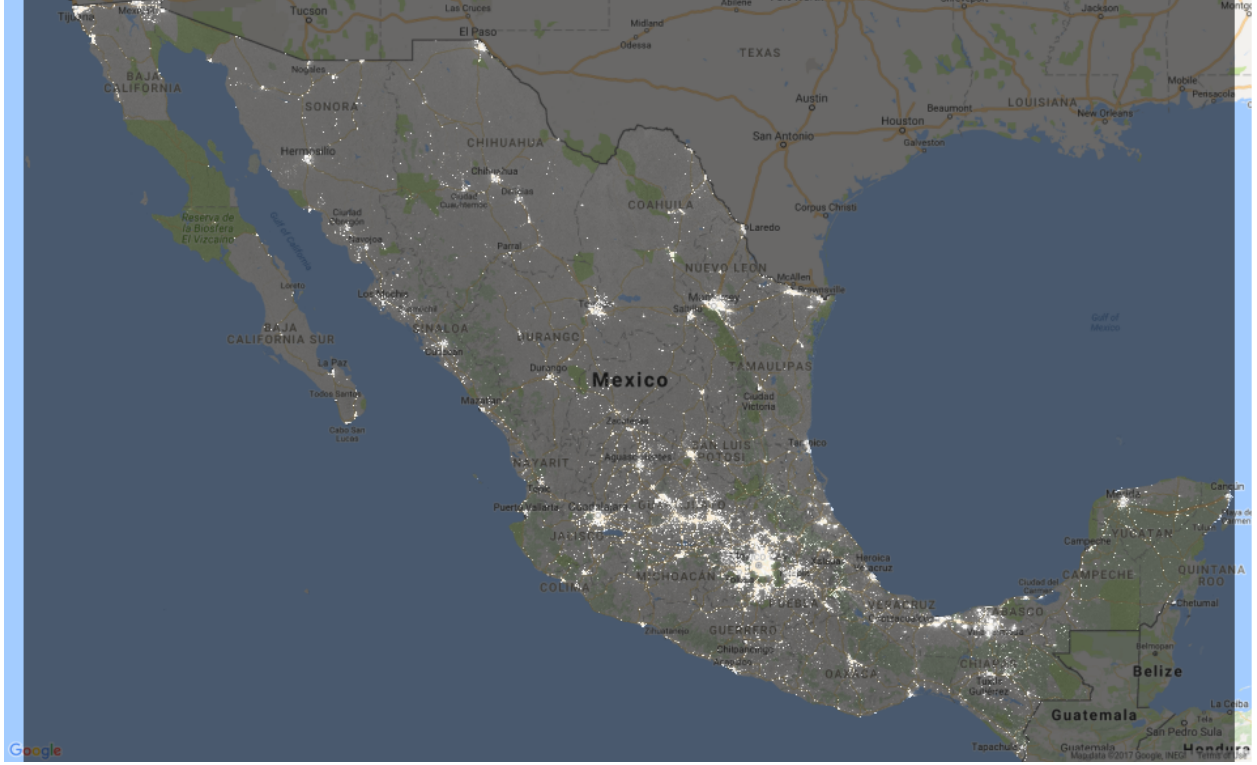


Figura 1: Foto satelital

Las fotos que se utilizaron son generadas con tecnología reciente que se llama *Suite de Radiometría de Imágenes Visibles Infrarrojas*, VIIRS. Estas imágenes tienen una resolución de  $0.55 \text{ km}^2$  y usan unidades de radiación  $\ell \sim \frac{\text{Watt}}{\text{cm}^2 \text{sr}}$ , es decir potencia entre área esférica. Estas mediciones las llamamos *luminosidad* y denotamos como  $\ell$ ,  $\lambda$  o  $\Lambda$ , dependiendo del nivel y agregación.

Estas fuentes satelitales proporcionan mediciones geográficas  $x_i, y_i$  y de luminosidad  $\ell_i$  que se indexan con  $i \in \mathcal{I}$ ; las mediciones corresponden a los pixeles de la cuadrícula o ráster. Relacionándolo con las regiones  $R \in \mathcal{M}, \mathcal{C}, \mathcal{E}$  se tiene que el área correspondiente es proporcional al número de pixeles contenidos en ellas

$$A(R) \propto \sum_{i \in \mathcal{I}} \chi_R(x_i, y_i).$$

De forma similar se calculan la luminosidad total y media de las regiones como

$$\Lambda(R) = \sum_{i \in \mathcal{I}} \ell_i \chi_R(x_i, y_i) \quad y \quad \lambda(R) = \frac{\Lambda(R)}{A(R)},$$

donde  $\chi_R(x_i, y_i)$  es la indicadora de cada pixel  $i$  en cada región  $R$  de acuerdo a sus coordenadas  $(x_i, y_i)$ .

Mas aún dado que MAGDA medirá la actividad económica en las zonas metropolitanas, se considera la luminosidad urbana como la restricción de las luminosidades a las localidades urbanas. Tómese  $U = \bigcup_{\mathcal{C}} L$  como la unión de todas las localidades urbanas, y se indice mediante el superíndice  $(\cdot)^U$  las restricciones correspondientes

$$A^U(R) = A(R \cap U), \quad \Lambda^U(R) = \Lambda(R \cap U), \quad \lambda^U(R) = \lambda(R \cap U).$$

La restricción se hizo con software especializado de geolocalización<sup>4</sup> utilizando archivos con formato específico para cartografía digital (*shapefiles*). Dichos *shapefiles* provienen del Marco Geoestadístico Nacional.

Finalmente se impone una segunda restricción a las zonas urbanas por considerar. Esta restricción consiste en limitar  $U$  a las zonas en donde el valor de  $\ell \leq 175 \frac{\text{Watt}}{\text{cm}^2\text{sr}}$ . El motivo de esto es que se encontraron regiones de pocos píxeles con luminosidades sumamente altas en el orden de millares, cuyas zonas se sabe que no tienen tanta actividad económica. Específicamente, las zonas se identificaron en sitios de actividad petrolera, y la luminosidad se asocia a la combustión de los procesos implicados. El tope de 175 se toma relativo al máximo de la Zona del Valle de México.

Manteniendo la notación simple, se redefine la zona urbana  $U$  como libre de combustión donde solamente se consideran los píxeles con  $\ell \leq 175$ . La luminosidad correspondiente sigue esta consideración.

Con estas medidas se calcularán los niveles de MAGDA en las zonas metropolitanas.

## Comisión Nacional Bancaria y de Valores

La CNBV proporciona datos mensuales de transacciones y otras variables bancarias a nivel de localidad y por institución. Se utilizaron series de transacciones en cajeros automáticos que se denota como  $\mu$ . Las series más básicas llevan subíndices  $\mu_{m,\tau,b}$  y corresponden al municipio  $m \in \mathcal{M}$ , periodo  $\tau \in \mathcal{T}$  e institución bancaria  $b \in \mathcal{B}$ .

Las agrupaciones en cada índice se obtienen como sumas de los datos individuales. Se agregan las operaciones tanto por trimestre como por grupos de instituciones bancarias. Estos grupos se especifican para cada estado  $B_e \subset \mathcal{B}$ ; se seleccionan a partir de bancos principales: Banamex, BBVA Bancomer, Otros Bancos o Todos Bancos y algunas posibles combinaciones de ellos.

El motivo de esta especificación es que se encontraron series  $\mu_{m,t,b}$  con anomalías ocurriendo en diferentes regiones. Debido a esto se aplicó el proceso de X11 y se eligió un grupo de bancos  $B_e$  para cada estado  $e \in \mathcal{E}$ ; a partir de ellas se obtuvieron las series de los municipios correspondientes. El proceso detallado de esta selección se puede ver en el apéndice.

Por ejemplo para  $e = \text{CDMX}$  se determinaron las series Banamex y Otros Bancos como el grupo  $B_e$ . Entonces para los municipios correspondientes  $m \subset e$  se suman las series de Banamex y Otros Bancos.

$$\mu_{m,t} = \sum_{\substack{\tau \sim t \\ b \in B_e}} \mu_{m,\tau,b}$$

y donde  $\tau \sim t$  son los meses de cada trimestre.

Para las ciudades y estados  $R \in \mathcal{C}$ ,  $\mathcal{E}$  se suman sobre los municipios correspondientes

$$\mu_{R,t} = \sum_{m \subset R} \mu_{m,t}.$$

Por último, las series de transacciones de cajeros aproximan a la actividad económica a través de los crecimientos proporcionales. Se escribe el crecimiento trimestral mediante el superíndice  $(\cdot)^\Delta$ ,

$$\mu_{R,t}^\Delta = \frac{\mu_{R,t} - \mu_{R,t-1}}{\mu_{R,t-1}}.$$

## Modelado

La estrategia para desarrollar MAGDA consiste en entrenar un modelo a nivel estatal, para después replicarlo con los datos correspondientes de las zonas metropolitanas. Denotamos este índice de actividad económica como

<sup>4</sup>Se utiliza el sistema de información geográfica QGIS.

$\varrho_t(R)$  y se ajusta a diferentes regiones  $R \in \mathcal{M}, \mathcal{C}, \mathcal{E}$ , nótese la distinción con el PIBE y el ITAEE que miden el producto de los estados  $e \in \mathcal{E}$  y escribimos como  $\pi_{e,t}$  e  $\iota_{e,t}$ .

A su vez el modelo se separa en dos partes que corresponden al nivel y al crecimiento. El cálculo de niveles se basa en los datos de luminosidad urbana, mientras que el de crecimiento en los datos de cajeros automáticos. Esta división apoya los siguientes puntos considerables:

- La luminosidad refleja el tamaño de las economías. Los artículos de Chen y Henderson han apoyado esta tesis, y examinan características donde la relación se hace más o menos robusta.
- Un supuesto fundamental de MAGDA es que la actividad económica proviene mayoritariamente de las localidades urbanas de las regiones. Esto se debe a que hay zonas oscuras que por su gran extensión acumulan luminosidad, pero que no es acorde con la producción debido a que están despobladas.
- El crecimiento en la actividad económica de las ciudades está ligado a las transacciones monetarias que hacen los usuarios de la banca, y específicamente a su demanda de efectivo. Este punto concuerda con el cálculo del ITAEE que incluye dichas transacciones a la vez que se permite una agregación refinada.

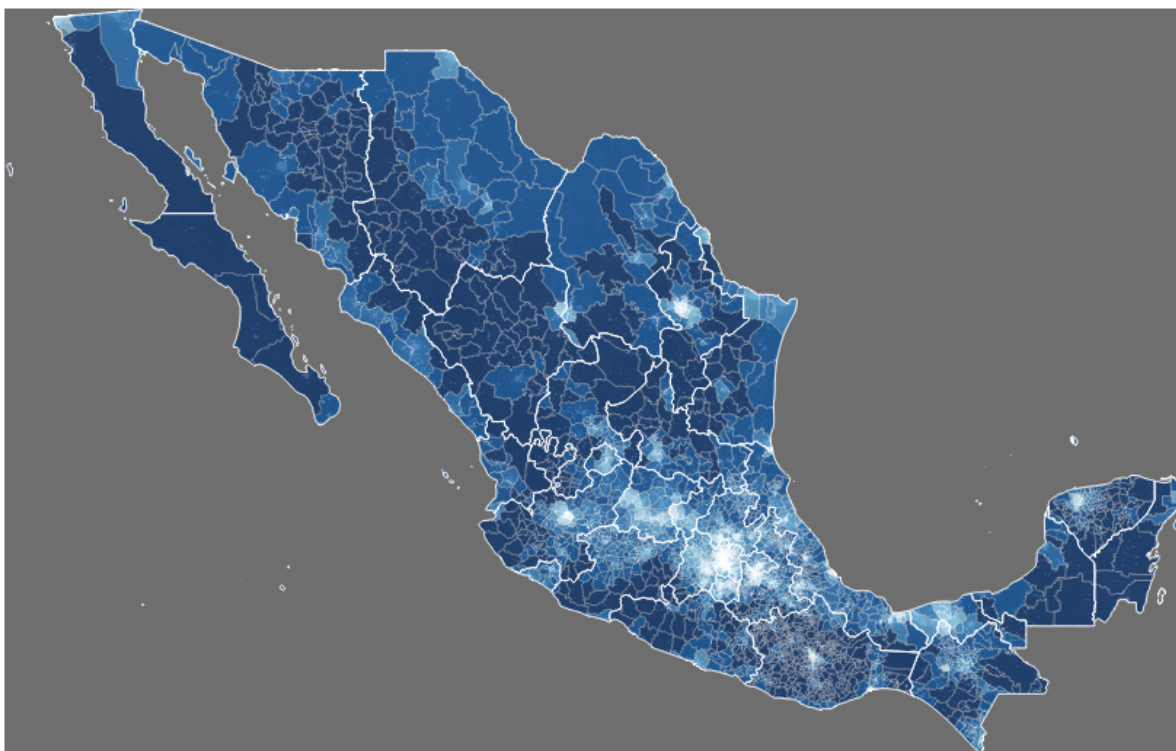


Figura 2: Luminosidad municipal

La intensidad del color de los municipios está relacionada con su luminosidad logarítmica. Además se indican las fronteras de las entidades federativas, a partir de la cual se estima el modelo.

En los siguientes apartados describimos los detalles técnicos.

## Niveles de actividad económica

Este cálculo utiliza los datos de PIBE de los estados y de luminosidad urbana de los municipios. Para un periodo inicial  $t_0 = 2014$ , se toman los datos correspondientes y se distribuye el PIBE con respecto a la luminosidad. Para las ciudades o zonas metropolitanas se agregan los municipios correspondientes.

Para cada municipio  $m \in \mathcal{M}$  se calcula su actividad económica como

$$\varrho_{t_0}(m) = \frac{\Lambda^U(m)}{\Lambda^U(e)} \pi_{e,t_0}$$

donde  $e = e(m)$  es el estado al que pertenece el municipio  $m$  y la luminosidad  $\Lambda^U(\cdot)$  es la correspondiente a la luminosidad urbana libre de combustión con  $\ell \leq 175$  que se introdujo anteriormente.

La actividad económica de las zonas metropolitanas  $c \in \mathcal{C}$  se obtiene sumando sobre sus municipios

$$\varrho_{t_0}(c) = \sum_{m \in c} \varrho_{t_0}(m).$$

En el caso de las metrópolis que se concentran en un solo estado esta estimación es igual a la ponderación por luminosidad del PIBE correspondiente

$$\varrho_{t_0}(c) = \frac{\Lambda^U(c)}{\Lambda^U(e)} \pi_{e,t_0}.$$

Cabe mencionar que éstas no son todas las zonas metropolitanas, pues hay algunas que se dividen en dos o más entidades federativas. Éstas son pocas por lo que podemos mencionarlas: La Laguna, La Piedad-Pénjamo, Puebla-Tlaxcala, Puerto Vallarta y el Valle de México.

Si bien la representación en términos de sus municipios permite estimar la actividad económica para estas zonas metropolitanas, también se controla por dicha separación a la vez que las representamos con sus partes correspondientes.

Después de considerar el nivel de actividad económica en  $t_0$ , se modela los niveles subsecuentes a partir del modelo de crecimiento de las transacciones de cajeros automáticos que explicamos a continuación.

## Crecimiento

El crecimiento de actividad económica está ligado a los datos de transacciones de cajeros automáticos. Esta consideración se puede justificar tanto teóricamente como en la práctica si utilizamos el ITAEE como el indicador base.

En la teoría, las transacciones de cajeros automáticos son una medida simplificada del crecimiento de la actividad económica en la sociedad, y éste tiene una relación dinámica con la producción. En la práctica, el cálculo del ITAEE también utiliza datos de la banca comercial provistos por la CNBV<sup>5</sup>. En la gráfica 3 se ve la similitud de las transacciones con la actividad económica que representa el ITAEE en los estados.

Cuantificamos esta similitud con el siguiente modelo

$$\iota_{e,t}^{\Delta} \sim \alpha_0^e + \alpha_1^e \mu_{e,t}^{\Delta} + \alpha_4^e \mu_{e,t-4}^{\Delta}$$

donde análogamente se tienen los crecimientos proporcionales  $(\cdot)^{\Delta}$ . Los coeficientes  $\alpha_{(\cdot)}^e$  con superíndice están asociados a cada estado y el término  $\alpha_4^e \mu_{e,t-4}^{\Delta}$  captura el efecto estacional.

Los resultados de la estimación los expresamos con  $\hat{\alpha}$ 's como coeficientes y  $\varepsilon$ 's como error

$$\iota_{e,t}^{\Delta} = \hat{\alpha}_0^e + \hat{\alpha}_1^e \mu_{e,t}^{\Delta} + \hat{\alpha}_4^e \mu_{e,t-4}^{\Delta} + \varepsilon_{e,t}.$$

Estos modelos de crecimientos son los que aplicaremos a las zonas metropolitanas para después estimar los valores de MAGDA para las ciudades  $\varrho_t(c)$ .

<sup>5</sup>Indicador Trimestral de la Actividad Económica Estatal. Fuentes y metodologías.



Figura 3: Crecimiento de actividad

## Integración de crecimiento

Después de la estimación del crecimiento de los estados se aplica la relación a las zonas metropolitanas. Sustituyendo los datos observados en estados  $e$  por ciudades  $c$  queda,

$$\varrho_t^\Delta(c) = \hat{\alpha}_0^e + \hat{\alpha}_1^e \mu_{c,t}^\Delta + \hat{\alpha}_4^e \mu_{c,t-4}^\Delta$$

donde  $e = e(c)$  es el estado al que pertenece la ciudad  $c$ , o las partes correspondientes cuando las ciudades pertenecen a más estados.

A partir de los niveles base  $\varrho_{t_0}(c)$  que fueron calculados, se aplica la relación del cambio porcentual  $\varrho_t(c) = \varrho_{t-1}(c)[1 + \varrho_t^\Delta(c)]$  repetidamente y se puede generalizar como

$$\varrho_t(c) = \varrho_{t_0}(c) \frac{\prod_{s \leq t} [1 + \varrho_s^\Delta(c)]}{\prod_{s \leq t_0} [1 + \varrho_s^\Delta(c)]}.$$

Finalmente se juntan dichos niveles en las ciudades que cruzan las fronteras estatales, que nos da el índice MAGDA.

## Resultados

MAGDA tiene una  $R^2$  de 59.7 % en su estimación con ITAEE. Debido a la forma independiente en cada estado, se encontró que la estimación para la Ciudad de México es la más acertada con 84.6 %, mientras que encontramos estados en los que el ajuste es bajo como Campeche que apenas tiene  $R^2$  de 0.06 %. Esto se debe a la alta proporción de actividad petrolera en dicho estado.

Con referencia al ITAEE, calculamos también el crecimiento anual acumulado de MAGDA para los estados que se define como el promedio de los crecimientos con respecto al año anterior. Con fórmulas, si  $\varrho_t^{\Delta,4} = \frac{\varrho_t - \varrho_{t-4}}{\varrho_{t-4}}$  es el

crecimiento con respecto al año anterior, el crecimiento anual acumulado es

$$\Delta_T(e) = \frac{1}{4} \sum_{t \sim T} \varrho_t^{\Delta,4}(e)$$

donde  $T$  es el año en cuestión y la suma se toma sobre los trimestres asociados  $t \sim T$ .

A continuación se muestra las tablas para las zonas metropolitanas de MAGDA y su crecimiento anual acumulado.



Id	Zona metropolitana	MAGDA 13	MAGDA 14	MAGDA 15	Cambio 14	Cambio 15
1	Aguascalientes	156 593.3	172 290.4	179 781.5	10.0	4.3
2	Tijuana	171 041.2	175 627.3	184 787.3	2.7	5.2
3	Mexicali	222 352.5	225 315.0	230 129.6	1.3	2.1
4	La Laguna	265 039.6	271 718.1	276 790.2	2.5	1.9
5	Saltillo	139 524.4	142 135.4	137 577.8	1.9	-3.2
6	Monclova	66 264.9	68 581.1	69 740.7	3.5	1.7
7	Piedras Negras	46 278.0	43 976.0	49 826.6	-5.0	13.3
8	Colima	48 709.8	50 554.3	51 819.0	3.8	2.5
9	Tecomán	11 008.6	11 794.5	12 255.9	7.1	3.9
10	Tuxtla	101 494.8	102 481.2	105 476.1	1.0	2.9
11	Juárez	218 911.2	227 321.6	247 744.2	3.8	9.0
12	Chihuahua	128 991.2	136 924.2	149 324.9	6.2	9.1
13	ZMVM	3 785 685.7	3 871 794.6	3 997 769.1	2.3	3.3
14	León	215 268.3	227 248.8	242 195.1	5.6	6.6
15	San Francisco	22 217.4	22 880.9	24 574.0	3.0	7.4
16	Acapulco	103 740.3	99 554.7	98 179.4	-4.0	-1.4
17	Pachuca	77 378.7	80 260.0	83 711.6	3.7	4.3
18	Tulancingo	23 669.8	23 523.2	24 517.9	-0.6	4.2
19	Tula	46 422.4	47 151.4	39 752.3	1.6	-15.7
20	Guadalajara	734 893.4	759 680.1	795 853.0	3.4	4.8
21	Puerto Vallarta	58 604.3	58 029.3	59 568.2	-1.0	2.7
22	Ocotlán	13 502.9	13 765.8	13 955.8	1.9	1.4
23	Toluca	204 266.4	211 753.9	220 260.8	3.7	4.0
24	Morelia	102 804.3	106 587.1	108 089.1	3.7	1.4
25	Zamora	20 832.8	21 975.9	21 454.3	5.5	-2.4
26	Pénjamo	25 267.4	26 110.2	26 380.8	3.3	1.0
27	Cuernavaca	97 260.7	97 350.1	97 376.2	0.1	0.0
28	Cuautla	53 166.6	53 840.0	53 451.8	1.3	-0.7
29	Tepic	56 653.8	57 654.3	59 502.5	1.8	3.2
30	Monterrey	1 003 410.3	1 043 565.9	1 080 284.9	4.0	3.5
31	Oaxaca	79 542.9	81 852.5	82 685.8	2.9	1.0
32	Tehuantepec	35 971.1	36 833.2	37 059.9	2.4	0.6
33	Puebla	340 600.2	344 533.5	348 243.8	1.2	1.1
34	Tehuacán	27 448.7	27 346.9	28 123.2	-0.4	2.8
35	Querétaro	251 230.3	265 207.8	287 395.3	5.6	8.4
36	Cancún	159 655.0	153 772.2	158 634.2	-3.7	3.2
37	San Luis	204 463.9	206 884.6	211 894.8	1.2	2.4
38	Cd. Fernández	9 268.1	9 339.9	9 173.0	0.8	-1.8
39	Guaymas	32 223.0	35 262.3	35 608.7	9.4	1.0
40	Villahermosa	219 290.7	220 488.4	222 092.1	0.5	0.7
41	Tampico	94 680.8	96 801.4	97 655.5	2.2	0.9
42	Reynosa	143 709.5	151 369.0	156 542.1	5.3	3.4
43	Matamoros	77 252.0	79 371.3	78 210.1	2.7	-1.5
44	Nuevo Laredo	106 954.2	106 640.6	107 702.7	-0.3	1.0
45	Tlaxcala	48 193.4	48 557.5	50 358.6	0.8	3.7
46	Veracruz	169 547.3	164 142.5	170 142.7	-3.2	3.7
47	Xalapa	77 922.9	81 370.0	77 001.0	4.4	-5.4
48	Poza Rica	57 882.0	54 759.0	58 031.9	-5.4	6.0
49	Orizaba	63 432.6	63 418.4	62 245.1	-0.0	-1.9
50	Minatitlán	90 851.8	88 169.7	91 648.8	-3.0	3.9
51	Coatzacoalcos	132 819.5	127 480.6	130 441.5	-4.0	2.3
52	Córdoba	23 872.9	23 611.6	23 554.2	-1.1	-0.2
53	Mérida	185 467.1	190 835.1	196 067.6	2.9	2.7
54	Zacatecas	49 495.1	51 973.2	54 162.9	5.0	4.2
55	Ensenada	52 851.3	53 153.4	54 620.4	0.6	2.8
56	La Paz	57 315.0	54 449.0	52 853.2	-5.0	-2.9
57	Los Cabos	40 731.6	42 977.6	44 777.9	5.5	4.2
58	Campeche	278 823.3	272 618.3	261 470.9	-2.2	-4.1
59	Cd. del Carmen	225 105.3	222 263.1	213 270.2	-1.3	-4.0
60	Manzanillo	33 429.4	35 034.1	36 142.0	4.8	3.2
61	Tapachula	21 538.2	21 699.3	21 604.2	0.7	-0.4
62	Durango	63 749.9	66 386.9	69 134.2	4.1	4.1
63	Celaya	86 191.0	92 227.3	97 490.8	7.0	5.7
64	Guanajuato	22 286.2	23 697.9	25 195.2	6.3	6.3
65	Irapuato	67 502.7	73 739.6	80 489.1	9.2	9.2
66	Salamanca	47 588.8	51 361.5	56 266.7	7.9	9.6
67	Uruapan	32 486.1	33 347.6	31 929.7	2.7	-4.3
68	San Juan del Río	42 921.7	45 669.0	49 259.2	6.4	7.9
69	Los Mochis	51 662.2	49 300.2	50 440.5	-4.6	2.3
70	Culiacán	122 377.7	130 614.1	142 390.0	6.7	9.0
71	Mazatlán	75 874.8	75 422.8	75 824.1	-0.6	0.5
72	Cd. Obregón	54 357.6	50 351.4	60 196.9	-7.4	19.6
73	Hermosillo	170 813.3	160 501.0	164 437.1	-6.0	2.5

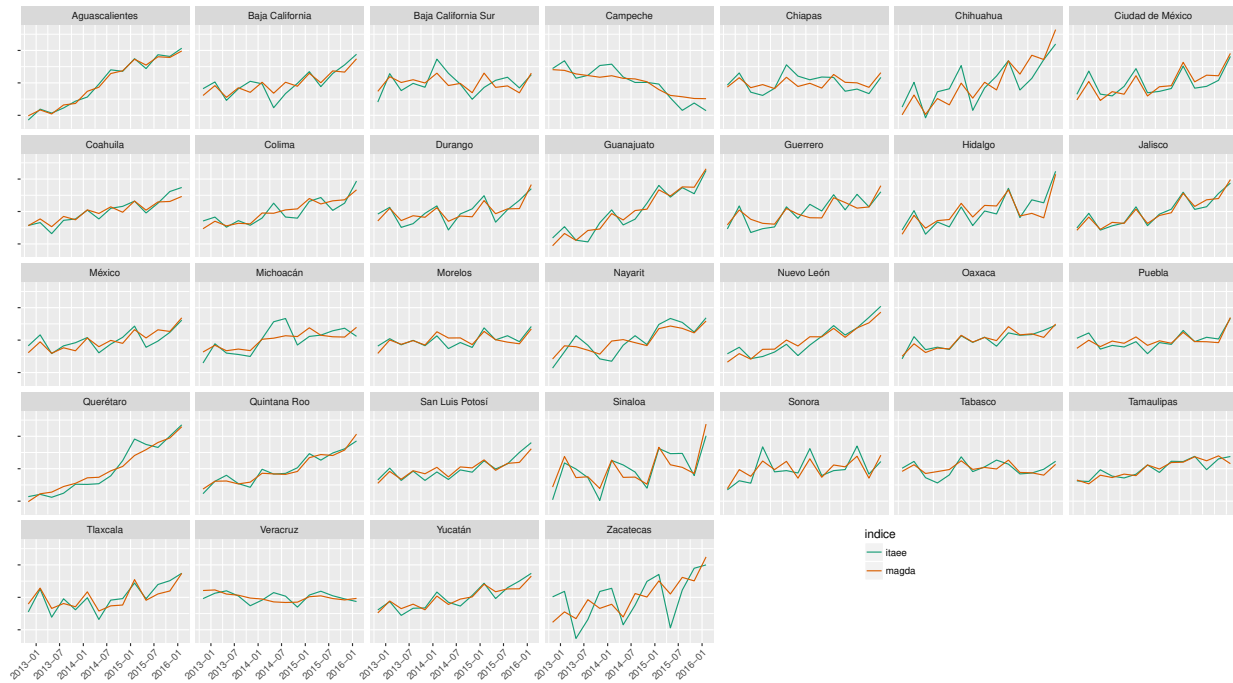


Figura 4: Comparación estados

## Apéndice

### Serie de transacciones bancarias

Se aplicaron filtros estadísticos a las 6,000 series de demanda de efectivo de la CNBV para que fueran útiles para este análisis. La precisión de las estimaciones mejorará en la medida en que los bancos comerciales produzcan mejor información del uso de cajeros automáticos y la reporten correctamente a la CNBV, o estos reportes se integren y publiquen de manera enteramente automatizada.

A continuación se detalla el proceso de estos datos.

1. Las series de transacciones bancarias se descargan del sitio de CNBV. Siguiendo el formato del archivo se separan los distintos bancos  $b \in \mathcal{B}$ , para cada mes  $\tau \in \mathcal{T}$  y cada colonia  $\zeta \in \text{Col}$ .
2. Se agregan las colonias en los municipios  $\zeta \sim m$  y distinguen las series de bancos principales, Banamex y BBVA Bancomer. Con ello se obtienen las series básicas  $\mu_{\tau}^{m,\beta}$  donde  $\beta$  es una de cuatro opciones: banamex, bbva\_bancomer, demás\_bancos o todos\_bancos. Estos grupos se determinaron de acuerdo al volumen de las transacciones y su presencia en los municipios de todo el país. BBVA Bancomer y Banamex en su conjunto representan la mayoría de las transacciones para casi todos los estados.
3. Para cada municipio y grupo de bancos, se genera una serie que se le aplica el proceso de análisis X11<sup>6</sup> que la descompone en tres partes

$$\mu_{\tau} = \mu_{\tau}^{\text{tend}} + \mu_{\tau}^{\text{ciclo}} + \mu_{\tau}^{\text{out}}$$

cuyas componentes tend, ciclo, out corresponden a tendencia, ciclos y outliers. Se reconstruya la serie omitiendo el tercer componente

$$\mu_{\tau}^{\text{sin.out*}} = \mu_{\tau}^{\text{tend}} + \mu_{\tau}^{\text{ciclo}}.$$

<sup>6</sup><http://www.census.gov/srd/www/x13as/>

Este proceso corrige anomalías de pocos periodos, sin embargo no corrige anomalías de periodos prolongados.

4. Profundizando en el análisis, se compara la desviación proporcional de las nuevas series con sus originales para encontrar aquellas que tengan más de la mitad de las desviaciones mayores que cierto corte. Sea  $\nu$  la mediana de las desviaciones

$$\nu = \text{mediana} \left\{ \frac{\mu_{\tau}^{\text{sin.out*}} - \mu_{\tau}}{\mu_{\tau}} \mid \tau \in \mathcal{T} \right\},$$

de tal forma que sólo se ajustan las series en las que la mediana  $\nu_{m,\beta} < 0.7$ . Este corte se encontró bajo inspección de las series anómalas.

Nos queda la serie sin *outliers*,

$$\mu_{\tau}^{\text{sin.out}} = \begin{cases} \mu_{\tau}^{\text{sin.out*}} & \text{si } \nu \leq 0.7 \\ \mu_{\tau} & \text{si } \nu > 0.7 \end{cases}$$

que se agregan para cada trimestre

$$\mu_t^{\text{sin.out}} = \sum_{\tau \sim t} \mu_{\tau}.$$

5. Para seleccionar la serie de cada municipio con respecto a los posibles grupos de bancos, se agrupan por estado y comparan con las series de ITAEE. Es decir para un estado  $e \in \mathcal{E}$  considérense las series sin *outliers*

$$\mu_t^{e,\beta} = \sum_{m \subseteq e} \mu_t^{m,\beta}$$

donde ahora  $\beta$  es alguna de banamex, bbva\_bancomer, demás\_bancos, y algunas combinaciones de sus sumas.

Se elige como base la  $\beta$  que corresponde a banamex, dado que es la más estable, y añadimos otros grupos dependiendo de acuerdo al ajuste de correlación con el ITAEE. La serie final de cada municipio se toma de acuerdo a la  $\beta_e$  correspondiente

$$\mu_t^{\text{banco}} = \mu_t^{\beta_e}.$$

Con este procedimiento se calculan las series de transacciones de cajeros y se modela el crecimiento de MAGDA.

## Bibliografía

Adrián González-Estrada, Gilberto Gallegos-Cedillo. 2014. "The Gross Domestic Product of Municipalities of Mexico: II. Estates M-Z." *Revista Mexicana de Ciencias Agrícolas* 5 (8): 1405–21.

Christopher D. Elvidge, Mikhail Zhiszhin, Kimberly Baugh. 2013. "Why Viirs Data Are Superior to Dmsp for Mapping Nighttime Lights." *Proceedings of the Asia-Pacific Advanced Network* 35: 62–69.

J. Vernon Henderson, David N. Weil, Adam Storeygard. 2012. "Measuring Economic Growth from Outer Space." *American Economic Review* 102 (2): 994–1028.

Neal Jean, Michael Xi, Marshall Burke. 2016. "Combining Satellite Imagery and Machine Learning to Predict Poverty." *www.sciencemag.org* 353 (6301): 790–95.

Petricioli, Carlos. 2015. "México Desde El Espacio: Cómo Obtener, Procesar Y Calibrar Una Serie de Tiempo de Imágenes Satelitales de Iluminación Nocturna Para Estimar El Crecimiento Económico de Las Entidades Federativas." B.S. Thesis. Rio Hondo #1, México D.F.: Instituto Tecnológico Autónomo de México.

"Sistema de Cuentas Nacionales de México." 2013. *INEGI*.

Tilottama Ghosh, Christopher D. Elvidge, Rebecca L. Powell. 2010. "Shedding Light on the Global Distribution of Economic Activity." *The Open Geography Journal*, no. 3: 148–61.

Xi Chen, William D. Nordhaus. 2011. "Using Luminosity Data as a Proxy for Economic Statistics." *Proceedings of the National Academy of Sciences* 108 (21): 8589–94.